Fusing Home Infrastructure Sensors with Smart Home Sensors for Automatic Fixture Level Disaggregation

ABSTRACT
Light fixtures account for 11% of energy usage in homes and are good indicators of activities and room usage. Despite their essential role in the home, however, it remains a challenge to monitor the usage of light fixtures. This paper presents a new approach called LightSense that uses data fusion to cheaply and accurately infer light fixture usage. It combines data from a single power meter on the electrical mains and a light sensor in each room, neither of which is sufficient alone. LightSense is evaluated using in-situ sensor deployments in 4 homes for 10 days each, monitoring 41 light fixtures across a total of 27 rooms. It detects light fixture events with 81% recall and 86% precision, and estimates the total energy used by each fixture with 91.1% accuracy on average.

Smart water meters will soon provide real-time access to instantaneous water usage in many homes, and disaggregation is the problem of deciding how much of that usage is due to individual fixtures in the home. Existing disaggregation techniques require additional water sensing infrastructure and/or a manual characterization of each water fixture, which can be expensive and time consuming. In this paper, we describe a novel technique called WaterSense to perform fixture-level disaggregation using only a handful of inexpensive motion sensors. WaterSense automatically infers how many fixtures are in each room, and how much water each fixture uses. We evaluate the system using a 7-day in-situ evaluation in 2 diverse multi-resident homes with a total of 10 different water fixtures and 467 fixture events and observe that our approach achieves 86% classification accuracy in identifying individual water fixture events and 80-90% accuracy in determining the water consumption of individual water fixtures.

Author Keywords
ACM Classification Keywords
H.5.2 :

General Terms
INTRODUCTION
Set up motivating scenario. Proliferation of smart meters and smart home sensors. Need for fixture level usage and disaggregation for energy and activity recognition applications.

Existing research focusses exclusively on infrastructure sensing, or deploying specialized sensors per appliance along with infrastructure sensors. Existing techniques also require supervision. Lack of effective data fusion approaches to automatically infer number of fixtures, usage times, and resource consumption from existing infrastructure and distributed sensors. *also* - our data fusion approaches are not necessarily complete substitutes for these other approaches, but are orthogonal and could be used in combination with more sophisticated sensors to improve accuracy or reduce training.

we present a general framework for fusing data from infrastructure and indirect sensors in homes based on our experiences with fixture disaggregation. highlight key challenges faced and solutions used.

we evaluate our framework through two case studies - light fixture and water fixture disaggregation. highlight any key differences of these two instances of the framework.

summarize deployments, evaluation results.

Light fixtures play an essential role in the home, and monitoring their usage can serve many purposes. For example, task lighting near a chair, countertop, or table can be a good indicator of activities of daily living (ADLs) such as reading, preparing food, and eating dinner. In fact, light fixture usage may be the primary indicator of certain activities, at least from a sensing perspective. For example, the status of the light in a bedroom may be an indicator of certain activities, at least from a sensing perspective. Monitoring light usage is also important from an energy conservation viewpoint. Light fixtures account for 11% of energy consumption in US homes on average[4] and 25% in California because heating and cooling energy is lower due to the mild climate [20]. This energy usage is equivalent to that used by 15 million homes or about 20% of all automobiles on the road in the US. Homeowners can reduce lighting energy by installing energy-efficient CFL or LED bulbs, daylight harvesting hardware, or occupancy-automated lighting controls. However, these solutions are costly relative to the energy saved: some of the homes we analyzed contained over 50 bulbs, including specialized lighting (e.g. recessed lighting or under-cabinet lighting) that is partic-
ularly costly to retrofit. Homeowners need concrete data about individual light fixture usage to quantitatively assess the expected return on investment of any given solution.

Unfortunately, light fixture monitoring remains a challenge today. Non-intrusive load monitoring (NILM) approaches have long been used to identify usage of major appliances by analyzing their electrical signature, using only a single power meter on the electrical mains of the home. However, NILM is not effective for low-power appliances such as light fixtures: the wattage of an incandescent bulb (30-100W) is similar to the electrical noise levels of a typical home. Recently, several alternatives have been developed, but each has important limitations. For example, they require a sensor to be installed on every fixture [11, 19, 14, 12], require a manual training and labeling period [17], or are limited to light fixtures that involve switched-mode power supplies (SMPS) [8] such as those with CFLs or dimmer switches.

In this paper, we present a new approach called LightSense that uses data fusion to cheaply, easily, and accurately infer light fixture usage. In addition to a power meter on the electrical mains of the home, LightSense uses a light sensor in each room to capture both the electrical signature and the light signature of each fixture. For a typical home, this approach requires only 3-8 more sensors than conventional NILM techniques. Furthermore, light sensors are inexpensive and can be easily installed without touching the electrical wiring. However, light sensors alone are not sufficient to detect fixture usage: the light levels produced by a typical fixture are similar to or even smaller than the light levels caused by window shades opening and closing, people walking past the sensor, and clouds passing overhead. Light fixtures account for only a small fraction of the light level changes in a room. To address this problem, LightSense exploits several insights about light fixture usage: 1) noise on the power lines and noise in light are typically independent 2) light fixtures typically exhibit a similar power/light signature over time, even when using dimmers 3) light usage always occurs in ON/OFF pairs: a light cannot be turned ON twice in a row. In this paper, we present a probabilistic data fusion framework that combines power and light data streams, automatically learns the long-term signature of each light fixture, and accurately recognizes the real-time usage events.

LightSense is completely unsupervised: it does not require a training process in which users manually label light fixture usage. Furthermore, it does not require user configuration: users do not need to indicate the number of fixtures in a home or their individual wattages. These parameters are automatically inferred. We envision two types of LightSense deployments. In a short-term energy audit, light sensors would be placed on a shelf or countertop in each room for a period of several days or weeks to assess the total energy usage of each light fixture. Power data would be collected from a smart power meter, which are already deployment in many parts of the US by power utility companies for billing purposes. For longer-term activity recognition applications, the light sensors would simply be integrated with other sensors. For example, several commercial motion sensors already have integrated light sensors for calibration purposes [?, ?]. If a motion sensor is placed in every room for activity recognition purposes anyway, the light sensor data can be used by LightSense to infer light fixture usage.

We evaluate LightSense with an in-situ deployment of power meters and light sensors in four multi-resident homes over 10 days each, monitoring a total of 41 light fixtures across 32 rooms. We deployed wireless light switches and smart plugs to collect ground truth information about light fixture usage. We found that the power meter alone detects light fixture events with only 15% precision, and the light sensors alone achieve only 25% precision. By combining the two data streams, however, the LightSense data fusion algorithm detects ON/OFF events with 81% recall and 86% precision. We also demonstrate that LightSense can disaggregate whole-house energy usage into fixture-level energy usage, estimating the total energy usage of most the commonly used light fixtures with 91.1% accuracy on average.

The world’s usable water supply is decreasing at a faster rate than it can be replenished. Household water conservation is important to ensure sustainability of fresh water reserves, to save energy from water treatment and distribution, and to prevent fresh water habitats from being affected through excessive water use [?]. Residents have a number of practical options to conserve water, ranging from replacing high flow toilets and showers with low flow replacements, to reducing water used for daily activites such as brushing teeth or washing dishes. To make informed decisions that maximize water savings, households first need a detailed understanding of how much water is used by each appliance and water fixture in the home.

Water utilities are increasingly installing smart water meters that provide real-time access to household water consumption, and 31 million smart water meters are expected to be installed by 2016 [?]. However, these meters are installed at the water mains and only provide aggregate water usage, primarily for billing purposes. Disaggregation is the problem of deciding how much of that usage is due to individual fixtures in the home, and is challenging because homes often have multiple sinks, toilets, showers, and other fixtures that produce similar rates of flow. For this reason, existing disaggregation techniques require additional sensing on the water piping infrastructure, and/or a manual characterization of each water fixture [5, 6, 7, 13, 16]. These techniques can be expensive, difficult to deploy, and time consuming.

In this paper, we present the WaterSense system that performs fixture-level disaggregation of smart water meter data using only simple motion sensors. Motion sensors are inexpensive ( $5 each), easy to install, and already prevalent in many homes as part of home security or home automation systems [?]. WaterSense does not require any additional sensing infrastructure on the water pipes or fixtures, and disaggregates fixtures in an unsupervised manner that does not
require the collection of training data. The WaterSense technique is based on two basic insights: 1) fixtures with similar flow signatures (e.g. identical toilets) are often in different rooms, and 2) fixtures in the same room often have different flow signatures (e.g. a toilet vs. bathroom sink). Based on these insights, WaterSense clusters all water usage events based on both flow signatures and motion sensor signatures, and each of these clusters represents a unique water fixture in the home. One limitation of this technique is that it is not likely to differentiate two identical fixtures in the same room, such as double sinks in a bathroom. However, such distinctions may also be less important for the purpose of water conservation decisions. We use a novel Bayesian clustering algorithm to create robust clusters despite noise in the motion sensor data caused by both wireless packet loss and multiple residents moving in multiple rooms simultaneously. We deployed the WaterSense system in two different, natural home environments for 7 days each and found that the system can disaggregate flow at the water mains to individual fixtures with an average 86% classification accuracy.

**RELATED WORK**

Several approaches have been developed for monitoring light fixture usage, but each has important limitations. Figure 1 lists the main techniques in this area, and compares them to LightSense in terms of several important metrics.

**Non-Intrusive Appliance Load Monitoring (NIALM)** [9] techniques can be used to identify the usage of electrical fixtures in the home based on electrical signatures. These signatures can be extracted using only a single power meter [7], which is already available in many homes today [2]. Some appliances have a unique profile of real and reactive power, while other appliances such as washing machines and dishwashers exhibit characteristic electrical patterns over time. However, NIALM techniques not effective for light fixtures that exhibit constant, low power values due to the large number of similar, low power appliances in homes, and due to low power state transitions from complex appliances such as the television or the HVAC system [9].

An alternative to NIALM is use **in-line current sensors** to directly measure the amount of energy consumed by individual appliances. Examples include pluggable power meters [7, 11] to monitor lamp fixtures, and smart switches [2] to monitor lights controlled by wall switches. However, direct sensing a large number of sensors to be integrated with the home’s electrical wiring, which is costly in terms of both hardware and installation time. In our own deployments across four homes, for example, installing in-line current sensors for ground truth consumed approximately 19 man hours per home on average, while the LightSense installations only consumed an average of 25 minutes per home.

Recently, new solutions have been developed that require only a single sensor on the power line. Gupta et al propose the **Electrisense system** [8], which uses an easy to install, plug-in sensor that leverages unique high frequency EMI (Electromagnetic Interference) signals on the power line to identify light fixtures. However, this approach is limited to appliances that use switched-mode power supplies (SMPS) such as lights with CFL bulbs or dimmer switches; it does not work for conventional incandescent bulbs. Patel et al [17] propose a related approach to identify **mechanically switched appliances** in homes based on a unique high frequency voltage transients. However, this approach requires users to manually train the system by labeling ON and OFF events of individual appliances. If lights are added or moved throughout the house, training must be performed again. Furthermore, the authors only evaluate the ability to classify manually labeled events that are explicitly input to the classifier (recall and classification accuracy). It is unclear if the reported upper bound classification accuracy for these approaches of 75-95% will be retained in a realistic, in-situ evaluation, where *false positives* in the power line signals could reduce precision. Moreover, neither of these approaches can currently infer the energy consumption of individual fixtures; they can only identify which fixture was used. LightSense is a completely unsupervised technique that can identify fixtures and measure their energy consumption. LightSense is also evaluated in a *in-site* experiment.

The systems most similar to LightSense include Ambient sensing approaches that fuse data from whole-house power monitors with environmental sensors deployed throughout the house. For example, Viridiscope [14] uses one specialized sensor node per appliance containing light, acoustic, and magnetic sensors with whole house power meter data to infer the energy costs of individual appliances in the home. However, the Viridiscope algorithm makes the strong assumption that the power consumed by unmonitored appliances is constant, essentially requiring complete coverage of every electrical fixture and appliance in the home. A contactless sensing system proposed by Rowe et al [19] also requires specialized environmental sensors per appliance, and additionally uses high frequency circuit level power meter.

<table>
<thead>
<tr>
<th>Technique/ Feature</th>
<th>Easy to Install?</th>
<th>Cost</th>
<th>Energy feedback?</th>
<th>Accuracy</th>
<th>User training?</th>
<th>Applicability?</th>
<th>In-site evaluation?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrisense</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>High</td>
<td>No</td>
<td>CFUs</td>
<td>No</td>
</tr>
<tr>
<td>Using mechanical properties of switches</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>High</td>
<td>Yes</td>
<td>All</td>
<td>No</td>
</tr>
<tr>
<td>NIALM</td>
<td>Yes</td>
<td>Low</td>
<td>Yes</td>
<td>Very Low</td>
<td>Yes</td>
<td>All</td>
<td>Yes</td>
</tr>
<tr>
<td>Direct sensing</td>
<td>No</td>
<td>High</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>All</td>
<td>No</td>
</tr>
<tr>
<td>Ambient sensing</td>
<td>No</td>
<td>High</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>All</td>
<td>No</td>
</tr>
<tr>
<td>Light Sense</td>
<td>Yes</td>
<td>Low</td>
<td>Yes</td>
<td>High</td>
<td>No</td>
<td>All</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 1. Comparison of existing techniques to infer usage and power consumption of light fixtures.
data, which requires high installation effort at the breaker box. Both of these approach require far more sensors and a more involved installation process than LightSense. LightSense can detect light fixture usage without the need to monitor any other fixtures or appliances in the home. Further, it does not require a sensor on each fixture, but only requires full coverage of all light fixtures of interest (typically, one sensor per room).

Several approaches have been proposed in the literature to infer fine-grained water fixture use in homes. The most basic approach is to use flow signatures, such as flow rate, flow duration or, in the case of washing machines and dishwashers, patterns of flow to identify types of fixtures and appliances [16]. However, flow signatures alone cannot disambiguate between different instances of identical sinks, toilets, or showers in the same home. Fogarty et al [5] use microphones installed in a basement to classify water fixture use, and achieve good accuracy in identifying high consumption appliances, but low accuracy in differentiating between different instances of the same fixture category. Another recent approach [13] uses vibration sensors on pipes to disaggregate total flow measured at a central location in a novel, unsupervised manner. Both of these techniques, however, require additional sensing of the water infrastructure, which may require access to pipes in crawl spaces or walls. Furthermore, microphones and accelerometers are more power intensive than motion sensors and would either have a short battery life or would require wired power. Finally, microphones may be considered privacy-invasive sensors.

Patel et al [6] avoid extensive sensing by using a single water pressure sensor that samples at 500Hz, plugged into a free spigot or water outlet in the home. However, this approach requires significant training data on the order of several days in a real world setting [7] to achieve high accuracy in inferring individual water fixture events. Our approach is an unsupervised technique that does not require training data.

A DATA FUSION ALGORITHM
(flow diagram architecture figure)

Tier I: Event Detection
Do automatic edge detection on each infrastructure and indirect sensor. highlight key common challenges.

Light Fixtures
discuss briefly about edge detection challenges for light sensor and power meter.

Water Fixtures
discuss briefly about edge detection for water flow meter and motion sensor.

Tier II: Data Fusion
Motivate two main reasons why data fusion is important: 1. eliminate false positives 2. infer correct power/water consumption for a fixture event for energy applications

use temporal distance as a criteria for data fusion, since the assumption is that fixture usage causes both indirect and infrastructure sensor events within a small time window. but indirect sensor events such as motion sensors are not always in sync with infrastructure events , or exact time sync may not be possible because infrastructure and smart home sensor might be from different systems.

Key challenge: dealing with a large number of possibilities for computing infrastructure sensor, direct sensor; event pairs due to high false positive rates from either direct sensors or infrastructure sensors or both.

Solution: Bayesnet approach to automatically keep track of commonly occuring pairs of infrastructure, direct sensor; events incorporating both instantaneous and historical sensor evidence

Light Fixtures
data fusion eliminates many false positives from both light sensor and power meter. bayesnet addresses high false positives from power meter and computing correct light, power; event pairs

(example data figure from light sense)

Water Fixtures
bayesnet helps disambiguate from simultaneous events from motion sensors in fixture rooms.

(example data figure from water sense)

Tier III: Event Matching
goal: match ON and OFF fixture usage events. discuss bipartite matching approach.

Light Fixtures
use both probability of power and light reading matching along with joint prob from bayesnet in tier II

Water Fixtures
matching is more trivial since simultaneous use is rare

Tier IV: Fixture Clustering
goal: automatically compute set of fixtures sensed by each indirect sensor using clustering approaches.

Light Fixtures
gives set of light fixtures. possible to compute nominal wattage, power consumption, usage.

Water Fixtures
gives set of water fixtures. use simple heuristics to classify as sink or flush or shower.

A CASE STUDY: MOTION AND LIGHT SENSORS
Inferring Water Fixture Usage
Inferring Light Fixture Usage
To detect whole house power consumption edges from the power meter time series $P$, we apply a sliding window based edge detection algorithm, that computes all possible power edges with a maximum window bound. In particular, each power edge $e$ is defined by a time window $[e_f, e_s]$, and the edge value is given by $e_v = (P[e_f] - P[e_s])$. We place two main constraints on time windows which can constitute a power edge $e$:

1. $(e_f - e_s) \leq \text{maxwinP}$
2. $|e_v| \geq dP$

Condition 1 ensures that all edges have a time window length bounded by $\text{maxwinP}$. A small parameter for $\text{maxwinP}$ parameter ensures that we eliminate a significant number of slow power intensity changes that do not originate from artificial lighting. In our current implementation, we fix $\text{maxwinP} = 5$ seconds across all homes. *Condition 2* determines the lowest wattage light fixtures detected by LightSense, parameterized by $dP$; in section 2, we evaluate the recall precision tradeoff as $dP$ is varied.

Our power edge detection algorithm is a multi-edge detection approach, since multiple power edges could simultaneously be detected at any time instant from the different time window lengths bounded by $\text{maxwinP}$. We adopt this approach because the power meter data typically contains several closely spaced simultaneous low power events, and a single edge window is insufficient to achieve high recall; a larger edge window is ineffective when there are closely spaced events ON or OFF events, while a smaller edge window does not capture slowly increasing power events from dimmable lights as reported by the power meter, which reports rms power every second from the AC current and voltage inputs. We address the problem of pruning power edges in section 2, by leveraging the joint light and power signatures of light fixtures. Also, we bound the maximum power edge magnitude $|e_v|$ to 1000W to prevent interference from high power appliances such as the stove or the washing machine. The edges output by our edge detection algorithm are partitioned to rising and falling power edges $RP$ and $FP$, respectively, and the edge value $e_v$ is set to $|e_v|$ in the partitioned subsets.

Figure 3 shows the timestamps of edges output by the power edge detection algorithm. We observe that the power edge detection approach shows poor precision in detecting light fixture events, due to false positives from numerous low power appliances, and low power state transitions of complex appliances such as the HVAC system or the television.

**Light Edge Detection**

To detect light edges from each light sensor stream, we design the light edge detection algorithm outlined below. Figure 2 shows the timestamps of rising(ON) and falling(OFF) edges output by the two steps in the light edge detection algorithm. The first step uses a sliding window-based edge detection algorithm similar to the power edge detection algorithm to eliminate noise due to slowly changing light events, such as daylight changes from the movement of the sun. The second step uses three novel adaptive noise filters to filter the
typical noise events seen in home light sensors.

\[
\text{for each light sensor } i \in S \\
L^p_i = \text{medfilt}(L^i, \text{window}) \\
E^i = \text{windowdetect}(L^p_i, \text{maxwin}, dL) \\
F^i = \text{null}
\]

\[
\text{for each } e \in E^i \\
b = \text{noise\_filter}(e) \land \text{onoff\_filter}(e) \land \text{consensus\_filter}(e) \\
\text{if } b = 1 \text{ then add } e \text{ to } F^i
\]

partition \(F\) to \((RL, FL)\)

In the algorithm above, for every light sensor \(i \in S\), where \(S\) is the set of light sensors in the home, we first perform median filtering on the light sensor time series \(L^i\) with a fixed window size of 10 samples across all four homes. Median filtering eliminates impulse noise from our light sensors. We then apply a window-based edge detection algorithm \(\text{windowdetect}\) on the smoothed time series \(L^p_i\) to output a sequence of edges \(E^i\). This algorithm uses the two conditions mentioned above for the window-based power edge detection algorithm, with new parameters \(dL\) and \(\text{maxwinL}\) instead of parameters \(dP\) and \(\text{maxwinP}\). Setting a small bound on the \(\text{maxwinL}\) parameter ensures that we eliminate a significant number of similar light intensity changes that do not originate from artificial lighting. In our current implementation, we fix \(\text{maxwin} = 2.5\) seconds across all homes, since that is the maximum time required for typical residential dimmers to switch to their target light intensity level. Parameter \(dL\) determines the lowest intensity of light edges detected by LightSense; in section 4, we evaluate the recall precision as \(dL\) is varied. We also add two additional constraints on edges \(e \in E^i\) to ensure that multiple light edges are not detected simultaneously:

1. \(\forall e \in E^i, (e \neq d) \rightarrow ((e_x, e_y) \cap [d_x, d_y] = \emptyset)\)
2. \(\forall x, y \in \{L^p_i[y] - L^i[x] \geq dL\} \land ((x, y) \in [e_x, e_y]) \rightarrow (e_x \ast \left\{ (L^p_i[y] - L^i[x]) \right\} > 0)\)

The above conditions together ensure that the edge window sizes are adjusted automatically depending on the variation in the light signal; smaller edge windows are used when light intensity changes are dense, while larger edge windows are used when the light intensity changes are sparse. Due to space constraints, we do not discuss how the \(\text{windowdetect}\) algorithm ensures the constraints above to output \(E^i\), but we note that it is a straightforward algorithm that is in \(O(|L^i|)\).

For each edge \(e \in E^i\) output by the \(\text{windowdetect}\) algorithm, we apply three additional adaptive filters to eliminate noisy edges, namely the \(\text{noise\_filter}\), the \(\text{onoff\_filter}\), and the \(\text{consensus\_filter}\); in figure 3, we see that these adaptive filters eliminate a significant number of noisy edges output by the \(\text{windowdetect}\) algorithm.

Firstly, lights do not turn on and off very frequently; if we detect a large number of similarly sized light edges, we assume they are ambient noise. For each edge \(e\), the \(\text{noise\_filter}\) algorithm first \(\text{counts}\) the number of comparably temporally adjacent edges \(x\) within time \(dT_a\) of edge time \(e_x\), whose magnitude is at least \(\alpha\) times the current edge value \(e_x\) under consideration. The filter returns \(false\) if the number of such comparable edges exceeds a \(\text{maxcount}\) parameter. We fix parameters \(dT_a = 20\) minutes, \(\text{maxcount} = 20\), and \(\alpha = 0.8\) across all homes.

Secondly, when people or clouds pass by the light sensor, a rising edge and falling edge occur closely together in time; we filter these pairs of events out. For each edge \(e\), the \(\text{onoff\_filter}\) algorithm returns false if an \(\text{opposing}, \text{comparably}, \text{temporally adjacent}\) edge \(y\) exists within time window \(dT_b\) of edge time \(e_x\), whose magnitude \(y_x\) is at least \(\alpha\) times the current edge \(e\) under consideration, and whose polarity is \(\text{opposite}\) to edge \(e\). We fix parameter \(dT_b = 12\) seconds.

Thirdly, redundant light fixture events from adjacent rooms are detected in multiple light sensors. So, when we observe light edges in multiple light sensors simultaneously with the same polarity, we only retain the light edge with maximum intensity, since that edge is most likely from the sensor in the same room as the light fixture. Assuming a coarse time synchronization among the multiple light sensor streams based on timestamps at the receiver base station, the \(\text{consensus\_filter}\) eliminates redundant light edges within one second of each other. Finally, the filtered light edges are partitioned into rising and falling edges based on their polarity to \(RL\) and \(FL\), respectively.

**Tier II: Data Fusion and Matching**

The goal of Tier II is to combine the low precision light and power edges from Tier I to compute a sequence of matched ON-OFF events \(M^i\) for each light sensor \(i\). Each ON-OFF event \(m \in M^i\) is defined by a four tuple \((m_s, m_f, m_p, m_l)\); \(m_s\) and \(m_f\) denote the ON and OFF time, respectively, while \(m_p\) and \(m_l\) denote the average power consumption and light intensity increase for event \(m\). Firstly, we use data fusion to eliminate temporally isolated power or light edges as false positives. Secondly, we use Bayesian matching to prune unmatched ON or OFF light and power edges as false positives. Finally, we use a novel Bayesian clustering approach to accurately assign power costs to each ON-OFF event, by automatically learning the typical power and light signatures in a room.

**Data Fusion**

The Data Fusion step combines the rising and falling power edges \(\{RP, FP\}\) and light edges \(\{RL, FL\}\) to compute a set fused light-power edges \(\{RLP^i, FLP^i\}\) for each light sensor \(i \in S\). We first describe how the rising light-power
edges $RLP^i$ are computed from $RP$ and $RL^i$. We first compute a power edge set $PS_i$ for each light edge $e \in RL^i$, where $PS_i$ denotes all the power edges from $RP$ within time $dT_{add}$ of the light edge timestamp $e_s$. If no power edges are found, we discard the light edge $e$, otherwise we add the light-power edge $(e, PS_e)$ to $RLP^i$. Thus, temporally isolated light or power edges are eliminated as false positives, while temporally adjacent light and power edges are added to $RLP^i$. We similarly compute the falling light-power edges $FLP^i$ from $FP$ and $FL^i$. We currently set $dT_{add}$ to 12 seconds across our deployments to account for time synchronization errors between the light sensor and power meter data streams. Since the power meter data is noisy, multiple power edges in $PS_e$ are typically associated with a single light edge $e$; the problem of assigning the correct power edge to each light edge is addressed as part of our Bayesian Clustering step below.

**Bayesian Matching**

For each light sensor $i \in S$, the Bayesian matching step matches rising light-power edges $RLP^i$ to falling light-power edges $FLP^i$ to compute the sequence of matched light fixture events $M^i$. Bayesian matching prunes false positive edges from $RLP^i$ and $FLP^i$ that cannot be matched with high probability. The algorithm is shown below:

```
for each light sensor $i \in S$

Weights$^i$ = null
for each edge $(d, PS_d) \in RLP^i$
for each edge $(e, PS_e) \in FLP^i$

c1 = ($d_s > e_s$)
c2 = $\min(L[d_s : e_s]) > \gamma \ast d_e$
if (c1 \lor c2)
$pmatch_{d,e} = 0$
else
compute $lprob_{d,e} \ast pprob_{d,e}$
pmatch$^i_{d,e} = lprob_{d,e} \ast pprob_{d,e}$
Weights$^i[d,e] = -\log(pmatch^i_{d,e})$
```

$$M^i = \text{Hungarian(Weights$^i$)} \text{ with weight threshold } w_{max}$$

The Bayesian Matching algorithm shown above first constructs a weighted bipartite graph with edges from the set of rising-light-power edges $RLP^i$, to the set of falling-light-power edges $FLP^i$. The edge weight $Weights^i[d,e]$ between any rising edge $d \in RLP^i$ and any falling edge $e \in FLP^i$ is set to $-\log(pmatch^i_{d,e})$, where $pmatch^i_{d,e}$ represents the probability of edge $d$ being matched to edge $e$. We perform a min-cost bipartite matching on the weighted bipartite graph represented by $Weights^i$, which returns an optimal matching $M^i$ with maximum match likelihood, i.e. $M^i = \arg \max_x (\prod_{(d,e) \in x} pmatch^i_{d,e})$.

In the algorithm above, we see how the match probabilities $pmatch^i_{d,e}$ are computed. Firstly, we set $pmatch^i_{d,e} = 0$ if any of the two conditions $c1$ or $c2$ are satisfied. Condition $c1$ ensures that rising edges in an ON-OFF event occur before falling edges. Condition $c2$ ensures that rising and falling edges from two different ON-OFF event pairs are not matched together by leveraging the observed additive nature of light intensities in a room; if $a + 100$ rising edge and $a - 100$ falling edge are separated by a time interval where the total light intensity is only $5$, then the two edges are likely from different ON-OFF events. In the algorithm, $\gamma$ is set based on empirical experiments to $0.8$ across all our deployments. If conditions $c1$ and $c2$ are false, the two edges $d, e$ under consideration could potentially be matched.

As shown in the matching algorithm, the match probabilities $pmatch^i_{d,e}$ are computed as $lprob_{d,e} \ast pprob_{d,e}$, where the light match probability $lprob_{d,e}$ denotes the probability that the light edges $d$ and $e$ are matched, while the power match probability $pprob_{d,e}$ denotes the probability that the two corresponding power edge sets $PS_d$ and $PS_e$ are matched.

We compute the light match probability $lprob_{d,e}$ as the likelihood that ON and OFF light edges $d, e$ belong to the same
light fixture; we observe that ON and OFF light edges from the same two-state light fixture have similar edge values. To leverage this observation, we first cluster all the light edge values from RLP and FLP using the QT (Quality Threshold) clustering algorithm [10] with a relative distance error threshold of 0.25 on the light edge values. QT Clustering outputs a set of light edge clusters $CL_i$ with a fixed cluster assignment for each edge. We assume each light edge cluster has a normally distributed light intensity distribution, and compute the mean and standard deviation for each light cluster $c \in CL_i$ as $cl_{mean}$ and $cl_{std}$, respectively, from the edges assigned to $cl$. The probability $lprob_{d,e}$ that the two light edges $d, e$ belong to the same light cluster $c \in CL_i$ is then computed as:

$$lprob_{d,e} = \sum_{cl \in CL_i} N(d_v, cl_{mean}, cl_{std}) \cdot N(e_v, cl_{mean}, cl_{std}) \tag{1}$$

Computing the power match probabilities $pprob_{d,e}$ is less straightforward since there are typically multiple rising and falling edges in the power edge sets $PS_d$ and $PS_e$. To simplify this problem, we choose the assignment $x \in PS_d, y \in PS$ that maximizes $pprob_{d,e}$ as follows:

$$pprob_{d,e} = \max_{x \in x, y \in y}(vprob_{x,y} \cdot pS(x, d_e) \cdot pS(y, e_e)) \tag{2}$$

In equation 2, the power value match probability $vprob_{x,y}$ denotes the probability that the two power edges $x, y$ are from the same light fixture, while the joint light-power probabilities $pS(x, d_e)$, $pS(y, e_e)$ denote the probability that a light fixture with light edge value $d_e$ or $e_e$ has a power consumption of $x$ or $y$, respectively. To compute the power value match probability $vprob_{x,y}$, we leverage the observation that two-state light fixtures produce ON and OFF power edges with similar values. We cluster all the power edge values in edge sets from RLP and FLP using the QT (Quality Threshold) clustering algorithm with a relative distance error threshold of 0.25 to obtain the set of power edge clusters $CP_i$; from $CP_i$, $vprob_{x,y}$ is computed similar to $lprob_{d,e}$ in equation 1. In equation 2, the light-power probability $pS(x, d_e)$ denotes the joint probability of observing power edge value $x$ together with light edge value $d_e$ from light sensor $i$. We use the Ipbayes Bayesian Clustering approach described below to compute the joint light-power probability values.

Bayesian Clustering

Our Bayesian Clustering approach computes the joint probability $pS(p_k, l_k)$ of observing any power edge value $p_k$ and any light edge value $l_k$ together in the same light-power edge from light sensor $i \in S$. The joint probability must effectively capture the typical light and power signatures that occur together in a room. For example, consider an experiment ON-OFF match from a 100W light bulb, containing a rising edge of $d = (40L, 40W, 103W, 500W)$ and a corresponding falling edge of $e = (41L, 40W, 95W, 1000W)$ sensed by light sensor $i$. We expect $lprob_{d,e}$ to be high for this match, since the two light edges are very likely to be from the same 40L light cluster. Among the multiple power edge assignments possible, the highly likely assignments based on $vprob_{x,y}$ from equation 2 are $(40W, 40W)$ and $(103W, 95W)$. If we use only $vprob_{x,y}$, we incorrectly assign the $(40W, 40W)$ power edges to the ON-OFF event, since they are more likely to be from the same cluster than the $(103W, 95W)$ edges which are more noisy. However, by leveraging the joint probability of occurrence of the light and power meter data streams using Ipbayes, we still make the correct 100W assignment to the ON-OFF match because the 103W event has a much higher likelihood than a 40W event of occurring with the 40L light event, i.e. $pS(103W, 41L) >> pS(40W, 41L)$.

Our Bayesian Clustering approach uses the Ipbayes Bayesnet shown in figure 4 to compute the joint probabilities $pS(p_k, l_k)$; in the Bayesnet, $CL^k$ and $CP^k$ are hidden random variables that denote the cluster to which the light edge value $l_k$ and power edge value $p_k$ belong to, respectively. The Bayesnet effectively captures the probabilistic relationship between individual light or power values and their cluster assignments, and more importantly, the conditional relationship between the light and power clusters, indicating how likely an element from power cluster $CP^k$ occurs given that an element from light cluster $CL^k$ has occurred. From the basic properties of Bayesnets, the joint light-power probabilities $pS(p_k, l_k)$ are computed as follows:

$$pS(p_k, l_k) = \sum_{cL^k} \sum_{cP^k} p(l_k | cL^k) \cdot p(p_k | cP^k) \cdot p(cP^k | cL^k) \tag{3}$$

The conditional probabilities $p(l_k | cL^k)$ and $p(p_k | cP^k)$ are computed using the Gaussian distributions computed for the light and power edge clusters during the QT clustering process described in the previous section, i.e. $p(l_k | cL^k) = N(l_k, cl_{mean}, cl_{std})$ and $p(p_k | cP^k) = N(p_k, cp_{mean}, cp_{std})$. The conditional probability $P(cp_k = a | cl_k = b)$ is computed by frequency counting the number of individual light edges $z$ assigned by QT clustering to light cluster $b$ that have at least one edge in $PS_z$ assigned to power cluster $a$. In other words, if we denote the set of light-power edges assigned to light cluster $b$ as $CA_b$, and the set of light-power edges in which at least one element in the power edge set is assigned to power cluster $a$ as $CA_a$, then:

$$p(cp_k = a | cl_k = b) = |CA_a \cap CA_b| / |CA_b| \tag{4}$$

The power assignments $x, y$ that maximize equation 2 are fixed as the optimal power assignments for the current match $(d, e)$ under consideration. From equations 1, 2, 3, and 4, we calculate the match probabilities $pmatch_{d,e}$. After computing the match probabilities, the edges from RLP and FLP are matched using the min-cost bipartite matching algorithm (Hungarian algorithm [15]). We limit the maximum match weight $w_{max}$ to 70 in our deployments to eliminate highly noisy matches which are unlikely to be true light fixture events. For each match $m \in M^i$, the start and end times $m_s$ and $m_f$ are set to the timestamps of the rising and


### Experimental Setup

In this section, we discuss details of our underlying sensing and ground truth hardware, and their deployment in homes.

### Physical Sensor Components

Our WaterSense approach requires a water flow meter at the whole house water input line, and also motion sensors in each room that contains water fixtures. In our home deployments, we use a Shenitech Ultrasonic water flow meter [3] that uses the doppler effect to measure the velocity and resulting flow of water through the pipeline. The flow meter reports instantaneous water flow (in cubic meters per hour) at a frequency of 2Hz using the home’s Wi-Fi connection to transmit data. We expect that utility water flow meters being deployed in a large scale [7] in homes will have a similar setup. Figure ?? shows the installation of the flow meter in one of our home deployments. In addition to the flow meter, WaterSense requires at least one motion sensor in each room containing water fixtures. In our deployments, we use the X10 motion sensors [?] inside rooms to detect occupancy, as seen in figure ??.

The X10 motion sensors send a binary ON message whenever motion is seen with a minimal damping interval of 7 seconds between ON messages.

### Real World Deployments

To evaluate the LightSense system, we deploy it in four multi-resident homes for 10 days each. We use the TED power meter [3] along with the cheap Hamamatsu photo diodes [1] interfaced with the telosb mote [18], as shown in figure 7a and figure 7b. The power meter samples at 1Hz to match the capabilities of prevalent smart utility meters, while the light sensors sample only at 2Hz to have a minimal impact on the lifetime of the sensing node. To measure ground truth, we replace the existing switches in each home with wireless ZWave smart switches [?], and connect each lamp to a wireless ZWave smart plug [?] that measures the power consumption of the appliance plugged into it. Figure 5 shows details of the four homes in which LightSense and the ground truth sensing system were deployed. We see that each home has a high proportion of rooms with windows, which introduces higher noise levels due to natural light changes for our LightSense approach to handle. During our deployments, we also encountered multiple light fixture types such as incandescent bulbs, CFLs, and halogen lights, and both dimmable...

---

### Table: Details of the four homes chosen for deploying LightSense

<table>
<thead>
<tr>
<th>House #</th>
<th># of residents</th>
<th># of rooms</th>
<th># of light fixtures</th>
<th>House Type</th>
<th># of rooms with windows</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>3 floor House</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>3 bedroom student apartment</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>8</td>
<td>9</td>
<td>1 bedroom condo</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>9</td>
<td>14</td>
<td>3 floor house</td>
<td>7</td>
</tr>
</tbody>
</table>

---

### Figure 5. Details of the four homes chosen for deploying LightSense

---

### Figure 8. WaterSense uses a three tier inference algorithm to find 1) water flow events 2) clusters of flow volumes that often co-occur with certain motion sensors, and 3) different types of fixtures in the same room.
and non-dimmable switches with different light intensities and wattages as seen in figure 11.

**Energy Usage Ground Truth**

Before we can evaluate the energy feedback accuracy of LightSense, we first need to map the set of ground truth light fixtures $GF$ monitored using the ZWave devices, to the set of light fixtures $LF$ inferred by LightSense. To perform this mapping, we first compute a match recall metric $mr_{g,l}$ between each fixture $g \in GF$ and each fixture $l \in LF$, that denotes the proportion of ON-OFF events from $g$ that are detected by at least one ON-OFF event from $l$. An event $x$ from fixture $g$ is said to be detected by an event $y$ from fixture $l$ if both the ON and OFF times of events $x$ and $y$ are within $dT_m = 10$ seconds of each other. We map each ground truth fixture $g$ to an unmapped LightSense fixture $\text{map}(g) \in LF$ with the maximum match recall metric, if the maximum match recall metric $mr_{g,max}(g) > .5$, i.e. at least 50% of the ON-OFF events from fixture $g$ are detected from fixture $l$. At the end of this matching process, some light fixtures in $GF$ and $LF$ may be unmapped, resulting in false negatives and false positives, respectively.

To provide an easily understandable economic interpretation of light fixture energy consumption, we transform the raw energy in KWh for each light fixture to a projected energy cost over 5 years, assuming a cost of $0.15 per KWh, and extrapolating our 10 day sample period to a 5 year period. We use the ZWave ON and OFF times to compute the energy cost of each ground truth light fixture. We set the threshold parameters $dL$(in raw ADC units), and $dP$(in watts), to both be 10 in order to identify even low intensity light fixtures; we perform a sensitivity analysis on these parameters in section .

**EVALUATION**

To evaluate our WaterSense approach, we deploy our system in two homes for 7 days each. Both homes had multiple residents, multiple bathrooms, and a wide array of water fixtures and appliances. Details of the two deployments are summarized in Figure 10. To execute WaterSense in these homes, we deployed a single water flow sensor on the water mains and a motion sensor in each room. In Home 2, one of the motion sensors in a bathroom malfunctioned during our week long deployment, so we used a motion sensor in an adjacent bedroom with a partial view of the bathroom in our analysis. To evaluate the system, we deployed ZWave contact reed switches [?], as shown in Figure ?? to record the actual times that each fixture was used. We compute ground truth water consumption for each fixture by integrating total water flow into the house when each fixture was used. In the case of simultaneous toilet and sink events when we do not observe explicit sink edges, we ignore the short duration usage of the sink. These cases constitute a small fraction of
### RESULTS

**Light Fixtures**
(figures 6, 10, 11) from LightSense

**Water Fixtures**
(figure 6) from WaterSense - briefly discuss confusion matrix, classification accuracy

### ASSESSING ENERGY USAGE

Given the above experimental setup, we assess the accuracy with which LightSense infers the energy usage of individual light fixtures. Figure 6 shows the projected energy cost of each light fixture for the ground truth approach and LightSense, across all four houses over the 10 day deployment period. For each light fixture in figure 6, figure 11 shows the corresponding room locations, wattages, and light intensity changes observed at our light sensors.

We see from figure 6 that our LightSense approach accurately computes the energy costs of the top energy consuming light fixtures in each home; we observe an average accuracy of 91.1% in determining the energy cost of light fixtures consuming 90% of the home’s lighting energy. All of the false negatives and positives have very low energy consumption and do not significantly interfere with decision making about which light fixtures to optimize. Interestingly, House 1, which is a large 3 floor house, uses less lighting energy than House 2, which is a 3 bedroom student apartment; the reason is that House 1 has large wall-sized windows, which negate the need for lighting during most of the day. Our feedback can also be used to apportion the lighting bill based on energy usage; for example, the resident in Bedroom1 of House 2 might be motivated to reduce her high lighting energy given a personal cost incentive to do so.

For each light fixture, LightSense uses Tier III to provide users with the (1) nominal wattage, (2) room location based on the light sensor room location, and (3) the total energy cost. LightSense reports the nominal wattage of each light fixture shown in figure 11 within ±5W accuracy for all the top 90% energy consuming light fixtures in each home. Using these three inputs, residents can optimize high energy light fixtures by replacing incandescent bulbs with CFLs or LED bulbs, by replacing traditional lights with motion activated lights, by implementing a daylight harvesting system, or just by using the lights in a more conservative fashion. Each of these techniques requires a different input cost and
Figure 9. WaterSense accurately classifies water flow events for most of the monitored water fixtures across the two homes as seen in the fixture level confusion matrices. Confusion between fixtures of the same type in different rooms is common due to overlapping occupancy. Confusion between fixtures of different types, such as a sink and a flush, due to overlapping flow signatures, is less common. (B stands for bathroom, K for kitchen, S for sink, and F for flush)

(a) Home 1

<table>
<thead>
<tr>
<th>Fixtures</th>
<th>K S</th>
<th>B1 S</th>
<th>B1 F</th>
<th>B2 S</th>
<th>B2 F</th>
</tr>
</thead>
<tbody>
<tr>
<td>K S</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>B1 S</td>
<td>1</td>
<td>19</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>B1 F</td>
<td>0</td>
<td>1</td>
<td>49</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B2 S</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>B2 F</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>13</td>
</tr>
</tbody>
</table>

(b) Home 2

<table>
<thead>
<tr>
<th>Fixtures</th>
<th>K S</th>
<th>B1 S</th>
<th>B1 F</th>
<th>B2 S</th>
<th>B2 F</th>
</tr>
</thead>
<tbody>
<tr>
<td>K S</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>B1 S</td>
<td>7</td>
<td>78</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>B1 F</td>
<td>1</td>
<td>0</td>
<td>85</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>B2 S</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>B2 F</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>13</td>
</tr>
</tbody>
</table>

Figure 11. Nominal Wattages and Room Locations for the light fixtures in the four houses along with the light intensity increase or decrease observed at our light sensor as the fixtures are switched ON or OFF, respectively. The light fixture numbers here match the light fixture numbers in figure 6.

<table>
<thead>
<tr>
<th>Light #</th>
<th>Room</th>
<th>Wattage</th>
<th>Light change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MasterBed</td>
<td>135</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>Livingroom</td>
<td>35</td>
<td>193</td>
</tr>
<tr>
<td>3</td>
<td>Livingroom</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>Kitchen</td>
<td>90</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>MasterBath</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>MidBathroom</td>
<td>10</td>
<td>62</td>
</tr>
<tr>
<td>7</td>
<td>BottomBath</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>MasterBath</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>BottomBath</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>Kitchen</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>TopRoom</td>
<td>80</td>
<td>923</td>
</tr>
<tr>
<td>12</td>
<td>Kitchen</td>
<td>56</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Light #</th>
<th>Room</th>
<th>Wattage</th>
<th>Light change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MasterBed</td>
<td>135</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>Livingroom</td>
<td>35</td>
<td>193</td>
</tr>
<tr>
<td>3</td>
<td>Livingroom</td>
<td>40</td>
<td>29</td>
</tr>
<tr>
<td>4</td>
<td>Kitchen</td>
<td>90</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>MasterBath</td>
<td>20</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>MidBathroom</td>
<td>10</td>
<td>62</td>
</tr>
<tr>
<td>7</td>
<td>BottomBath</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>MasterBath</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>BottomBath</td>
<td>40</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>Kitchen</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>TopRoom</td>
<td>80</td>
<td>923</td>
</tr>
<tr>
<td>12</td>
<td>Kitchen</td>
<td>56</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Light #</th>
<th>Room</th>
<th>Wattage</th>
<th>Light change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FrontRoom</td>
<td>55</td>
<td>311</td>
</tr>
<tr>
<td>2</td>
<td>Basement</td>
<td>325</td>
<td>353</td>
</tr>
<tr>
<td>3</td>
<td>Kitchen</td>
<td>250</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Bathroom</td>
<td>395</td>
<td>1028</td>
</tr>
<tr>
<td>5</td>
<td>Kitchen</td>
<td>280</td>
<td>125</td>
</tr>
<tr>
<td>6</td>
<td>Livingroom</td>
<td>80</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>DiningRoom</td>
<td>200</td>
<td>110</td>
</tr>
<tr>
<td>8</td>
<td>Bedroom</td>
<td>95</td>
<td>69</td>
</tr>
<tr>
<td>9</td>
<td>Bathroom1</td>
<td>95</td>
<td>649</td>
</tr>
<tr>
<td>10</td>
<td>Nursery</td>
<td>55</td>
<td>129</td>
</tr>
<tr>
<td>11</td>
<td>Bedroom</td>
<td>60</td>
<td>32</td>
</tr>
<tr>
<td>12</td>
<td>Kitchen</td>
<td>70</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>Kitchen</td>
<td>30</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>Kitchen</td>
<td>110</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 12. Projected cost savings based on LightSense recommendations and an optimal system with ground truth data.

reduces energy by a different amount, so the energy feedback from LightSense is valuable to the end user in order to maximize net cost savings.

Given a limit $n$ on the number of light fixtures that users can optimize, we compute the maximum projected energy cost savings achievable over 5 years (the typical lifetime of a CFL), assuming we optimize the top $n$ light fixtures based on the energy cost ordering computed by LightSense; we assume 75% energy reduction for each light fixture after optimization, which is similar to the energy savings obtained on switching from incandescent lights to CFLs. Figure 12 shows the maximum projected cost savings of LightSense for each house as $n$ is increased along the x-axis. We also show the cost savings for an optimal system that uses the ground truth energy costs. We see that LightSense closely tracks the optimal recommendation system for optimizing light fixtures across all houses. Also, in each house, the cost savings are quite different and become very small after 6-7 fixture are replaced: this suggests that users should carefully weight the cost vs. benefit of optimizing light fixtures, before deciding on a suitable $n$ for each house.

We see the value of LightSense by combining the results from figures 6, 11 and 12; many of the low energy, low return on investment light fixtures in each house are in similar rooms as the top energy consumers, and in some cases even have a higher or similar nominal wattage compared to the top energy consumers. Thus, optimizing light fixtures based on their actual energy usage is more profitable compared to using only their nominal wattages or room locations.

**ANALYSIS**

In this section, we provide detailed insights on the high accuracy achieved by LightSense in inferring the energy cost of individual light fixtures in the home. Firstly, we show that both the data fusion and Bayesian matching steps contribute equally to aggressively filter out false positives. Secondly, we show that our Bayesian Clustering approach is needed to accurately assign power measurements to ON-OFF events. Finally, we perform a sensitivity analysis of LightSense to the power and light thresholds $dP$ and $dL$.

**Impact of Data Fusion and Matching**

To understand the impact of Data Fusion and Bayesian Matching on improving LightSense accuracy, we implement five distinct inference approaches with increasing sophistication, namely: Light Edge Detection (LE), Power Edge Detection (PE), Light Edge Matching (LM), Power Edge Matching (PM), and finally Light Power Data Fusion (LP), that eliminates temporally isolated light or power edges from by applying our temporal intersection step. We also consider the LightSense approach, which effectively uses all the inference algorithms and sensor sources used in the five approaches above. To measure the accuracy of each inference approach above, we use the common evaluation metrics of edge detection recall and precision for light fixture ON and OFF edges. Figure 13 shows these metrics for all approaches; we fix the light and power thresholds $dL = dP = 10$, to be low enough to detect the low intensity fixtures seen in figure 11. We see that approaches LM and LP improve precision significantly compared to the simpler approaches LE and PE in each house, by eliminating unmatched or temporally isolated false positive edges. Interestingly, approach PM only has a marginal improvement over approach PE due to the large number of false positive power edges. Finally, we see that in each house, LightSense improves upon both LM and LP, by combining both noise elimination techniques to aggressively filter out false positives from the light and power meter data streams.

**Impact of Bayesian Clustering**

To understand the impact of using Bayesian Clustering in accurately assigning power costs to ON-OFF events, we consider two different power assignment approaches: (1) Power Only assignment uses only the power meter data to assign power edges to ON-OFF events, i.e. by only using $vprob_{x,y}^{LP}$ in equation 2, (2) Bayesian Clustering assignment also uses the joint probability of observing the light and power meter data as shown in equation 2 using our $lpbayes$ Bayesnet. For each approach, we consider the total energy cost error, normalized as a percentage by the total energy cost of all fixtures in the home. In figure 14, we observe that leveraging the long term temporal correlation between the light and power meter data reduces the total normalized energy cost error from 40-50% to less than 10% in multiple homes. We see almost no impact of our Bayesian Clustering approach in House 2, because this small apartment has a little noise on the power lines due to the lack of HVAC, television, or other
Complex appliances.

**Parameter Sensitivity Analysis**

The main variable parameters in our approach that affect classification accuracy are the light and power edge thresholds $dL$ (in ADC units) and $dP$ (in Watts). Figure 15 shows the recall-precision graph for all the five additional inference approaches from section and LightSense as $dL$ and $dP$ are simultaneously decreased from 500 to 10 in steps of 50 each; we set the the minimum values of the threshold parameters to be 10 instead of 0, and make the last step size 40. We plot the edge detection recall and precision for each parameter setting averaged across the four homes.

As the threshold parameters are decreased, we observe that (1) edge recall for all approaches increases as low intensity events greater than the decreasing threshold are detected, and (2) precision decreases as more false positives are introduced from low intensity edges. While approaches LM and LP achieve only about a 5-10% worse F1 score compared to LightSense at a threshold setting of 50, they achieve much lower precision when detecting low intensity events at less than 50 ADC light units or less than 50 W. However, identifying low intensity events is important, since we see from figure 11 that several light fixtures in the top 5 or 6 consumers in each home have either a wattage less than or close to 50W, or a light intensity less than 50 ADC units. Thus, our LightSense approach combines both noise elimination techniques, matching and data fusion, to accurately identify even important low intensity light fixture events.

**CONCLUSIONS**

In this work, we presented the LightSense system to infer the fine-grained energy usage of light fixtures in homes. Through 10-day deployments in four natural home environments, we show that LightSense effectively combines two noisy sensor sources, namely light sensors with 25% precision and whole house power consumption data with 15% precision, to infer the energy costs of individual light fixtures with 91.1% accuracy.

In this paper, we present an unsupervised approach to infer the fine-grained fixture level breakdown of water consumption in homes by effectively combining cheap occupancy sensors and whole water flow meter data. Our approach shows significant promise in an early evaluation carried out in two homes over 7 days each, and is able to accurately infer both the time of use and water consumption of individual fixtures. In the future, we expect to build upon the existing approach by performing an extensive evaluation of our system that includes ground truth sensing on fixtures such as the shower, washing machine, and dishwasher to address potential challenges posed due to simultaneous fixture use from these high consumption fixtures. There are also several interesting alternatives to explore in our multi-tier inference architecture to improve accuracy in a complete evaluation, including improvements to fixture identification in Tier III, improvements to the temporal distance features used in the bayesnet, and additions to the algorithm to handle simultaneous or compound flow events. With these additions, we expect our WaterSense approach to be a viable approach to provide fine grained water consumption information and recommendations to end users to help conserve water.

**REFERENCES**

3. Ted the energy detective.


