SmartSim: A Device-Accurate Smart Home Simulator for Energy Analytics

Dong Chen, David Irwin, and Prashant Shenoy

University of Massachusetts Amherst

November 9th, 2016

Ubiquitous Smart Metering and Energy Sensing

- Utilities rapidly deploying smart meters
  - \>50 million deployed in U.S.
  - Record energy usage data every 5-60 minutes
  - Next-generation meters use higher resolutions (1Hz)

- In-home energy sensing also becoming common
  - Included in many IoT devices, e.g., Belkin WeMo
  - Included as part of solar panel systems
  - Typically 1Hz

- Sensors now producing a **massive** amount of energy data

Exploiting Energy Data using Analytics

- Companies actively working to develop **energy data analytics**
  - How much energy are individual devices using?
  - Are any devices malfunctioning or being used inefficiently?
  - What brands/models of appliances do occupants own?
  - What are a home's occupancy patterns?
  - How often do occupants eat-in versus go out to eat?
  - **Preventing** such analytics is also an active research area

- Companies can apply energy analytics to utility-scale data to identify **energy-inefficiencies** or **profile customer behavior** (for marketing)

Exploiting Energy Data using Analytics

- Companies activel… …ergy data analytics
  - How much energy … ?
  - Are any devices ma… …inefficiently?
  - What brands/model… …nts own?
  - What are a home's
  - How often do occup… …o eat?
  - **Preventing** such a… …search area

- Companies can ap… …utility-scale data to identify **energy-ine…** …**ustomer behavior** (for marketing)

A sample of projects that we are working on and problems that we are trying to solve:

- **Appliance Energy Signatures.** Its the pattern recognition algorithms applied to the energy consumption signatures of various appliances to extract them from the whole house energy profile from Smart Meters.
- **Identify User Behavior.** Parameterize each appliance use based on user lifestyle and consumption and help them identify where to target energy reduction.
- **Identify Appliance Brands.** Use energy data to predict whether the user has GE or Maytag refrigerator. Very cool! Imagine the value of that information for Whirlpool to target this house for selling their appliance.
- **Handling Big Data.** With live data from millions of homes, imagine how interesting it would be to be able to predict the clustering of appliances based on model, year, geography, efficiency and user behavior.

bidgely plotwa… Save your ene… P☺WER belkin

Example Energy Data Analytics

- **Non-Intrusive Load Monitoring (NILM)**
  - Also called “load disaggregation”
  - Analyze smart meter data from an entire building….
  - ….infer the energy usage (over some interval) of the building’s individual devices
  - Many NILM algorithms proposed in prior work

- Many other variants of similar Non-Intrusive Analytics
  - Non-Intrusive Occupancy Monitoring (NIOM) – infer occupancy from smart meter data
  - Non-Intrusive Thermal Disaggregation – separate energy usage of temperature-dependent loads from temperature-independent loads

Problem

- Properly evaluating energy data analytics requires **ground truth energy data** from **many buildings**
  - For NILM, need building data and energy data from **every** device
  - Need data from many buildings with different characteristics
  - Collecting data is **time-consuming, expensive, and logistically hard**
    - Public datasets available, but none has ground truth data for many buildings
  - The **best** evaluations in prior work include only a few buildings
    - Many include only a single building, and have no evaluation

- Even given a large-scale public dataset, researchers **cannot**…
  - …rigorously vary building and device characteristics….
  - …to determine characteristics that affect analytics accuracy.

Solution: SmartSim

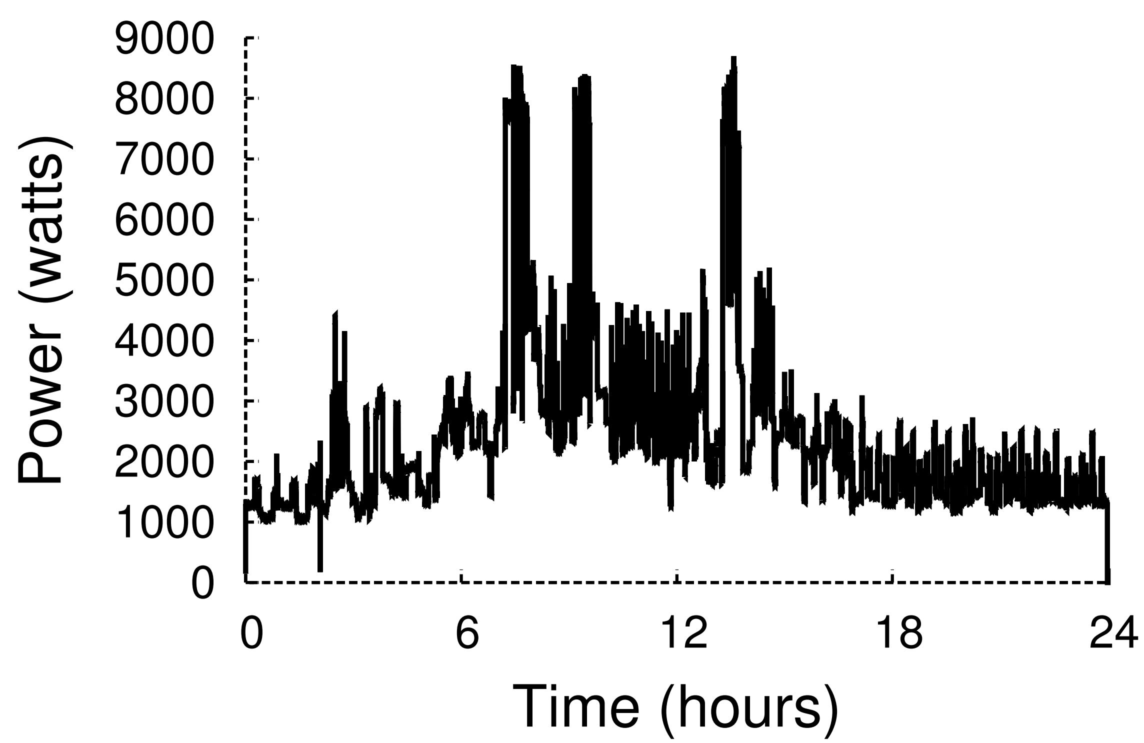
- SmartSim is a **device-accurate home energy trace generator**
  - Device-accurate -> Generates ground truth energy trace for each device
  - Home energy trace -> Sums device energy traces to generate home trace
  - **Goal:** Generate realistic home energy traces similar in complexity to real homes, but allow users to control home characteristics
    - Enable evaluation of energy analytics on synthetic data
    - Able to generate data for many homes and rigorously vary home characteristics

- Leverages **two primary components** in generating traces
  - Device energy model – describes how a device uses energy when active
  - Device usage model – describes when a device is active over time

Challenge

- Real energy data is highly complex, especially at high resolution
  - Modern devices have continuous/stochastic changes in energy usage
    - Often not clear "steps" in usage
  - Not trivial to generate synthetic data at similar complexity
  - Focus on 1Hz data, since this is the resolution most end-user sensors support (and are a focus of analytics)
    - Utility meters may support similar high resolutions in the future

Typical home's Energy usage is highly complex

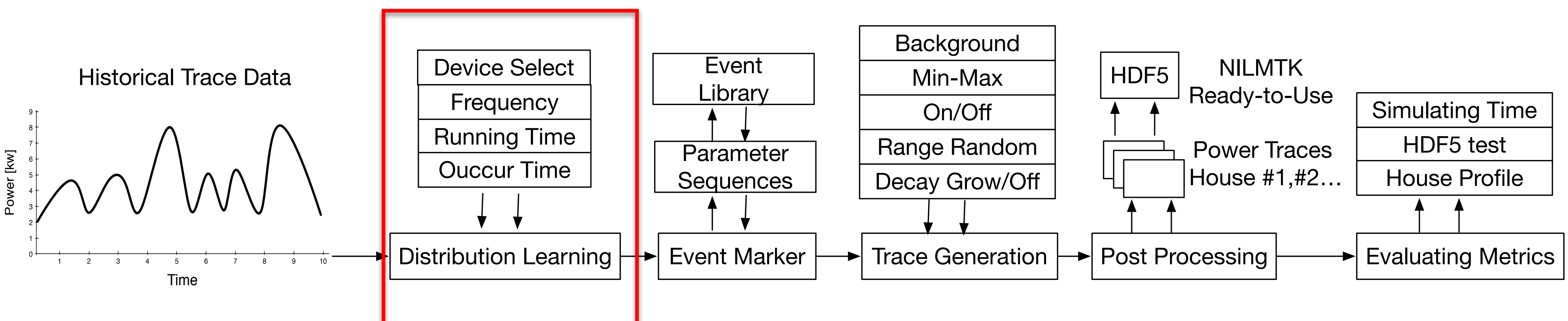


Outline

- ~~Motivation~~
- SmartSim Workflow Overview
- Implementation
- Evaluation
- Related Work
- Conclusion

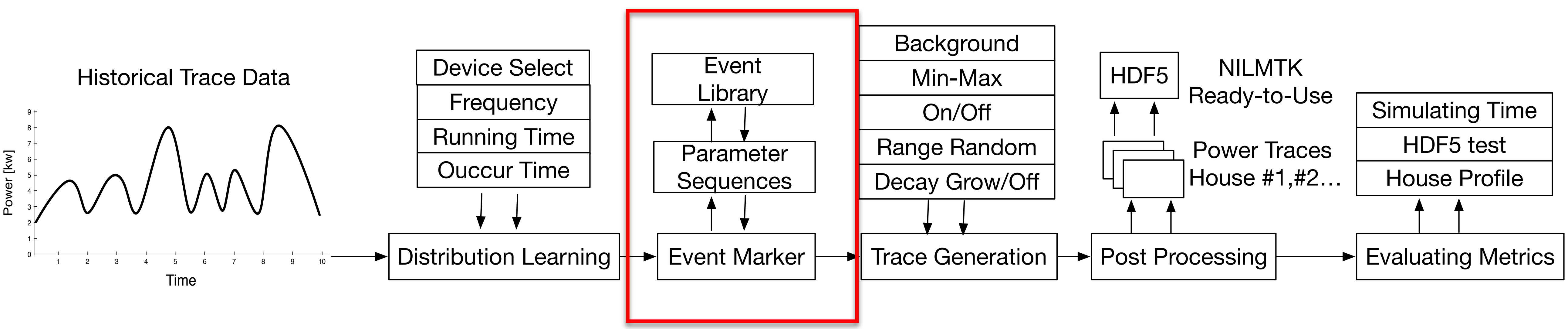
SmartSim Workflow Overview

- **1.** Select set of devices to include in energy trace



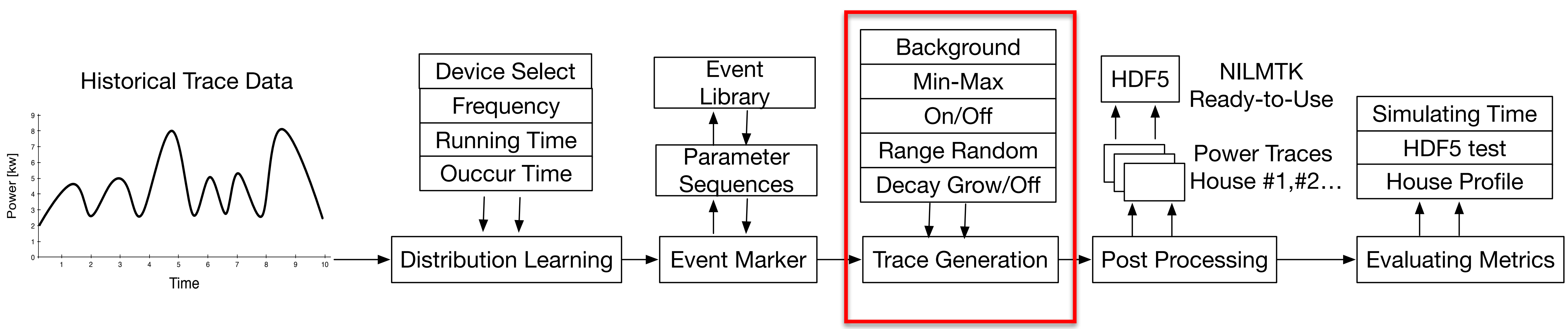
SmartSim Workflow Overview

- **1.** Select set of devices to include in energy trace
- **2.** Select usage model for each device
  - May select from well-known distributions, e.g., gaussian, etc.
  - May also empirically learn distribution from historical data (below)



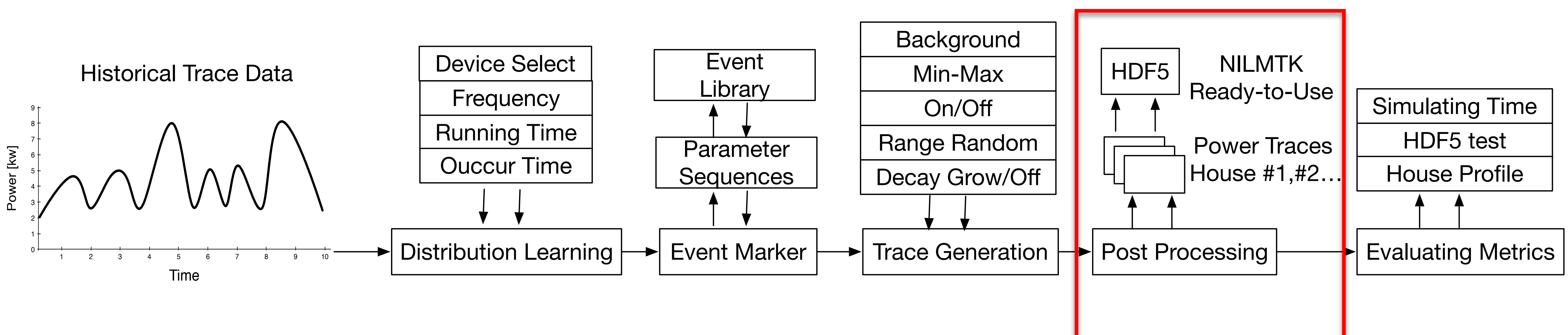
SmartSim Workflow Overview

- **1.** Select set of devices to include in energy trace
- **2.** Select usage model for each device
  - May select from well-known distributions, e.g., gaussian etc.
  - May also empirically learn distribution from historical data (below)
- **3.** Select energy model for each device from library



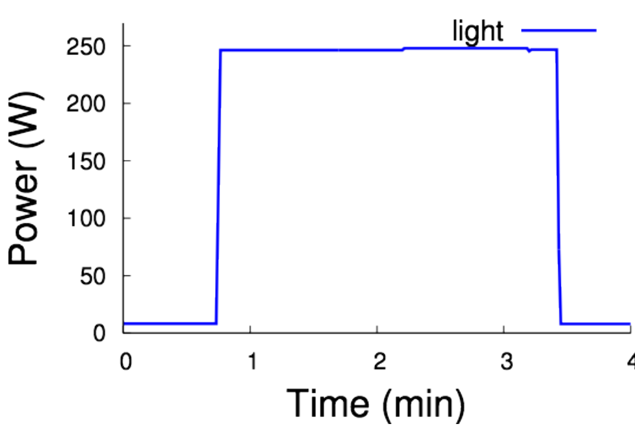
SmartSim Workflow Overview

- **1.** Select set of devices to include in energy trace
- **2.** Select usage model for each device
  - May select from well-known distributions, e.g., gaussian,etc.
  - May also empirically learn distribution from historical data (below)
- **3.** Select energy model for each device from library
- **4.** Generate random device energy traces based on usage distributions and energy model
  - Output in HDF5 format – ready-to-use by NILM-TK (open-source library)

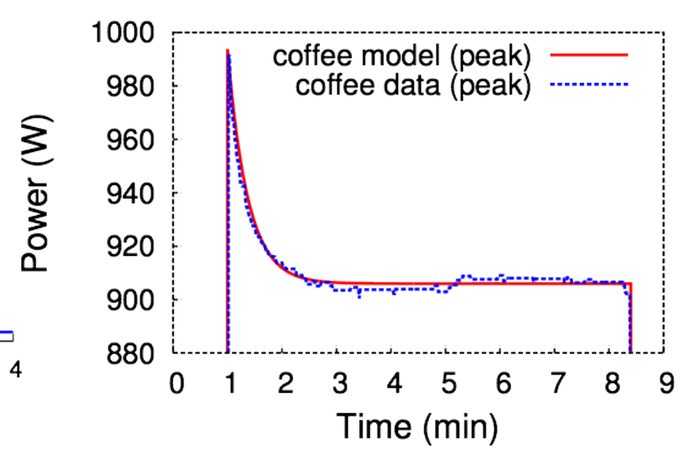


Device Energy Models

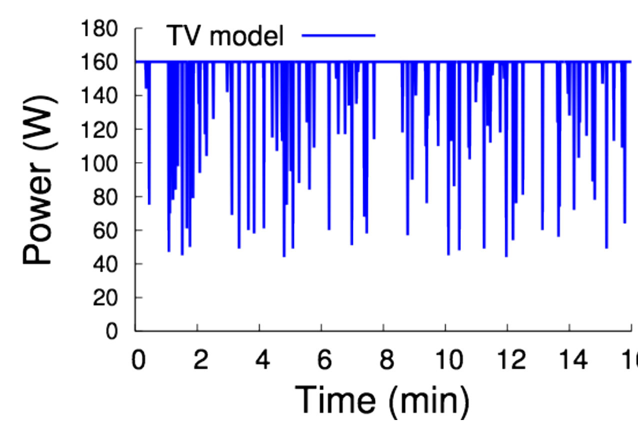
- Leverage modeling methodology from prior work [JSAC 2014]
  - Models device energy usage using four basic models
    - Models are device-specific: parameterized based on real data
  - Derived from fundamental AC characteristics:
    - Resistive (heating element), inductive (AC motor), or non-linear (electronics)
    - All devices are one or a combination of these basic elements
  - SmartSim includes a library of models to choose from



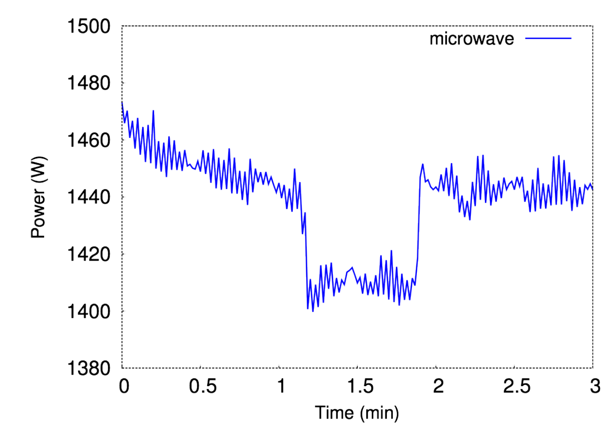
On-off



On-Off Growth/Decay



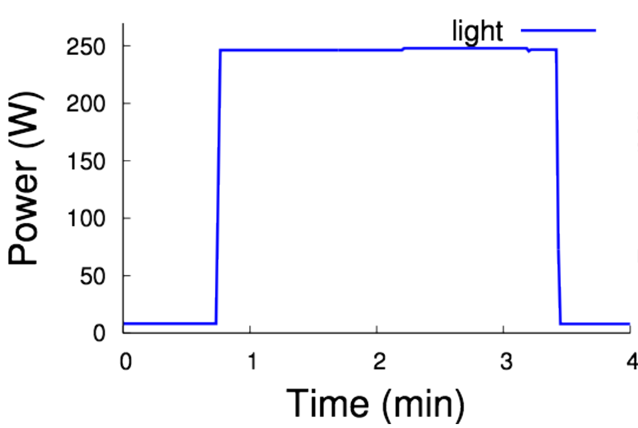
Stable Min-Max



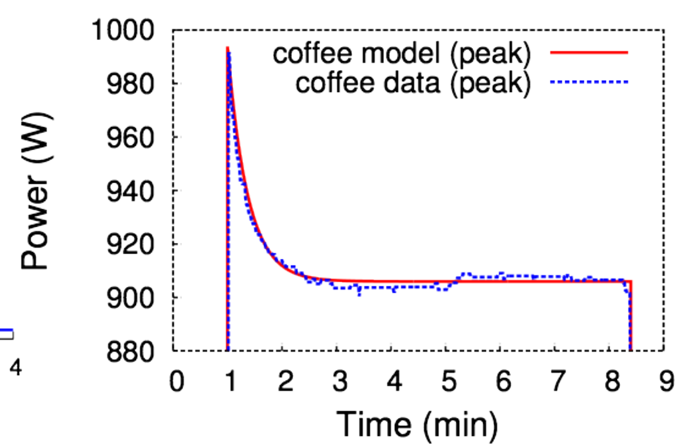
Random Range

Increasing Realism

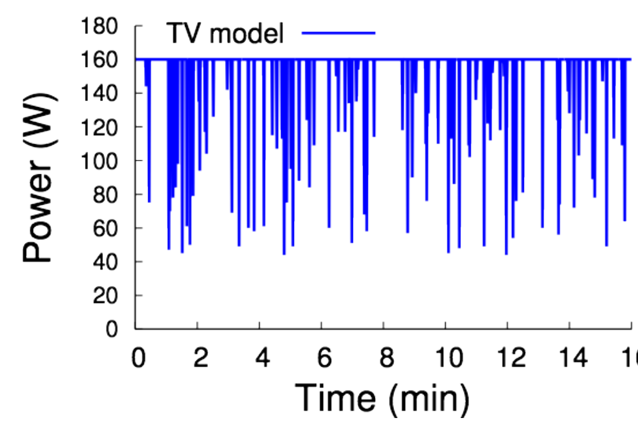
- Modeling Errors, Noise, and Dropouts
  - SmartSim includes support for adding **sensor error** and **data dropouts**
  - Most sensors have an error that is 1-2% of the total load
  - Data dropouts in real data occur frequently and affect analytics accuracy
  - Include these features to increase realism of the data



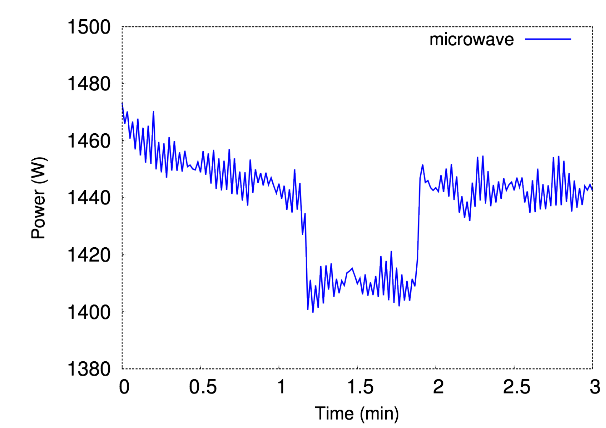
On-off



On-Off Growth/Decay



Stable Min-Max



Random Range

Device Usage Models

- Captures the frequency, timing, and duration of device activity
  - Dictates when to insert device energy model into trace
  - Separate **background loads** from **interactive loads**
  - Background load operation captured in device energy model
    - Devices operate cyclically every interval, e.g., refrigerator, air conditioner, etc.
  - Interactive load operation captured by device usage model
    - Define probability distribution for time-of-use, frequency-of-use, duration-of-use
    - 1) Use frequency-of-use distribution to determine frequency of use each day
    - 2) Use time-of-use distribution to determine when each use occurs
    - 3) Use duration-of-use distribution to determine each activity duration
  - For each device in our library, we automatically derive these distributions from device's historical data (or users may select standard distribution)

SmartSim Implementation

- Implemented in python
  - Current library includes 25 devices with device energy and usage models
  - Plan to expand the library using data from more devices
  - Includes tools to derive energy and usage models from existing data

- Output compatible with NILM-TK toolkit
  - NILM-TK -> open-source library of common NILM algorithms
    - Combinatorial Optimization (CO)
    - Factorial Hidden Markov Models (FHMM)
  - Uses a common dataset format (HDF5)
  - Many publicly-available energy datasets already available
    - Enables comparison between SmartSim data and real data

Evaluation

- Include **qualitative** and **quantitative** evaluation
  - Qualitative -> SmartSim generates traces similar to real data
  - Quantitative -> NILM algorithms (CO and FHMM) perform similarly on SmartSim-generated data and real data
    - Compare with Public Datasets: REDD and Smart*
- NILM Accuracy Metrics
  - F1 = 2 * (precision * recall) / (precision + recall)
  - Normalized Error in Assigned Power (NEP)

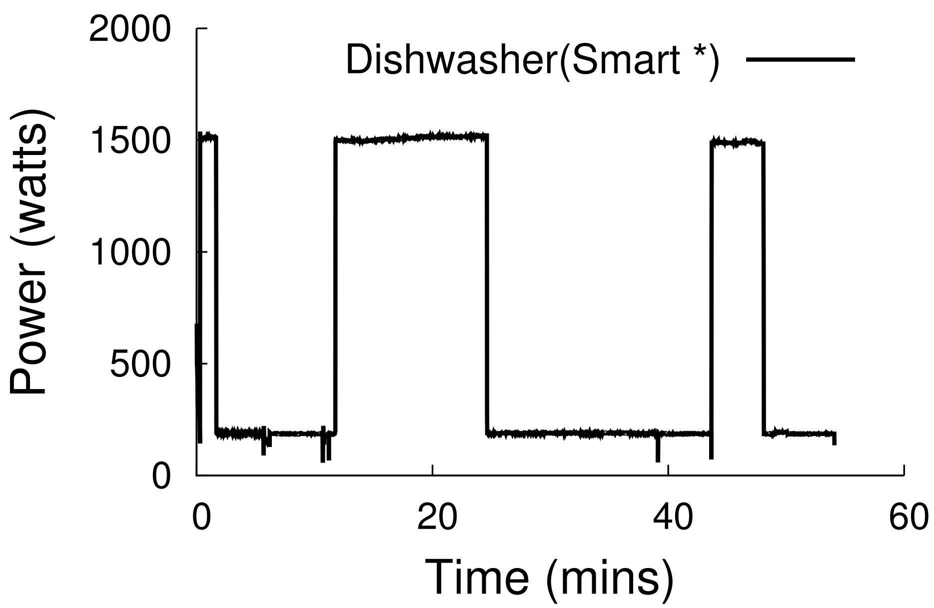
$$\delta = \frac{\sum_{t=1}^{T} |\tilde{p}_i(t) - p_i(t)|}{\sum_{t=1}^{T} \tilde{p}_i(t)}$$

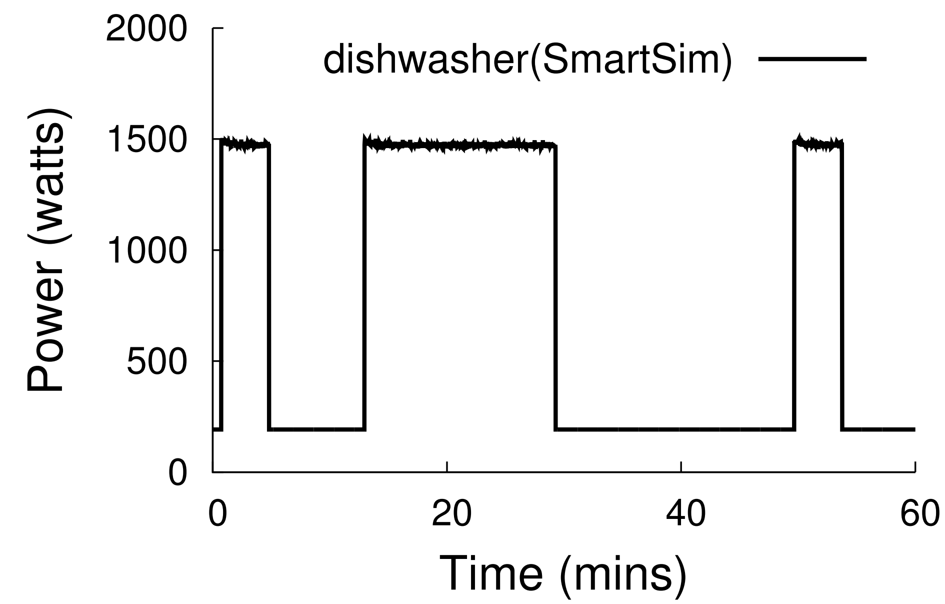
  - Matthews Correlation Coefficient

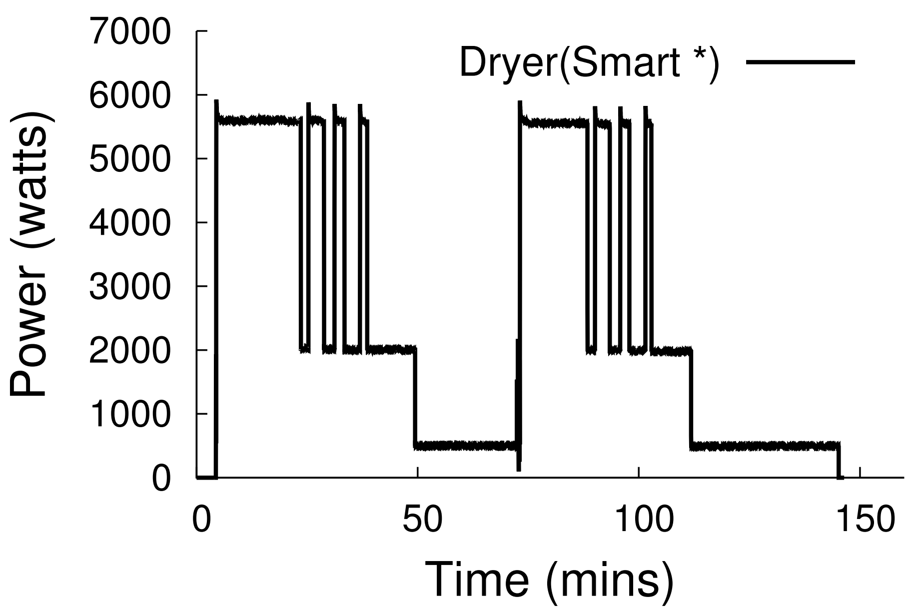
$$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

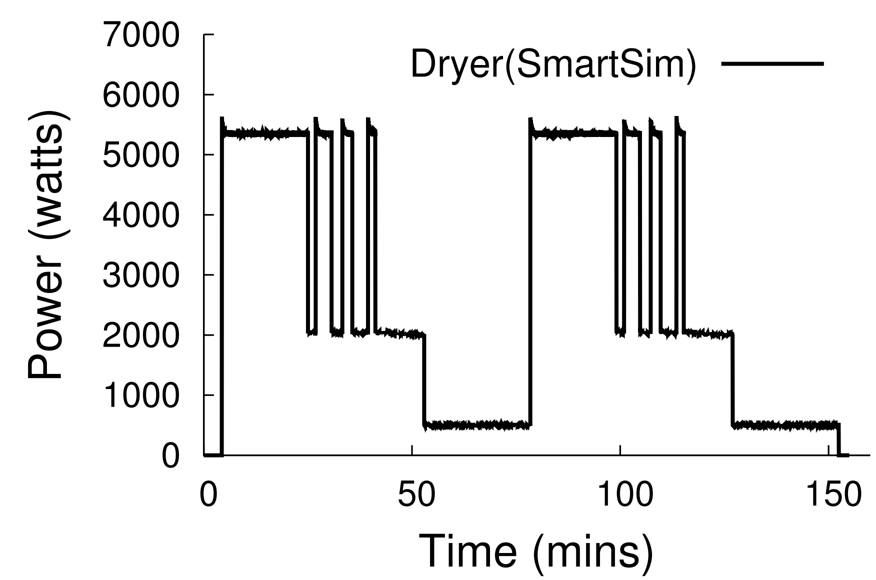
Qualitative Evaluation

- **Examples – more available in paper**



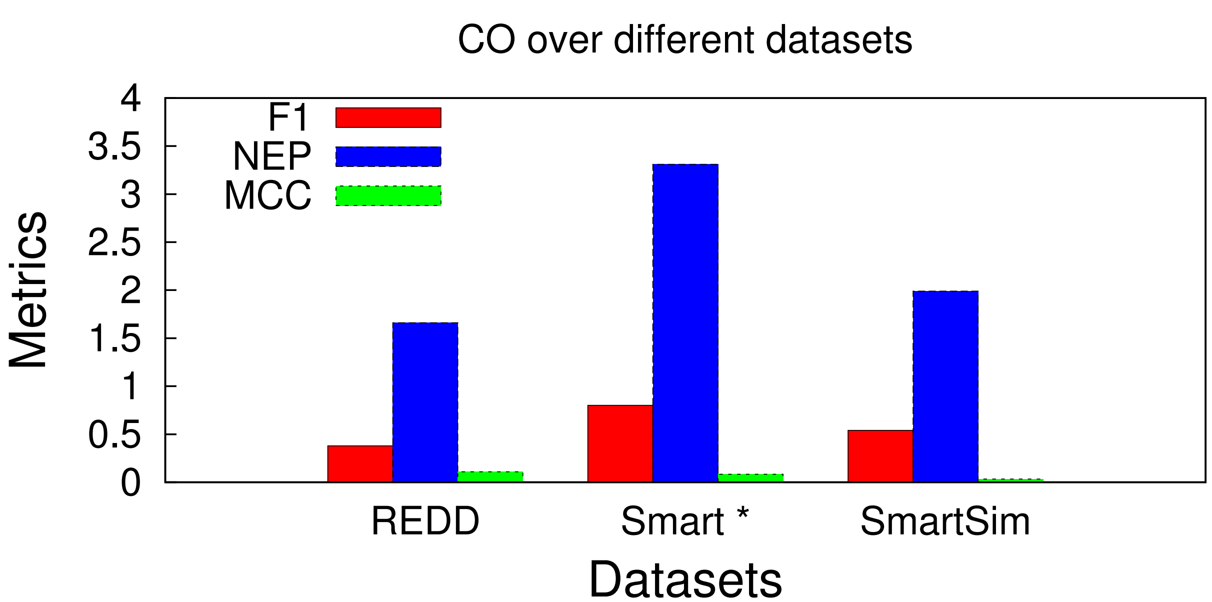


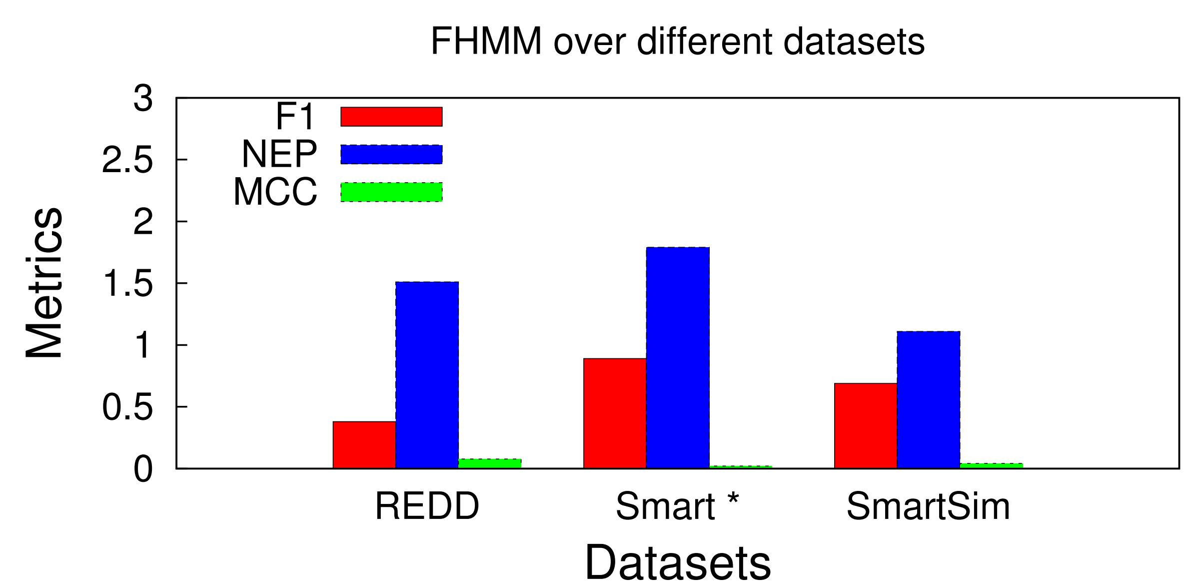




Quantitative Evaluation

- **Smart* and SmartSim+Noise have comparable accuracy and performance for CO and FHMM algorithms**
  - Performance: similar training and test times
    - ~10 seconds training for CO, ~900 seconds for FHMM
    - ~2-3 seconds testing for CO and FHMM
  - Accuracy: similar MCC and NEP





Related Work

- Significant prior work on NILM and other energy analytics
  - Most do little-to-no evaluation
  - Best evaluations are small-scale on a few homes
  - Recent survey demonstrates widely disparate results on different homes using same/similar algorithms (Armel et al.)
  - Must test on more/better/controlled data
- Also related to prior work on modeling energy usage
  - SmartSim is a **tool** that applies these models to generate device-accurate energy data traces
  - Prior work on modeling focuses on modeling methodology and not application

Conclusion

- SmartSim fills a gap in evaluative techniques for energy analytics
  - Moves beyond evaluating techniques on a random sampling of homes with no ability to control characteristics

- Enables researchers to…
  - …generate device-accurate 1Hz energy data for home….
  - …choose the set of devices (and usage patterns) to control home characteristics…
  - …and apply to existing NILM algorithms by producing data in a common format (HDF5) compatible with existing open-source implementations.

Questions

Questions?