SUSTAINABILITY LAB

CLIMATE CHANGE AI · VIRTUAL SUMMER SCHOOL 2026

AI for Buildings

Machine learning across the building lifecycle: sense, model, decide, verify

Nipun Batra · Sustainability Lab, IIT Gandhinagar

Module: AI for Buildings & Cities · Part 1

PART 1

Why buildings?

A third of global energy — and it's already metered

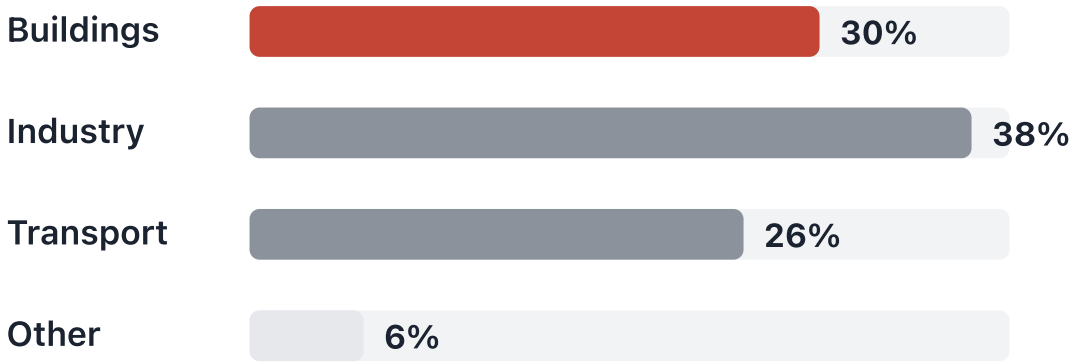
SENSE MODEL DECIDE VERIFY

AI for Buildings · Climate Change AI Summer School 2026

THE STAKES

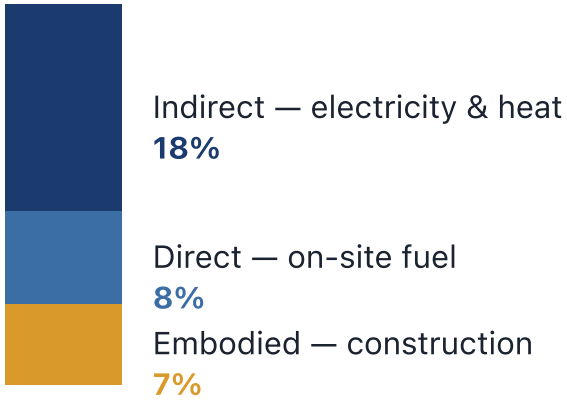
Buildings use about a third of global energy

SHARE OF GLOBAL FINAL ENERGY



Buildings = 34% once construction materials are included

BUILDINGS' ENERGY-RELATED CO_2



≈ one-third of energy-related CO_2

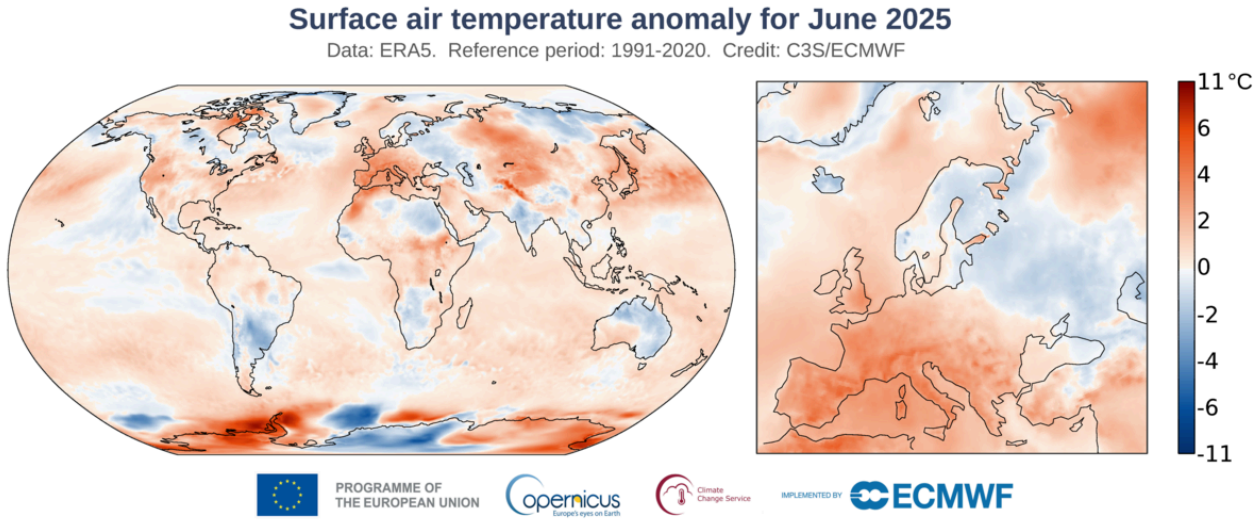
IEA, data year 2022

TAKEAWAY

Buildings are about **30% of global final energy** and roughly **one-third of energy-related CO_2** once construction is included.[1]

WHY NOW

Cooling is the fastest-growing use of energy in buildings



June 2025 — western Europe's warmest June on record. Credit: C3S/ECMWF[2]

1.6→5.6B

air-conditioners by 2050 — the stock grows ~3.5×[3]

+14%

Spain's daily demand; prices rose 108% in France and 175% in Germany during the 2025 heatwave[4]

~20%

of EU homes have AC (vs ~90% in the US) — demand will climb fast

POLL

What's the single biggest electricity user **in your home** — AC/heating, water heater, fridge, or something else?

ANSWERING THE QUESTION

It depends on climate — and on what you count

19%

US · air conditioning share of household **electricity**[5]

62.5%

EU · space heating share of household **final energy**[6]

~40%

India · fans, coolers, and AC share of surveyed household **electricity**[7]

40–60%

Singapore · AC share of total **building energy**[8]

TAKEAWAY

Read the denominator before the percentage. Climate, fuels, appliances, and construction change the dominant load — and make geographic transfer a first-class ML problem.

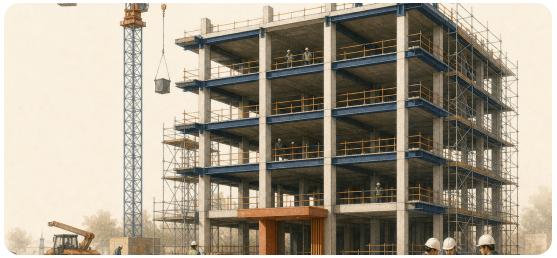
THE MISSING MAP

Four phases of a building — four different decisions



Design · choose

Massing, envelope, materials, systems



Build · check

Progress, quality, metadata, sequences



Operate · adapt

Weather, people, prices, faults



Retrofit · verify

Diagnose, prioritize, measure change

TAKEAWAY

Before operation, people and performance are assumptions. During operation they become evidence; retrofit should use that evidence to change the building.

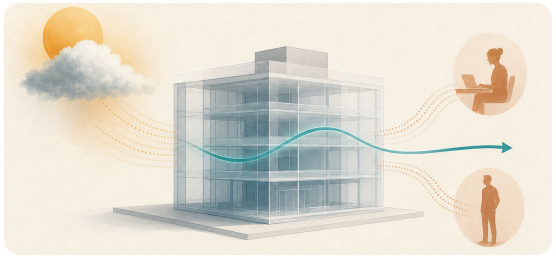
THE PLAN

The workflow: sense ⇢ model ⇢ decide ⇢ verify



Sense

What is happening?



Model

What happens next?



Decide

Which action meets the constraints?



Verify

Did it work — and for whom?

TAKEAWAY

We will run this loop through four applications. One question throughout: **can the result survive the next building, the real grid, and the people inside?**

APPLICATION 1 · STAR: SENSE

Energy disaggregation

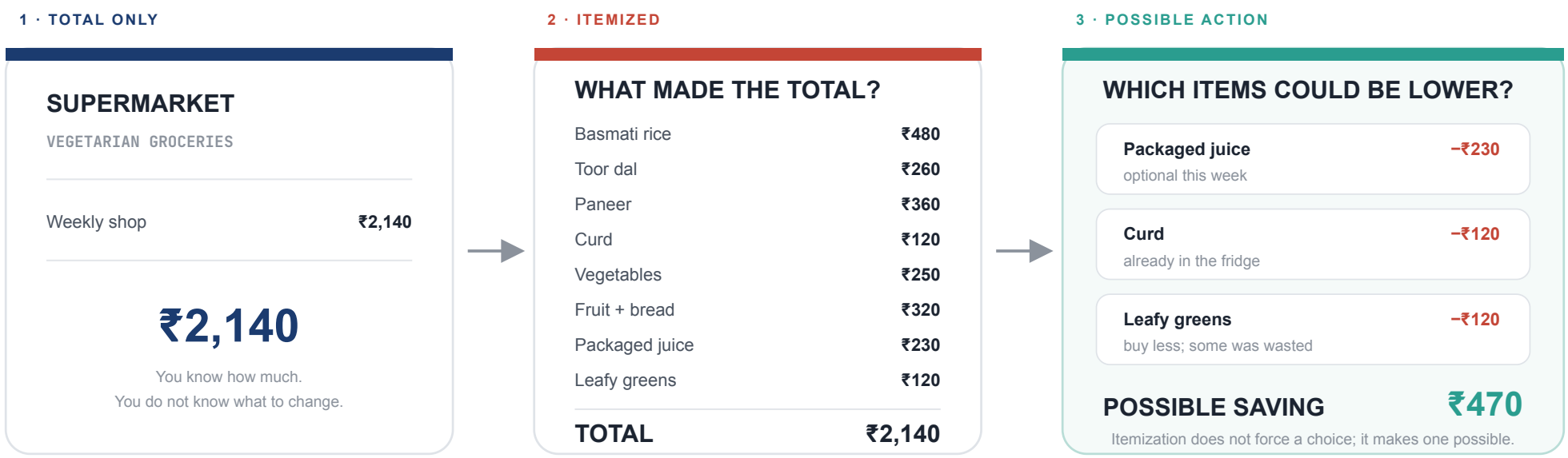
One cheap meter in — itemized, actionable feedback out

SENSE MODEL DECIDE VERIFY

AI for Buildings · Climate Change AI Summer School 2026

THE INTUITION BEFORE THE ALGORITHM

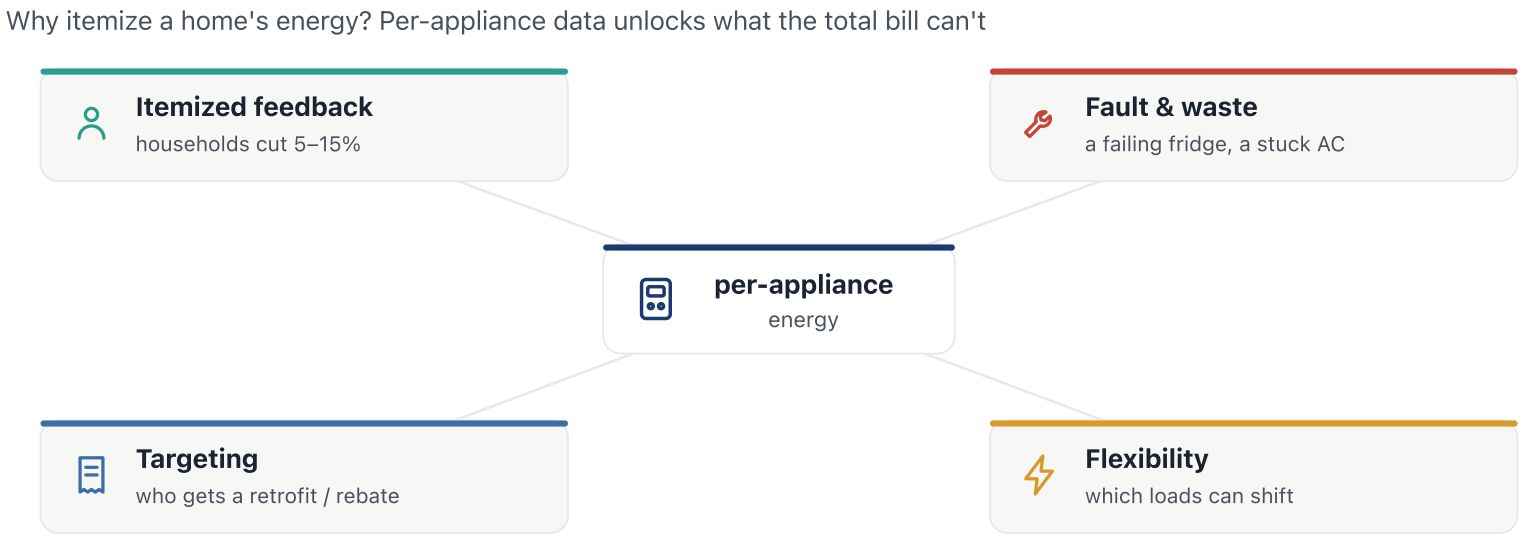
Your supermarket bill explains why itemization matters



TAKEAWAY

A total tells you how much; an itemized bill tells you where a decision is possible. NILM asks whether one whole-home electricity signal can become the equivalent of an itemized energy bill.

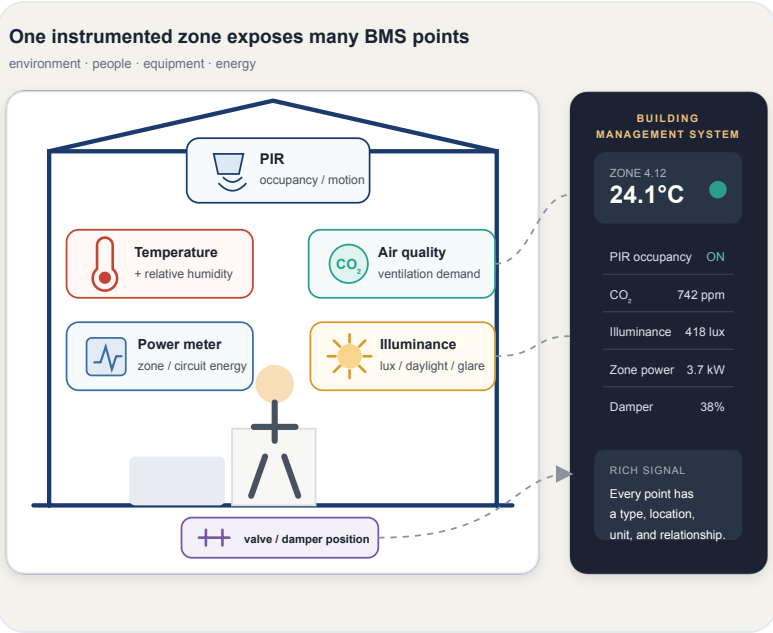
What per-appliance data unlocks



TAKEAWAY

Itemized feedback nudges households to cut **5–15%**, and the same breakdown powers fault detection, retrofit targeting, and demand-response eligibility. This is why so much of buildings-ML chases the appliance-level signal.[9, 10]

What you can measure — and what it costs

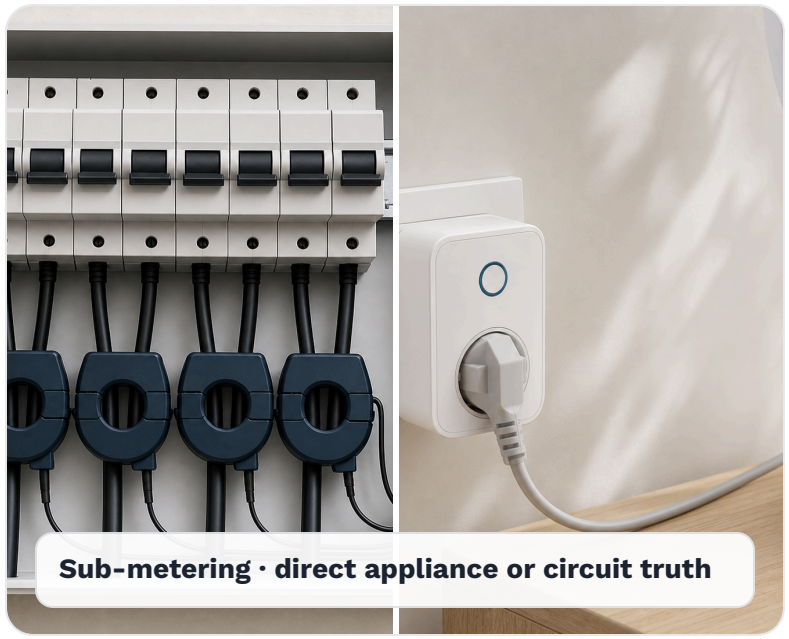


Utility bill	monthly · whole building	billing, benchmarking	nearly free
Smart meter	seconds–minutes · aggregate	NILM, forecasting, DR	wide reach
Sub-metering	circuit / appliance	ground truth, audits	higher cost
Instrumented building	power + temp + CO_2 + occupancy	control, comfort, fault diagnosis	**highest effort**

TAKEAWAY

Start at the bottom: rich sensing can answer almost anything, but it is expensive to install, label, and maintain across many buildings.

What you can measure — and what it costs



Sub-metering · direct appliance or circuit truth

Utility bill	monthly · whole building	billing, benchmarking	nearly free
Smart meter	seconds–minutes · aggregate	NILM, forecasting, DR	wide reach
Sub-metering	circuit / appliance	ground truth, audits	higher cost
Instrumented building	power + temp + CO_2 + occupancy	control, comfort, fault diagnosis	highest effort

TAKEAWAY

Sub-meters give the **answer directly**. They are ideal ground truth — and exactly what becomes difficult to afford and maintain at scale.

What you can measure — and what it costs



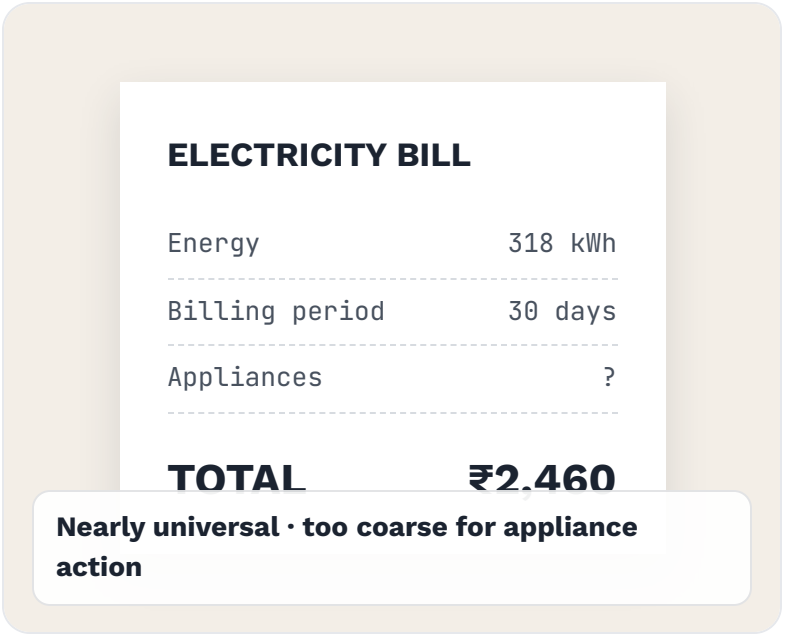
One aggregate waveform · millions of buildings

Utility bill	monthly · whole building	billing, benchmarking	nearly free
Smart meter	seconds–minutes · aggregate	NILM, forecasting, DR	**wide reach**
Sub-metering	circuit / appliance	ground truth, audits	higher cost
Instrumented building	power + temp + CO_2 + occupancy	control, comfort, fault diagnosis	highest effort

TAKEAWAY

A smart meter is the interesting middle: cheap and widespread, but it measures only the **sum**. NILM tries to infer the richer appliance signal from it.

What you can measure — and what it costs



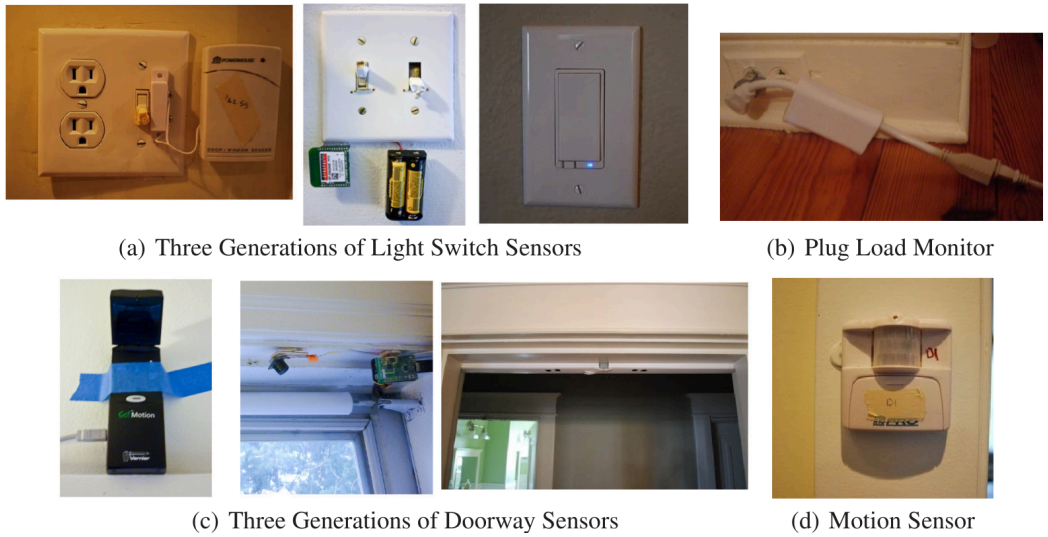
Utility bill	monthly · whole building	billing, benchmarking	**nearly free**
Smart meter	seconds–minutes · aggregate	NILM, forecasting, DR	**wide reach**
Sub-metering	circuit / appliance	ground truth, audits	**higher cost**
Instrumented building	power + temp + CO_2 + occupancy	control, comfort, fault diagnosis	**highest effort**

TAKEAWAY

The field exists because reach and resolution move in opposite directions. **Buildings-ML often infers an expensive signal from a cheaper one.**

REAL DEPLOYMENTS

What real homes do to your sensors



(a) Three Generations of Light Switch Sensors

(b) Plug Load Monitor

(c) Three Generations of Doorway Sensors

(d) Motion Sensor

Home sensors — Hnat et al., **Hitchhiker's Guide**, SenSys 2011[11]

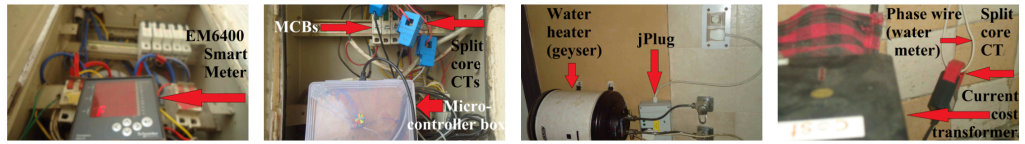
LESSONS FROM REAL DEPLOYMENTS

- **Maintenance outweighs installation** — sensors fail, labels rot, occupants change[11]
- **Effort grows super-linearly** with homes, nodes, and months in the field[11]
- **Assumptions travel badly** — what breaks when you deploy outside the lab's grid? (next slide)

TAKEAWAY

Deployment papers are as useful as algorithm papers, and far rarer — every model inherits these messes.

Delhi taught us: measure voltage, not just current



73 days, 33 sensors, one Delhi home — smart meter, per-circuit CTs, appliance plugs[12]

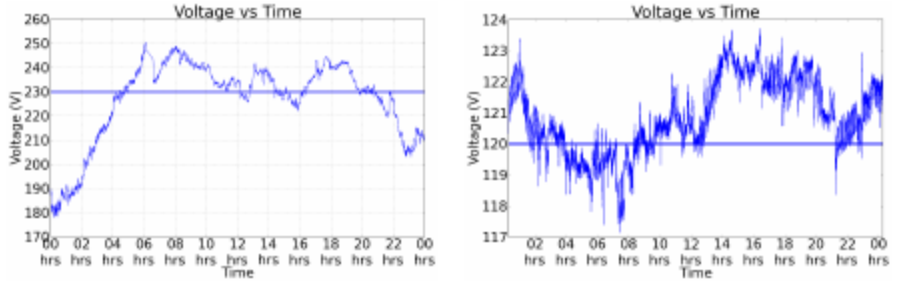


Fig. 5 · **Batra et al., BuildSys 2013** · doi:10.1145/2528282.2528293

TAKEAWAY

The grid itself moves: Delhi supply swung **~180–250 V** in a day (US comparison: pinned near 120 V). CT-based meters that record only current **assume** fixed voltage — here that assumption breaks, so **measure V and I, and normalize for voltage** before any NILM.[12]

WHY NOT JUST METER EVERYTHING

You can't put a sensor on every appliance

Two ways to know what each appliance uses:

Meter every appliance

- ✕ a plug sensor per device — cost adds up
- ✕ install labour in every home
- ✕ sensors fail, labels rot (deployment reality)
- ✕ intrusive — hardware in every socket

Use the whole-home meter

- ✓ already on ~1 B+ homes
- ✓ no new hardware, no install
- ✓ nothing extra to maintain
- → but you only see the sum

NILM: infer per-appliance power from the aggregate — no new sensors

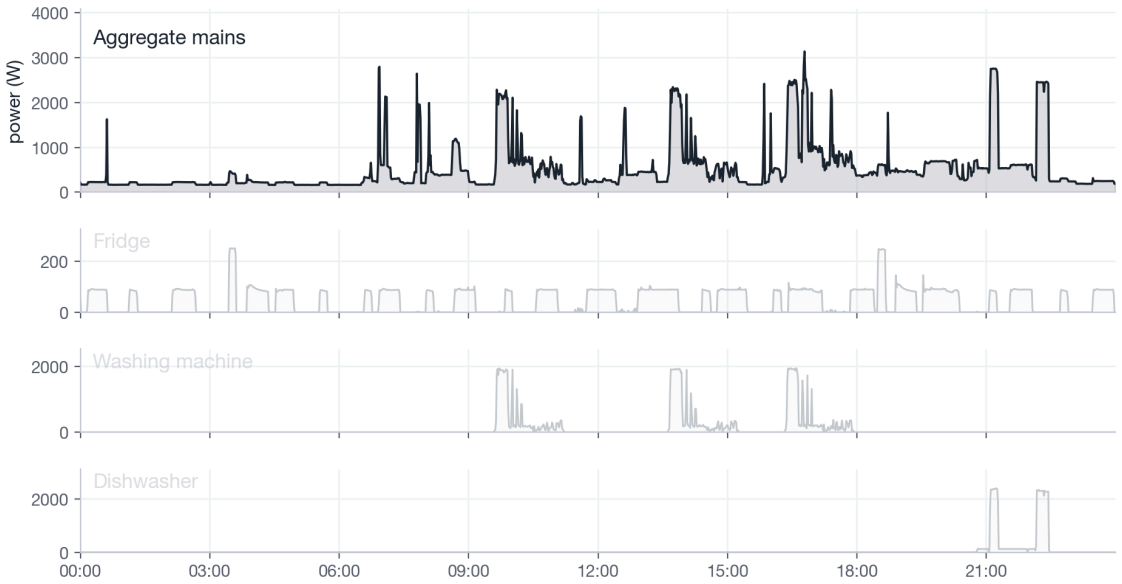
TAKEAWAY

Metering everything is accurate but doesn't scale; the meter you already have sees only the sum. **Non-intrusive load monitoring (NILM)** recovers the parts from that sum — our first deep dive.

THE PROBLEM

One aggregate signal is all you measure…

GIVEN	FIND	SUBJECT TO
aggregate meter y(t)	**per-appliance power $x_i(t)$**	**no per-appliance sensors**



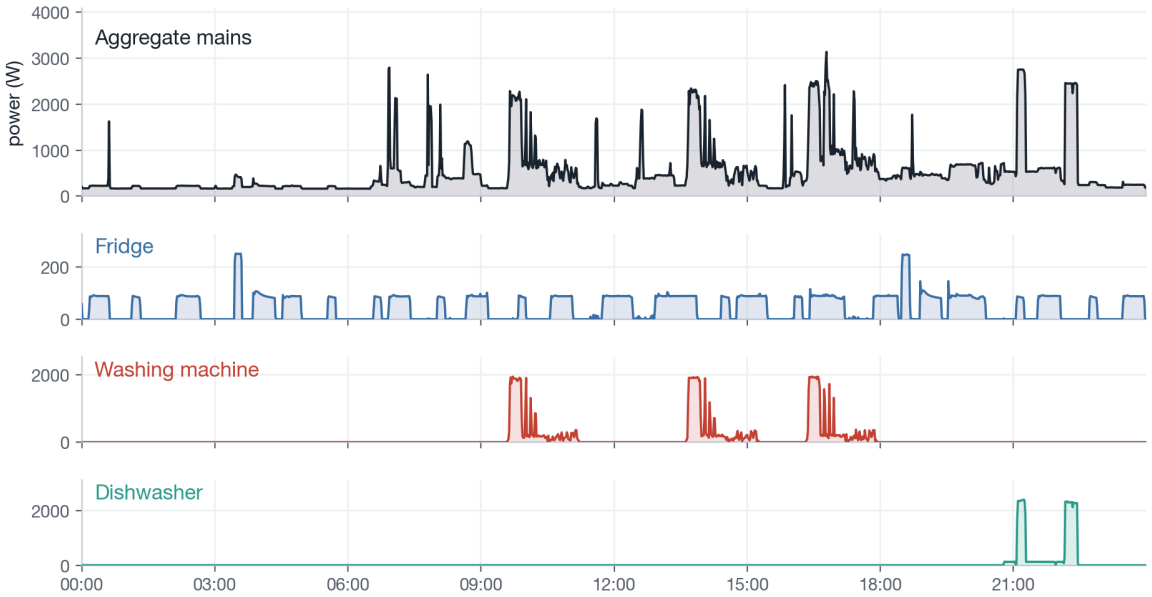
TAKEAWAY

At the meter you get **one** whole-home trace — the sum of everything on. The appliance panels below are ghosted because you never measure them directly.

EXERCISE · 60 SECONDS

Disaggregate by eye: where does the **washing machine** run? How many **fridge** cycles can you count? Which spikes are the **kettle**?

…pull each appliance back out of the sum



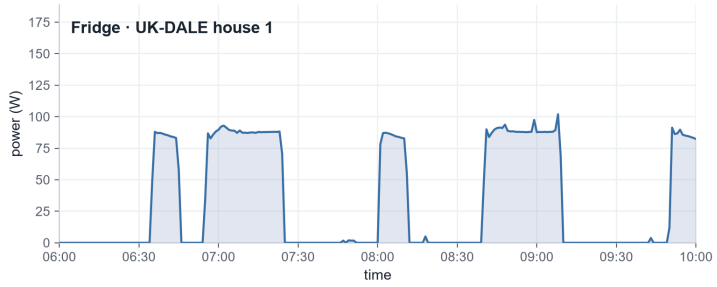
Real data · UK-DALE house 1

TAKEAWAY

Same signal, appliances revealed — with **no new sensors**. Posed by **Hart in the late 1980s** as blind source separation with structure.[13]

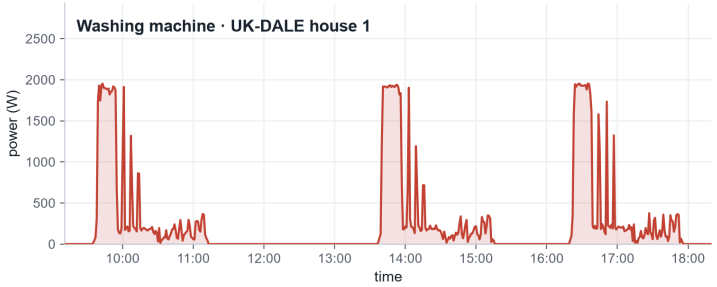
WHY IT'S HARD

Appliance signatures differ by load type



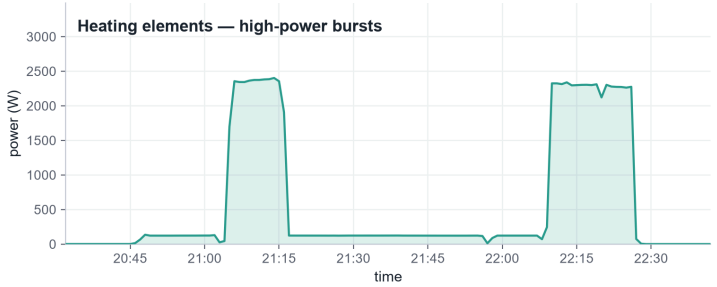
Fridge · periodic

Compressor cycles ~80–120 W; always-on, easy to spot.



Washer · multi-stage

Heat → agitate → spin; long, variable, many sub-states.



Heating elements · spiky

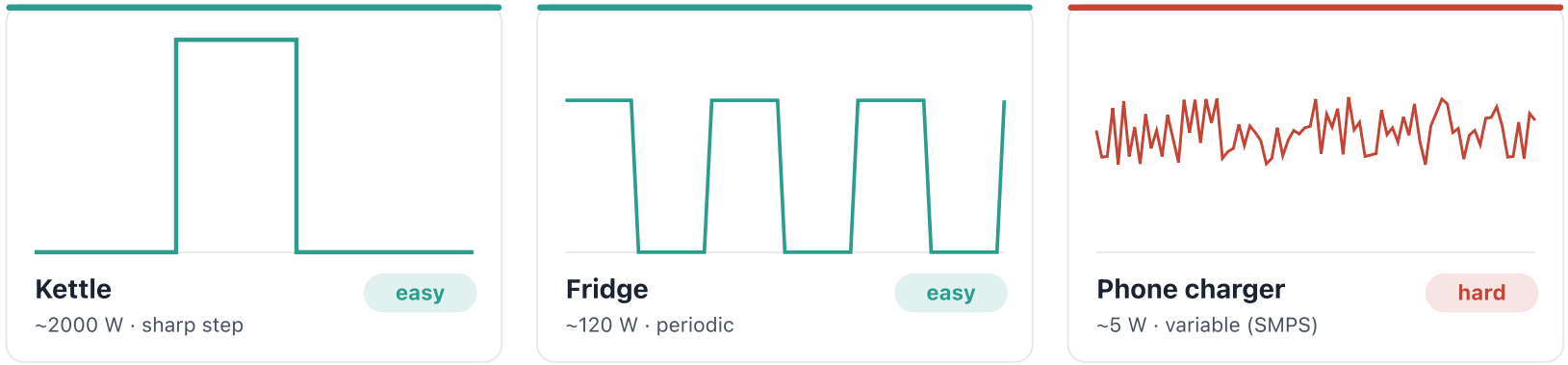
Dishwasher & kettle pulses look alike — the classic confusion.

Think: a kettle, a fridge, and a phone charger — which is hardest for NILM to recover, and why?

THE ANSWER

Small, variable loads are the hard ones

Hardest to recover: small, variable electronic loads — they barely move the meter



switched-mode supplies draw non-linear, varying current — no clean fixed step to latch onto

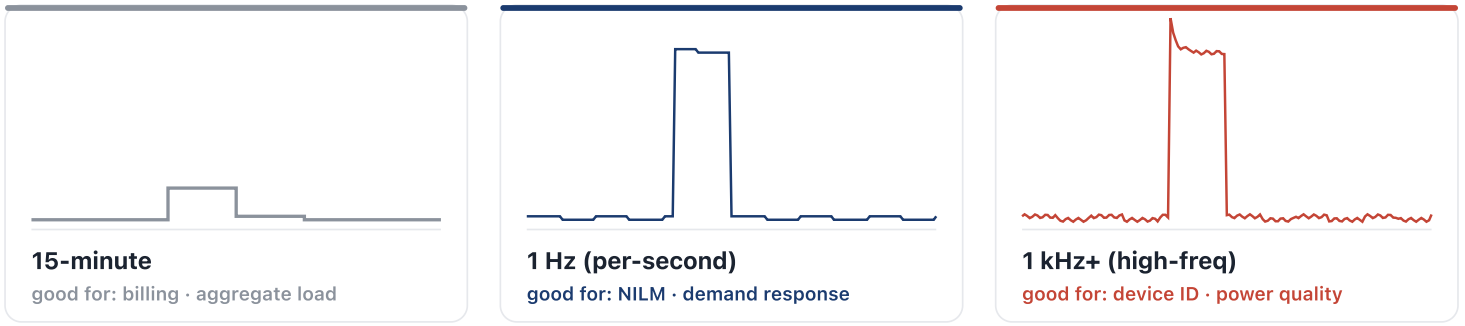
TAKEAWAY

The **charger** is hardest: its few watts can sit below meter resolution or disappear under background variation and overlapping loads. The kettle is large and sharp; the fridge is periodic. At low frequency, **amplitude and identifiability** matter more than architectural sophistication.

A DESIGN CHOICE

How fast you sample changes the problem

How fast you sample decides what you can even see — and what it's good for

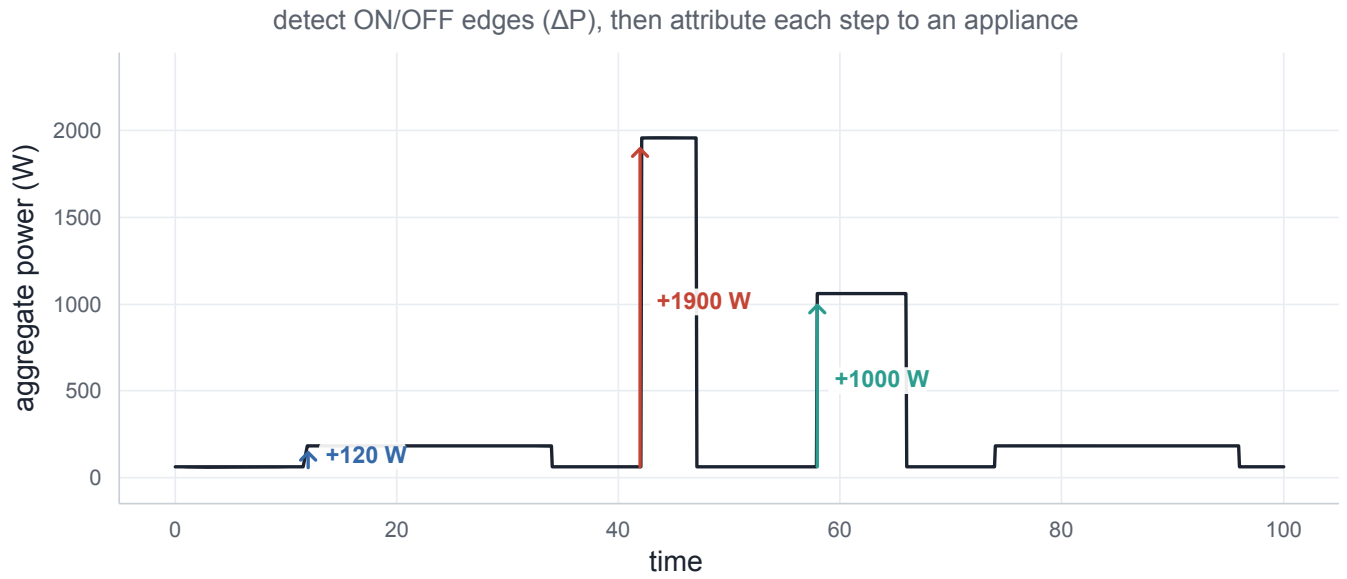


same window, three rates — coarse smears the kettle away, fine reveals its turn-on transient

TAKEAWAY

Coarse 15-minute data is useful for billing but smears short events away; **around 1 Hz** is a common low-frequency research setting; **kHz+** exposes transients and harmonics — at a storage and privacy cost. The useful rate depends on the appliance, meter, feature, and endpoint.

First idea: detect the ON/OFF edges

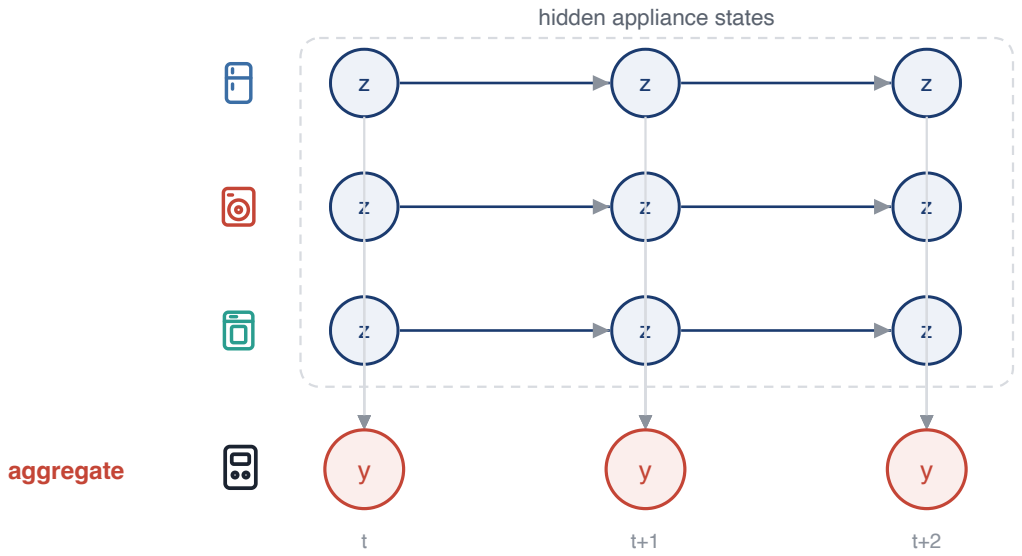


Appliances switch in steady **ΔP steps**. Hart's method: detect edges, cluster them by size, and match each ON to a later OFF.[13]

Breaks down when appliances overlap or vary continuously (dimmers, motors, inverters) — and the state-matching is combinatorial.

HISTORICAL · 2010–2012

Then: each appliance as a hidden state machine

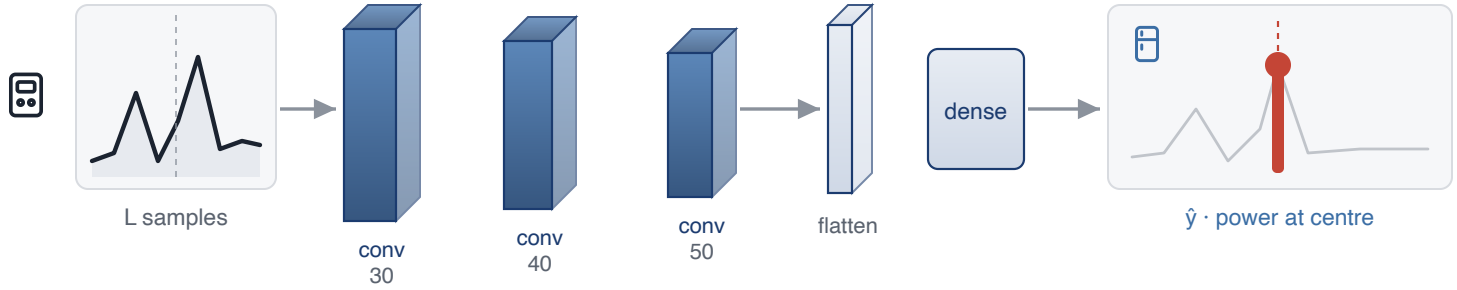


TAKEAWAY

Model each appliance as a **hidden Markov model**; the meter observes their sum. A factorial HMM is principled, but needs a model per appliance and makes inference combinatorial.[14]

Deep learning: map the mains window to one appliance

aggregate **window** (L) → one **appliance** at the centre

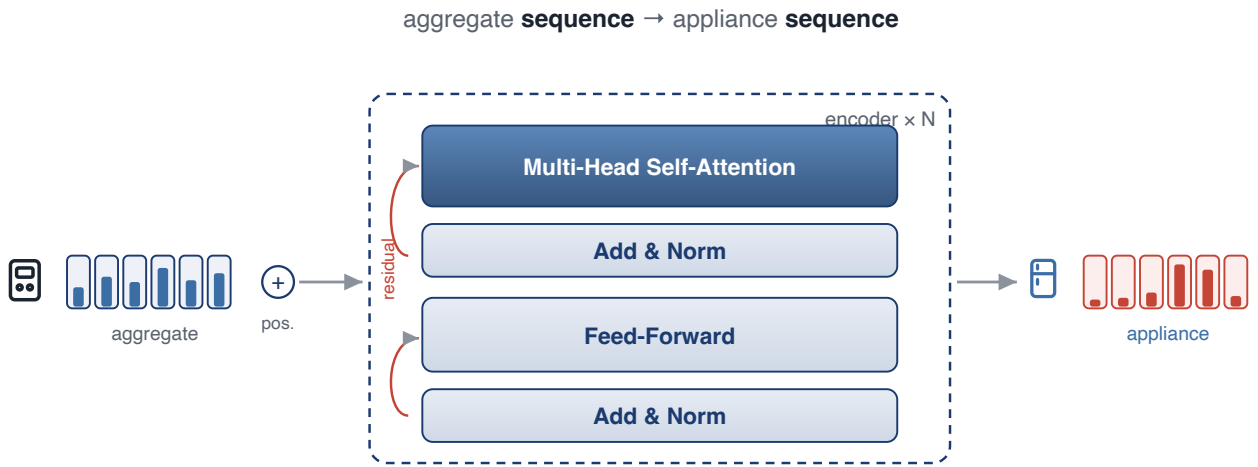


TAKEAWAY

Neural NILM reframed it as **regression**: a CNN maps a window of aggregate power to an appliance's power at the centre (**seq2point**).[15, 16] It learns signatures end-to-end — no hand-built edge rules.

MODERN · TODAY

Now: sequence models, attention, and pretraining



TAKEAWAY

Attention maps **sequence → sequence** and captures long context; the frontier is **pretraining across many homes**, then transferring to an unseen one — NILM's version of a foundation model.[17]

COMPARING METHODS

Which method wins depends on the metric

NILM OOD-first Leaderboard

Rank	Model	MAE (↓) T1 same home	MAE (↓) T2 new home	MAE (↓) T3 new dataset	F1 Score (↑) T1 same home	F1 Score (↑) T2 new home	F1 Score (↑) T3 new dataset	Params (M)	FLOPs (G)
1	**Model A**	**24.1**	**32.6**	**41.8**	**0.842**	**0.759**	**0.621**	**15.8**	**32.1**
2	Model B	22.7	34.8	48.7	0.830	0.731	0.572	28.3	55.4
3	Model C	23.9	36.2	49.5	0.821	0.722	0.560	9.7	18.9
4	Model D	25.8	38.9	53.6	0.797	0.689	0.512	42.6	81.3
5	Model E	27.1	41.0	56.8	0.781	0.653	0.489	6.3	12.7
6	Model F	26.4	42.3	59.7	0.768	0.628	0.456	18.2	29.6

Ranked by **cross-dataset generalization** — T1 same home → T2 new home → T3 new dataset

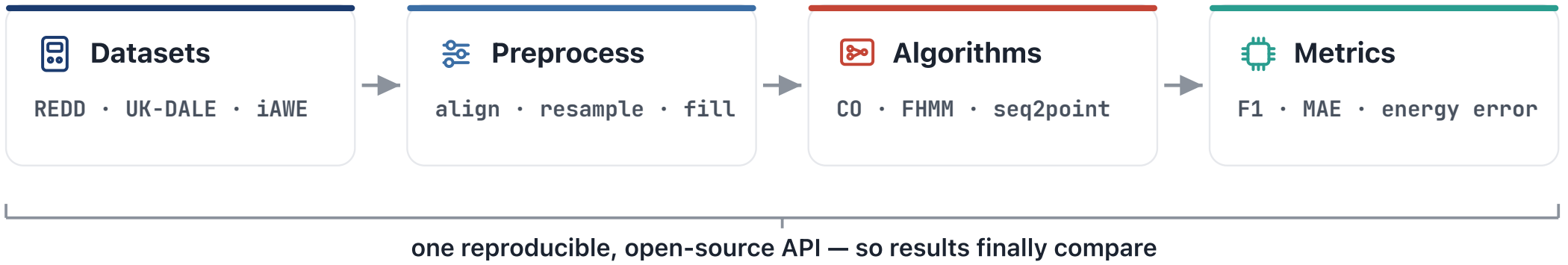
TAKEAWAY

Rank by **MAE** or by **F1** and you get different winners; score on the **same home** (T1) vs a **new dataset** (T3) and they flip again — and the biggest model is rarely the most efficient. Always ask **which metric, which house, seen or unseen**.[18] Models anonymized — NILMBench, our benchmark (in submission, 2026)

BENCHMARKING IS PART OF THE SCIENCE

NILMTK made methods comparable and reproducible

Before NILMTK every paper used its own data, metrics, and preprocessing

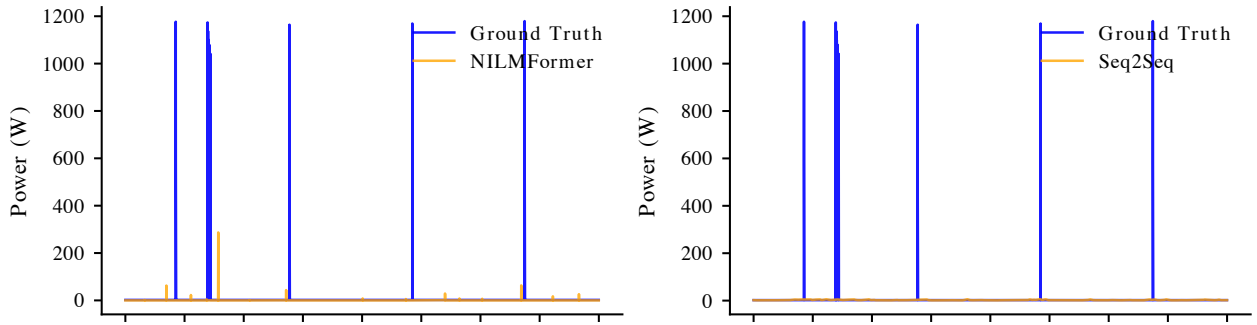


TAKEAWAY

A shared data model, preprocessing pipeline, algorithms, and metrics turned one-off demos into comparable experiments.[19, 20]
The broader lesson: **benchmarking infrastructure is itself a research contribution.**

HONEST SCORECARD

Accuracy is not the same as impact



Both models post low MAE yet miss almost every microwave event — NILMBench, our benchmark (in submission, 2026)

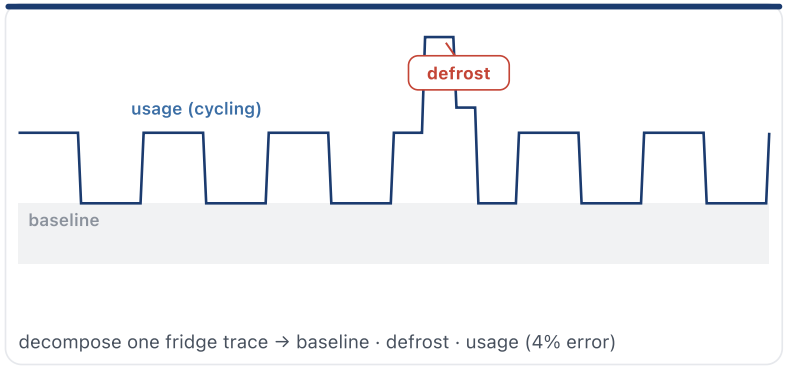
TAKEAWAY

A low **average error** can still miss the events that matter. On 240 homes, feedback accuracy stayed low even when disaggregation error was low — **"if you can measure it, you may not necessarily improve it."**[21]

WHERE THE VALUE ACTUALLY IS

The payoff is a specific action, not a number

Disaggregation pays off as specific, actionable feedback — not a generic number



decompose one fridge trace → baseline · defrost · usage (4% error)

Baseline too high → repair / replace	−26%
Excess defrost → service the unit	−25%
Heavy usage → habits & settings	−23%

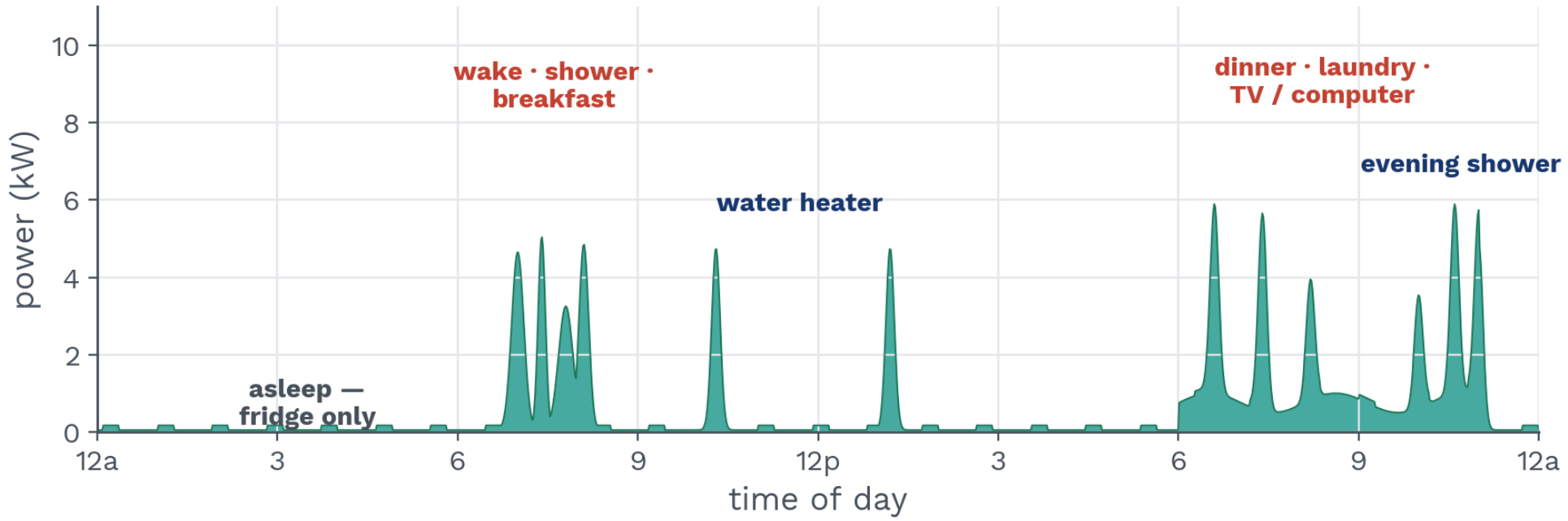
detects homes that need feedback with ~84% accuracy

TAKEAWAY

Split a fridge into baseline, defrost, and usage and you can tell a home **exactly** what to fix — worth **23–26%** of fridge energy.[21] This is Application 1's endpoint: **sense → model → a specific action a human can take.** Judge NILM on that job, not on MSE alone.

A smart meter reveals your whole day

Re-created from Molina-Markham et al., Private Memoirs of a Smart Meter, BuildSys 2010



TAKEAWAY

The same signals that cut energy expose when you wake, eat, and leave. Privacy is a design constraint, and the reason disaggregation raises real concerns.[22]

Three things that still break NILM in the wild



Identity

Two fridges can look alike electrically; a class label does not identify the failing unit.



Overlap

Small, variable electronics disappear under one another and the background load.



Transfer

UK → Delhi changes voltage, appliances, habits, and building stock.

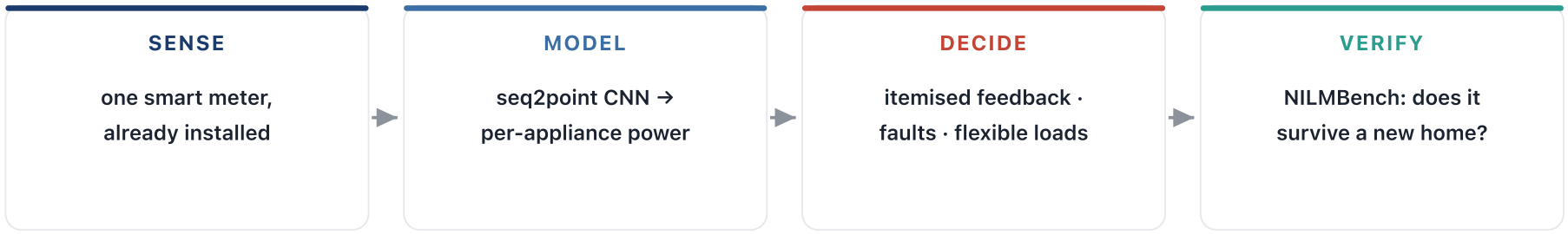
TAKEAWAY

A benchmark should test the **next device, next home, and next region** — not merely another window from the same meter.

SAME LOOP · APPLICATION 1

Disaggregation, end to end

Energy disaggregation — the same loop, end to end



swap the application, keep the loop — this is the transferable skill of the field

TAKEAWAY

One meter becomes an appliance estimate, then a targeted action, then a measured outcome. This **sense → model → decide → verify** loop is the template for the rest of the talk.

APPLICATION 2 · STAR: MODEL

Shifting demand to cleaner hours

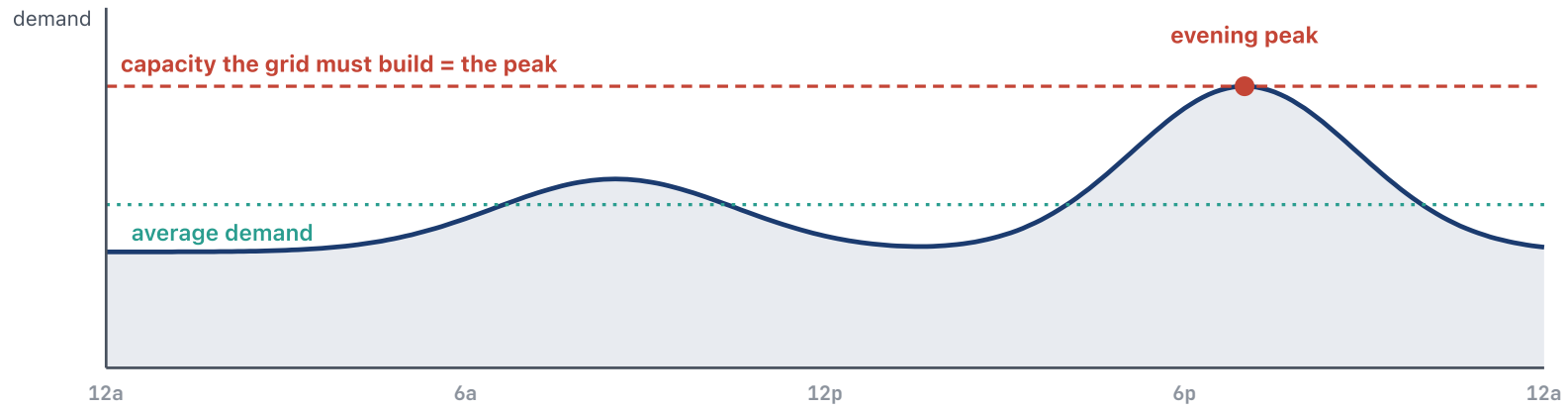
Forecast the constraint, move flexible load, protect the service

SENSE | **MODEL** | DECIDE | VERIFY

AI for Buildings · Climate Change AI Summer School 2026

Why the grid cares about one hour of your day

A day of city-scale demand — the whole grid is built for one evening hour



plants, wires, and transformers are all sized for the red line — even if it's touched for only a few hours a year

TAKEAWAY

Generation, wires, and transformers are sized around **peak demand**, not only average use. Lowering or shifting that peak can defer capacity and reduce cost. The carbon effect depends on **which generators respond** at the new hour.[23, 24]

WHY THOSE HOURS HURT

Peak demand is costly; its carbon is grid-specific



Capacity

A short peak can size years of generation, wires, and transformers.



Marginal resource

The next kWh may come from gas, hydro, imports, storage, or curtailed renewables.



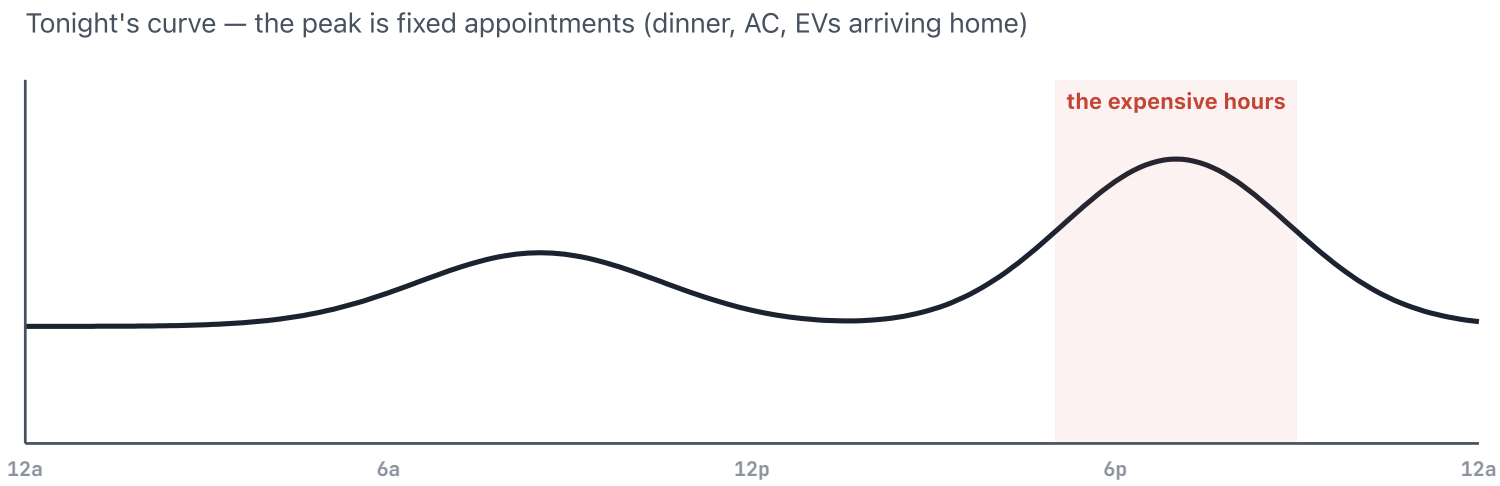
The bill

Demand charges price the highest kW, not only total kWh.

TAKEAWAY

A kWh at 6 pm is not equivalent to one at 2 pm: **price, congestion, and marginal emissions can all differ**.[24]

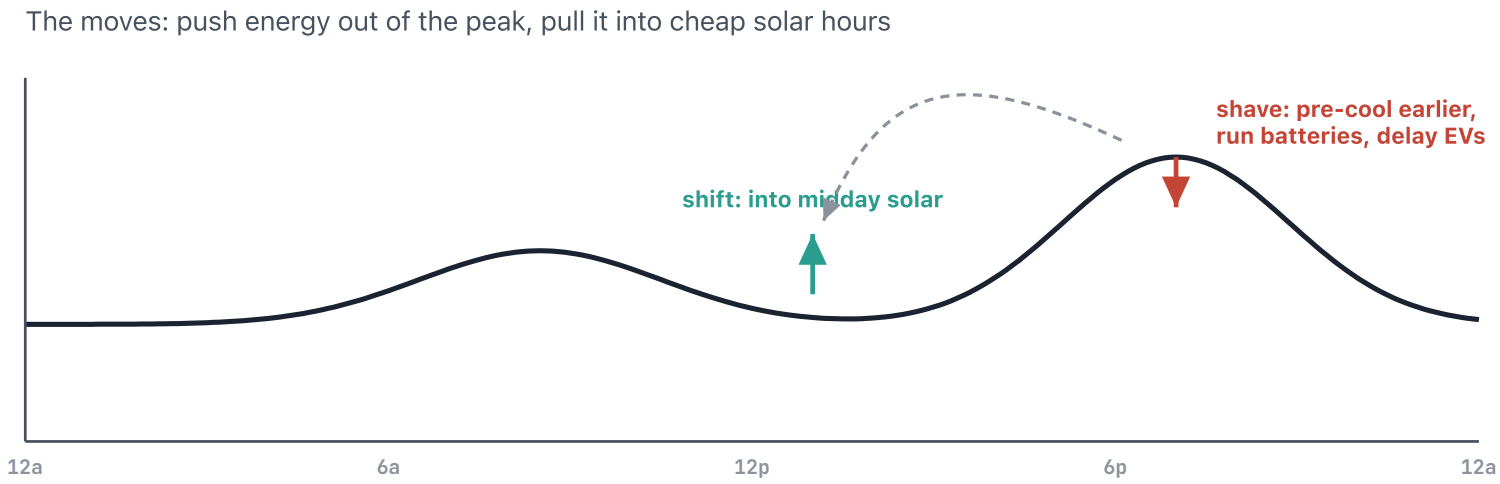
Shift energy, lower peak



the energy has to be served — but does it have to be served AT these hours?

Idealized first: dinner still gets cooked and EVs still get charged. What can change is **when** flexible energy is used.

Shift energy, lower peak

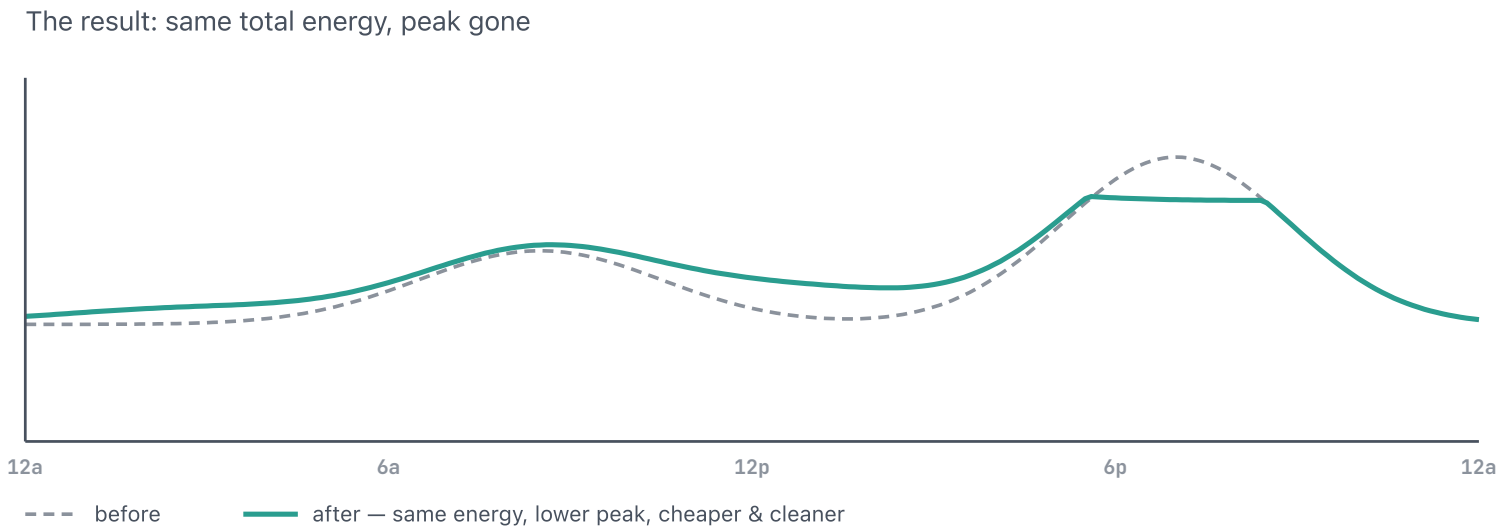


none of these change how much energy the building uses — only when

TAKEAWAY

Three levers: **shift** flexible demand, **shed** non-essential demand, or **store** energy. Pre-cooling and batteries introduce losses, so total energy can stay similar — or rise slightly — even while the peak falls.

Shift energy, lower peak



TAKEAWAY

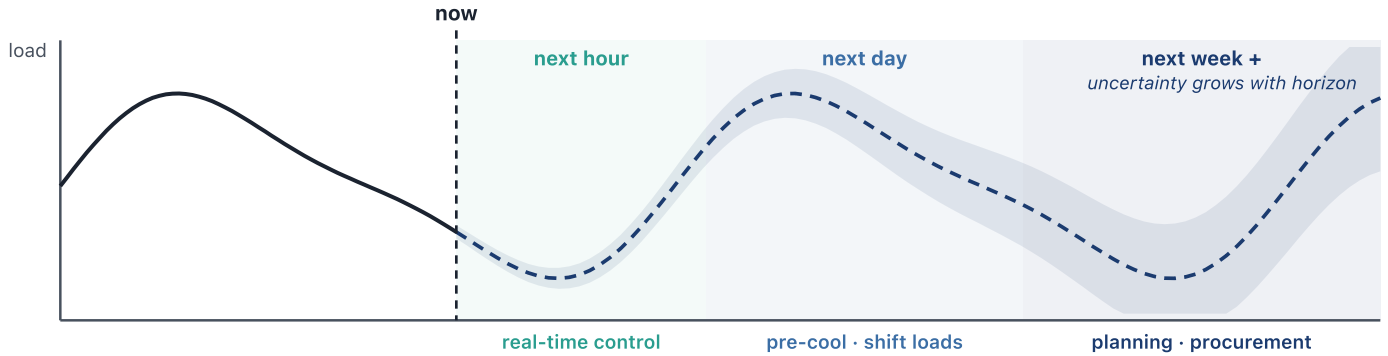
Peak and bill can fall. Carbon falls only if demand moves to hours with lower marginal emissions. Every claim therefore needs two checks: a forecast before the event and a counterfactual after it.[24]

THE ML PROBLEM

To act hours ahead, you must see hours ahead

GIVEN	FIND	USED FOR
weather · calendar · past load	**the load over the next hour to the next week**	**pre-cooling · load shifting · planning**

One forecast, many horizons — each feeds a different decision, each is less certain



TAKEAWAY

Pre-cooling needs the **next-day** curve; real-time control the **next hour**; procurement the **next season** — and error compounds with distance. This forecast is the input to everything that follows.

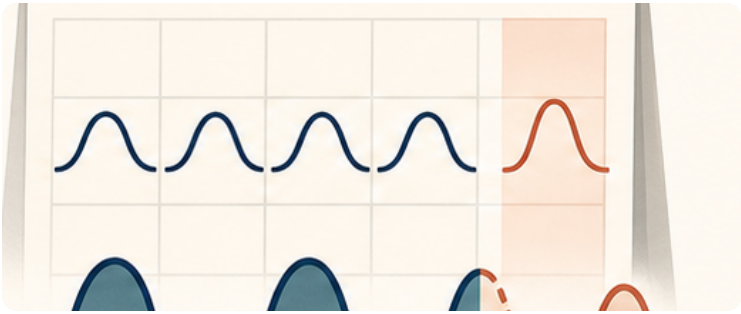
THE LONG-STANDING BASELINES

Regression and ARIMA, then gradient boosting



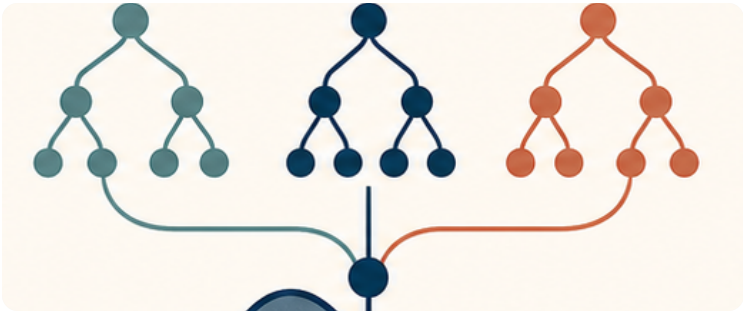
Weather regression

Temperature + calendar → an interpretable load response.



ARIMA / SARIMA

Continue autocorrelation and daily/weekly rhythm.



Gradient boosting

Combine engineered features with many small decision trees.

TAKEAWAY

Start with the baseline that matches the structure: weather, seasonality, then interactions. Gradient boosting won ASHRAE GEP III (~3,600 teams); deep learning is not automatically better.[25, 26]

The honest benchmark: persistence is hard to beat

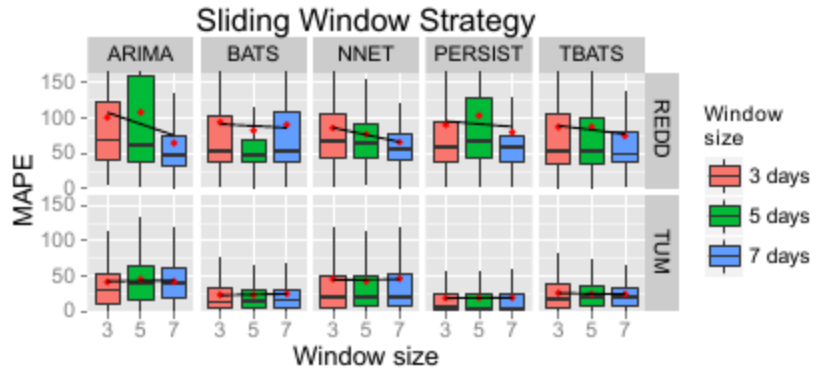


Fig. 1 · **Veit et al., e-Energy 2014** · doi:10.1145/2602044.2602082

TAKEAWAY

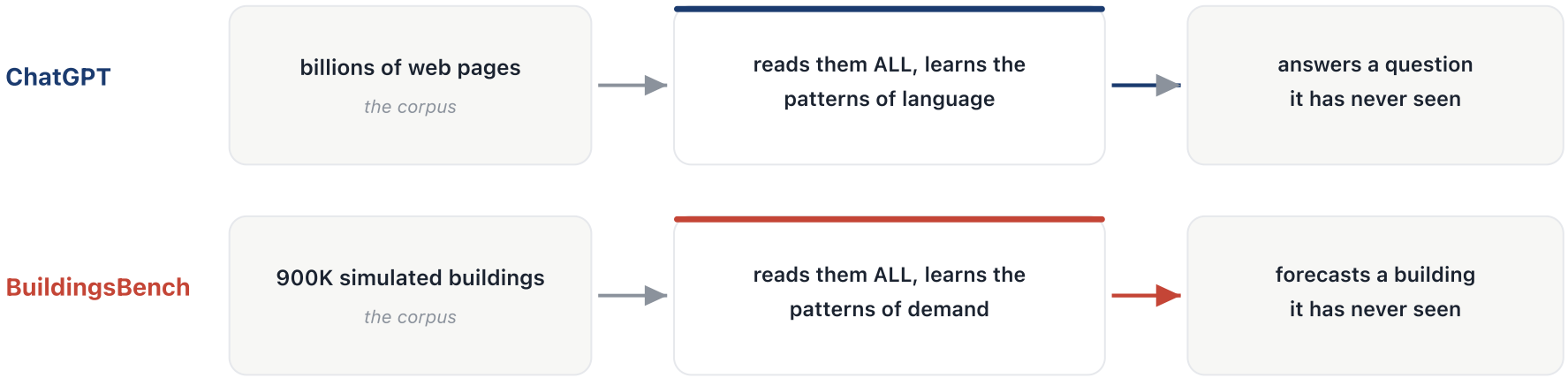
At **single-household** level, ARIMA, exponential smoothing, and neural nets **"rarely beat corresponding persistence forecasts"** — repeating the last observation sits in the same error range. [27]

And in a real office building, an occupancy sensor nudged forecast MAPE only 1.24% → 1.22% — while simply **re-forecasting every hour cut RMSE 73%**. Update frequency beat both the sensor and the model class.[28]

NOW: PRETRAIN, THEN TRANSFER

Pretrain on 900K simulated buildings

The same pretraining idea as ChatGPT — but load curves replace text

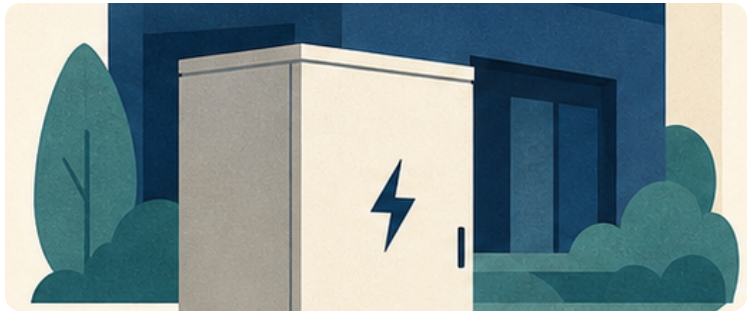


same strategy, different data — pretraining supplies a prior; transfer still has to be tested on real buildings

TAKEAWAY

The ChatGPT analogy is pretraining, not equivalence. BuildingsBench learns a broad prior from **900K simulated buildings**, then evaluates transfer on 1,900+ real buildings. The test is whether that prior helps a new building with little local history.[29]

Storage, EVs, and thermostats — a virtual power plant



Battery

Charge off-peak; discharge at peak.[30]



EV fleet

Share feeder capacity; meet every departure.[31]



Thermal loads

Shift AC, water heating, and refrigeration within service limits.[32]

TAKEAWAY

Coordinate many small assets and the grid sees one **dispatchable, constraint-aware resource** — a virtual power plant.[33]

APPLICATION 3 · STAR: DECIDE

Comfort-aware building control

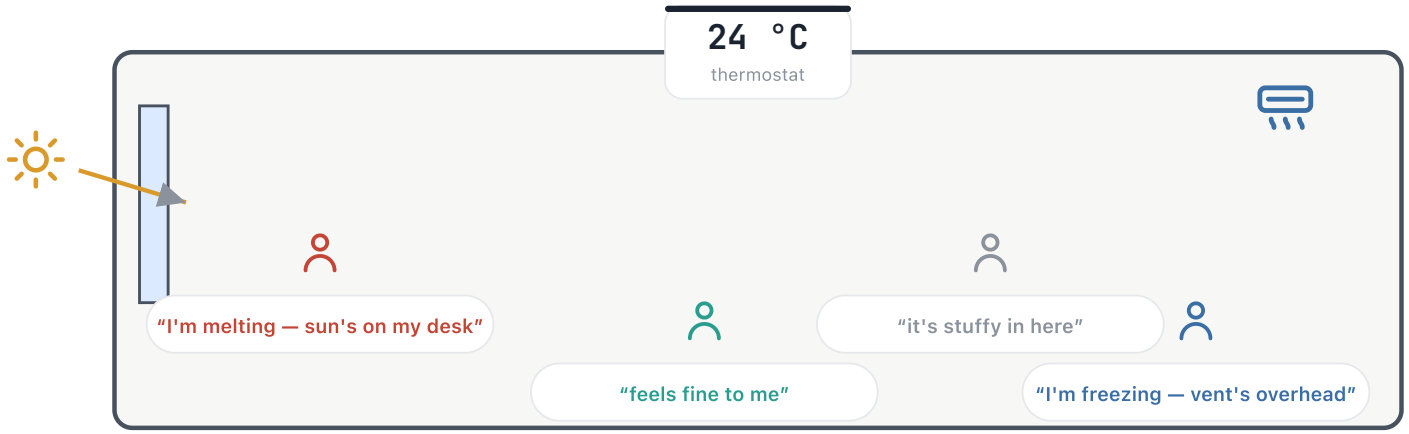
Occupants in the loop — safe targets out, measured outcomes back

SENSE MODEL **DECIDE** VERIFY

AI for Buildings · Climate Change AI Summer School 2026

The thermostat war: everyone knows this room

One meeting room, one thermostat — four different experiences



the thermostat reads ONE number; four bodies, four micro-climates, four verdicts

sun on your desk, a vent overhead, what you're wearing, what you just did — none of it is on the thermostat

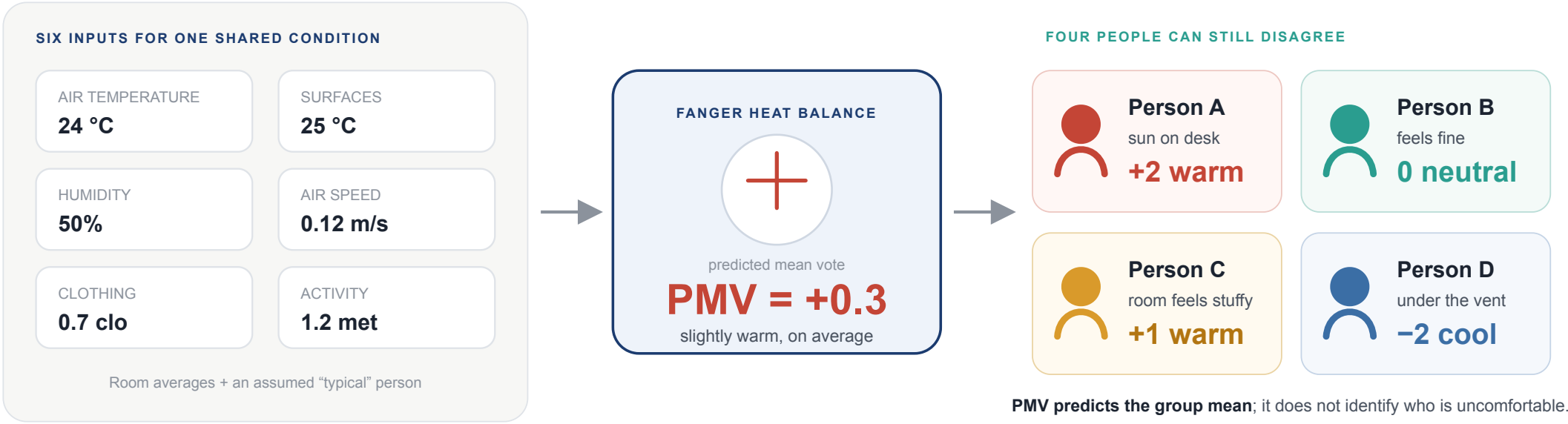
POLL

Hands up: who has fought over an office thermostat — or worn a jacket indoors in summer?

TAKEAWAY

The thermostat reads **one number**; the room contains four micro-climates and four bodies. Comfort is not a temperature — so what is it?

PMV predicts the group's average vote — not each person



PMV predicts the group mean; it does not identify who is uncomfortable.

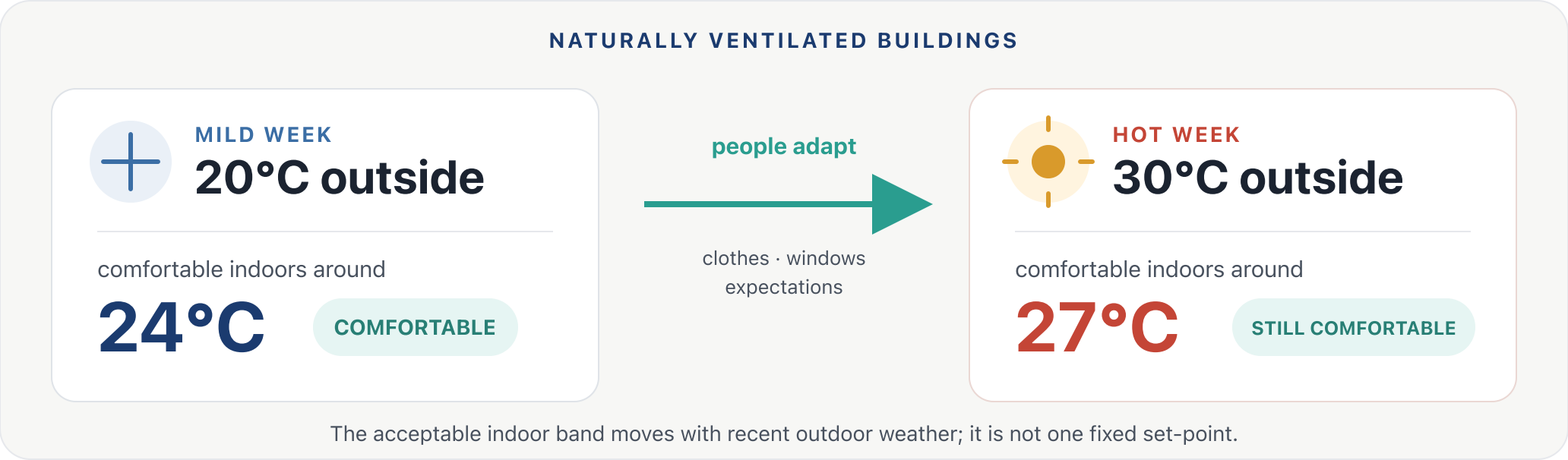
TAKEAWAY

PMV = Predicted Mean Vote. Fanger's heat-balance model combines air and radiant temperature, air speed, humidity, clothing, and activity. Here it predicts the group average is **+0.3: slightly warm** — while the four people still disagree.[34, 35]

Control implication: use PMV to set a reasonable starting band; use occupant feedback and local adjustment as stronger evidence for the person in front of you.

COMFORT IS PARTLY LEARNED FROM THE CLIMATE

On a hot week, 27°C can still feel comfortable



The acceptable indoor band moves with recent outdoor weather; it is not one fixed set-point.

TAKEAWAY

In **naturally ventilated buildings**, people change clothing, use windows and fans, and shift expectations. The acceptable indoor band rises with recent outdoor temperature; it is **not one fixed set-point** — and not a claim that every air-conditioned office should be 27°C.[36]

SET THE SCENE · ONE ROOM, ONE AFTERNOON

At 14:00, one room becomes a control problem



What is coming?

17:00: hotter weather, bright sun, high electricity price, eight-person meeting.

What is fragmented?

HVAC, blinds, and lighting each have their own sensors and local controller.

What must be coordinated?

Comfort, 300–500 lux without glare, low peak kW, and safe equipment limits.

INPUTS

room state · occupancy · weather · price/carbon · daylight

OUTPUTS EVERY 15 MIN

HVAC temperature target · blind position · light level

TAKEAWAY

The control problem: choose coordinated targets that keep people comfortable and the room usable while reducing energy, cost, or carbon. The forecast matters because it lets the building act **before** the peak.

THREE VERTICAL SYSTEMS · ONE COUPLED ROOM

HVAC, blinds, and lights cannot optimize independently



HVAC

More solar heat → more cooling.

Blind

Close for glare → lose daylight.

Lights

Brighten to replace daylight → add heat.

TAKEAWAY

The blind closes, lights brighten, and HVAC removes the lights' heat. Each vertical system meets its own target; the room wastes energy. Coordination is useful precisely because one action changes several outcomes.

THREE VERTICAL SYSTEMS · ONE COUPLED ROOM

HVAC, blinds, and lights cannot optimize independently



HVAC

More solar heat → more cooling.



Blind

Close for glare → lose daylight.

Lights

Brighten to replace daylight → add heat.

TAKEAWAY

The blind closes, lights brighten, and HVAC removes the lights' heat. Each vertical system meets its own target; the room wastes energy. Coordination is useful precisely because one action changes several outcomes.

THREE VERTICAL SYSTEMS · ONE COUPLED ROOM

HVAC, blinds, and lights cannot optimize independently



HVAC

More solar heat → more cooling.



Blind

Close for glare → lose daylight.



Lights

Brighten to replace daylight → add heat.

TAKEAWAY

The blind closes, lights brighten, and HVAC removes the lights' heat. Each vertical system meets its own target; the room wastes energy. Coordination is useful precisely because one action changes several outcomes.

One sensor, three names — so every application becomes bespoke

The same sensor — a zone temperature — in three vendors' languages



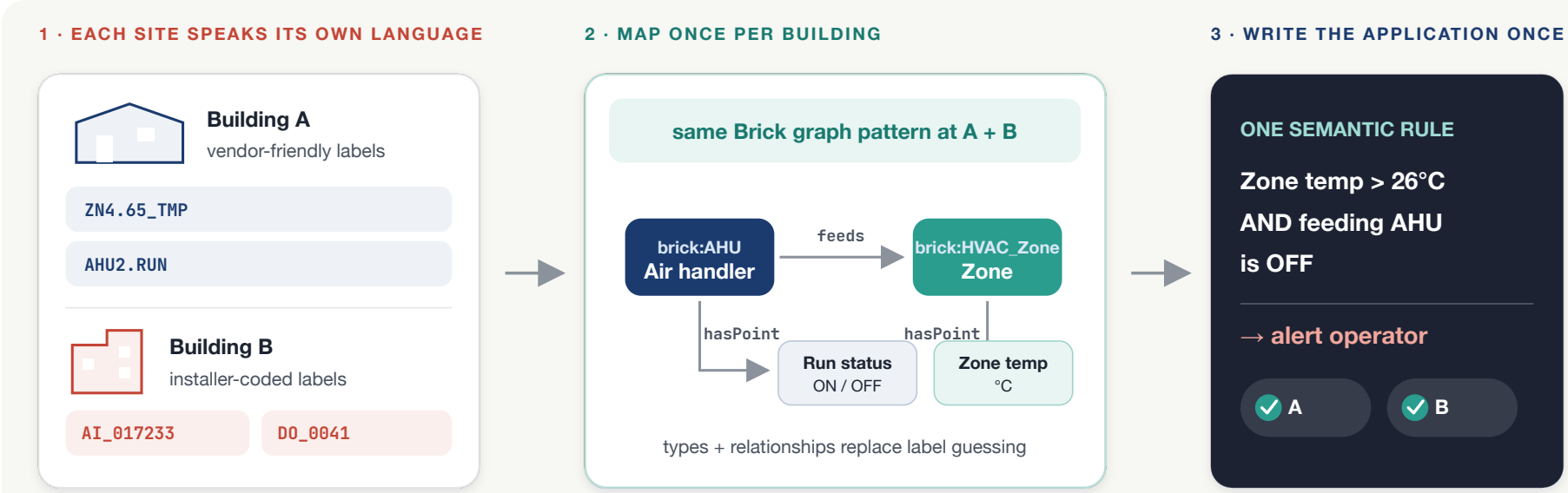
a large building has 1,500–2,500 sensor points; labels are typed by installers, per project

an app written for one building silently breaks on the next — this is why building software doesn't scale

TAKEAWAY

Open three building-management systems and the same idea can appear as three opaque point names. The names are project-specific, inconsistently enforced, and sometimes hidden behind proprietary APIs. **The analytics is reusable; the integration is not.** Motivation adapted from the Brick Schema webinar at 05:03.[37]

Brick separates a point's meaning from its local name

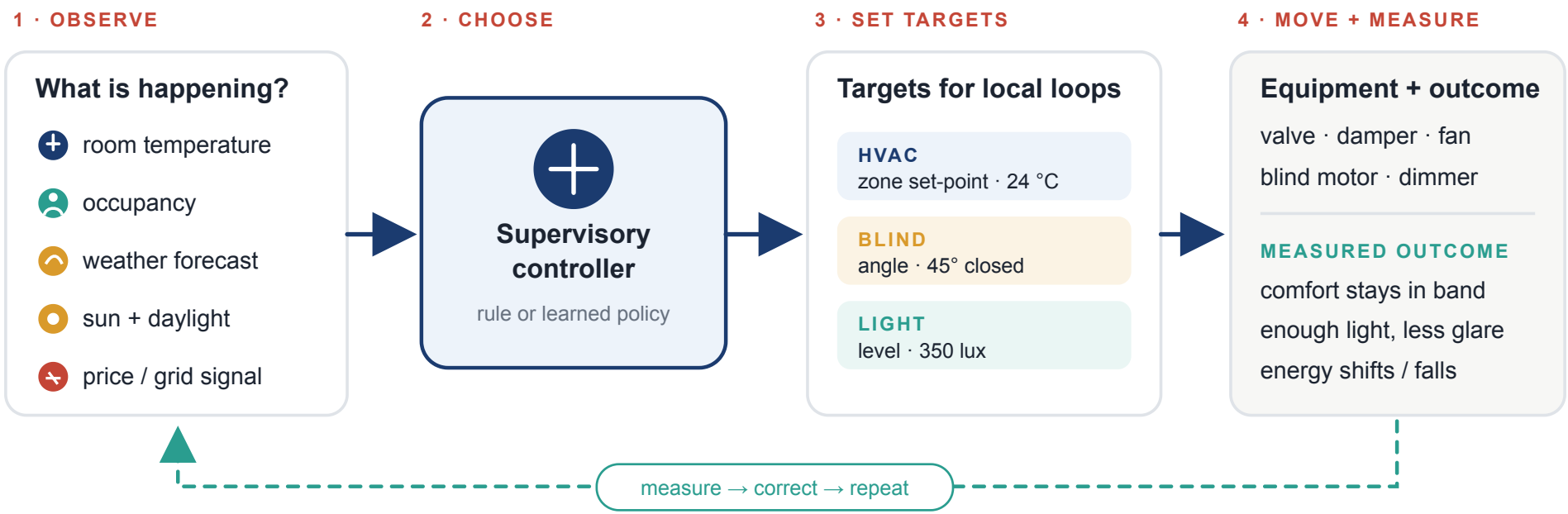


TAKEAWAY

Brick maps each site's local labels to shared **types and relationships**. The coordinator asks for “the zone temperature and blind actuator serving this room” — not a vendor tag. **Control chooses targets; metadata tells software where to read and write them.**[37]

FROM FORECAST TO PHYSICAL ACTION

Every controller does the same job: choose safe targets



TAKEAWAY

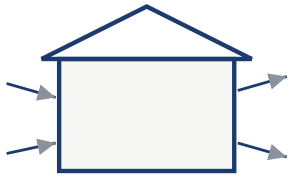
The AI does **not** directly drive a valve. It chooses the next safe targets; existing local loops track them, and sensors report whether comfort, light, and energy improved.

White, grey, or black box: three ways to predict the same room

Three ways to model the same building — how much physics do you keep?

WHITE BOX

full physics simulation



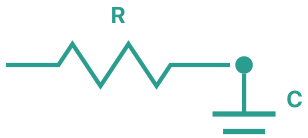
every wall, window & schedule modelled

`EnergyPlus / Modelica`

needs: blueprints + calibration

GREY BOX

a small circuit, fit to data



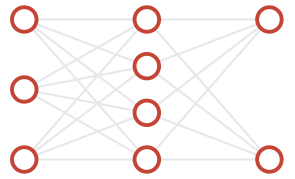
2 physical numbers, fit from data

`learn R and C from a week of logs`

needs: a little data + the analogy

BLACK BOX

any function the data supports



no structure assumed — data decides

`gradient boosting / neural nets`

needs: lots of data, no physics

more physics knowledge → more data

TAKEAWAY

These are models of the building, not controller types. A model-based optimizer can use detailed physics, a fitted thermal circuit, or a learned predictor; it needs a **useful** forecast of consequences, not perfect physics. Grey boxes are often the practical compromise.[38, 39]

Rules write the strategy; RL learns one from experience

RULES

A human writes the strategy

INPUTS NOW
occupied?
room temp
price signal

→

IF / THEN
IF occupied AND peak soon
THEN pre-cool 1°C

→

TARGETS
HVAC
blind
lights

GOOD WHEN
the logic is known, stable, and easy to bound

BREAKS WHEN many interactions create too many rules

REINFORCEMENT LEARNING

Experience writes the strategy

TRAINING · SEPARATE FROM DEPLOYMENT

simulator / logs → reward → update policy

SAME INPUTS
state + forecast + constraints

→

LEARNED POLICY
state → targets

→

TARGETS
HVAC · blind · lights

ONLY CREDIBLE WITH safe training · hard bounds · fallback · monitoring

TAKEAWAY

Start with rules. Consider RL only when coupled decisions repeat often, a credible simulator or offline dataset exists, and hard bounds, fallback control, and monitoring remain outside the learned policy.[40, 41]

APPLICATION 4 · DESIGN TIME

Searching the low-carbon design space

Every simulation is a point; the surrogate helps find the frontier

SENSE MODEL DECIDE VERIFY

AI for Buildings · Climate Change AI Summer School 2026

Same building, three ways to spend its carbon budget



Hold constant

Floor area, use, safety, and service life

Change

Structure, facade, shading, and HVAC

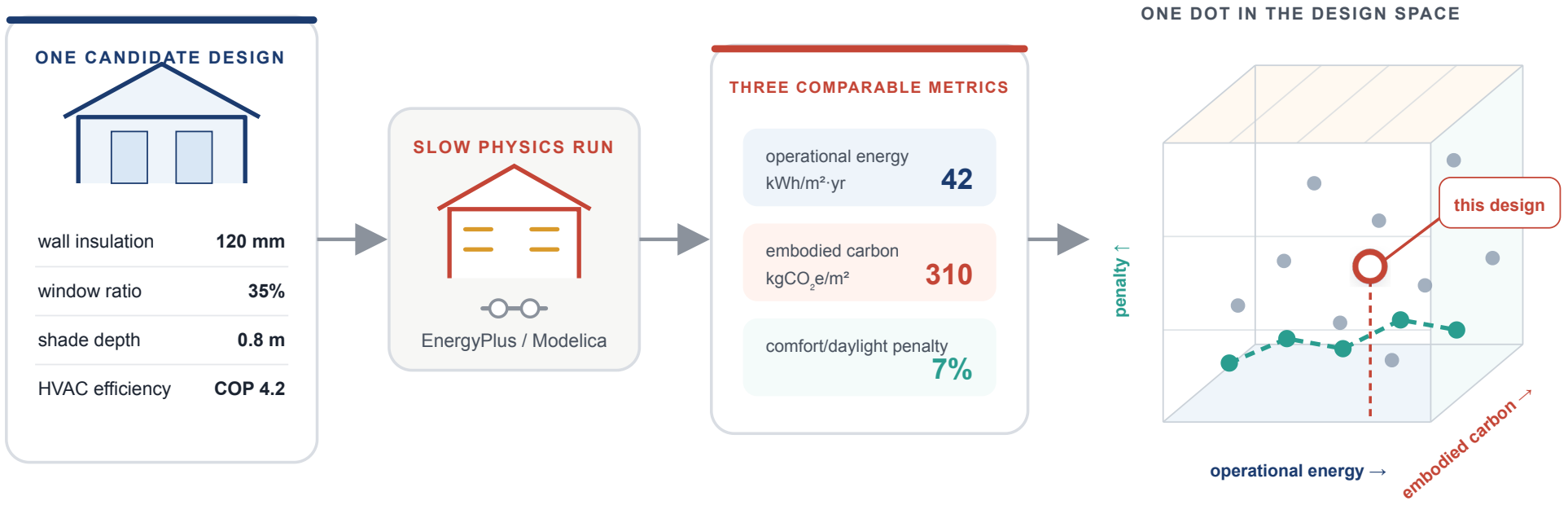
Compare

Upfront carbon + operating carbon over the chosen horizon

TAKEAWAY

The decision is not “timber good, concrete bad.” Product declarations, sourcing, durability, climate, and the future grid can change the ranking. **Compare whole-life options while design choices are still reversible.**[42]

Each simulation turns one design into one comparable point



TAKEAWAY

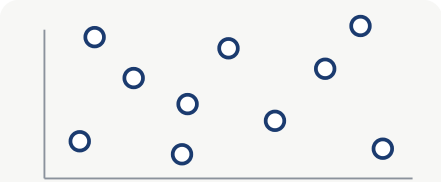
There is rarely one best building. The Pareto set is the small group worth discussing: improving one objective from any highlighted design would worsen at least one other objective.[43, 44]

Spend the slow simulation budget on 30 diverse designs

STEP 1 · BUY 30 HIGH-VALUE LABELS WITH PHYSICS

A · COVER THE DESIGN SPACE

Do not simulate 30 near-identical buildings.



30 diverse candidates

B · RUN THE SLOW SIMULATOR

Each run produces the trusted answer.



minutes → hours

run 19 of 30

C · KEEP INPUTS + VERIFIED OUTPUTS

This small table is the surrogate's training set.

design	energy	carbon	penalty
#01	42	310	7%
#02	51	244	5%
… #30	38	405	4%

30 expensive, trustworthy examples

The simulator is the labeler. Spend its limited budget on coverage, not repetition.

TAKEAWAY

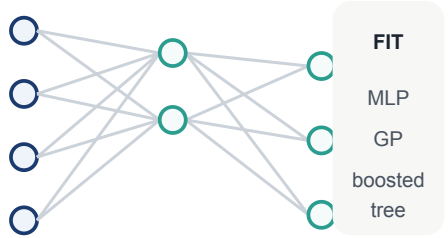
The simulator is the expensive labeler. Sample for coverage: combinations of envelope, glazing, shading, and HVAC that span the choices you might actually make.[43, 44]

Learn the map, screen 10,000 candidates, verify five

STEP 2 · LEARN THE MAP, SCREEN THOUSANDS, THEN VERIFY FIVE

A · FIT A FAST SURROGATE

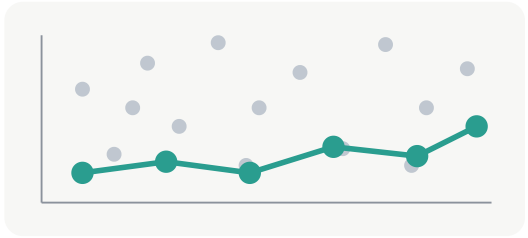
Learn design variables → three metrics.



milliseconds per prediction

B · PREDICT 10,000 CANDIDATES

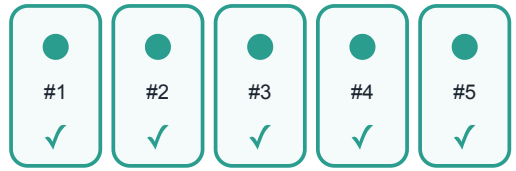
Keep only the non-dominated frontier.



seconds, not thousands of simulations

C · VERIFY THE FIVE FINALISTS

The surrogate proposes; physics confirms.



5 final physics runs

before a real design decision

The speed-up is useful only if the final shortlist returns to the trusted simulator.

TAKEAWAY

For scalar metrics, start with an **MLP, Gaussian process, or boosted tree**. A neural operator is more natural for **field → field** predictions. The surrogate proposes; **physics verifies the final five**.[43, 44]

THE THROUGHLINE

Better predictions matter only when they change a building decision

NILM
aggregate signal → appliance action

Forecasting
future constraint → timed schedule

Control
room state → safe system targets

Design
slow simulations → verified shortlist

`sense → model → decide → verify`

Thank you

Questions, collaborations, and student projects welcome

nipun.batra@iitgn.ac.in · sustainability-lab.github.io

AI for Buildings · Climate Change AI Summer School 2026

APPENDIX

Backup slides

Extra depth, kept for Q&A

AI for Buildings · Climate Change AI Summer School 2026

A compact objective makes the trade-offs explicit

$$\min_{u_{t:t+H}} \; J = \sum_{\tau=t}^{t+H} \left[w_E e_\tau + w_T D_\tau^{\text{thermal}} + w_V D_\tau^{\text{visual}} \right] + w_P \max_{\tau} P_\tau$$

Inputs

Zone state, occupancy, weather, solar, tariff, and carbon forecasts.

Decisions

HVAC set-point, airflow, blind angle, and light level within equipment limits.

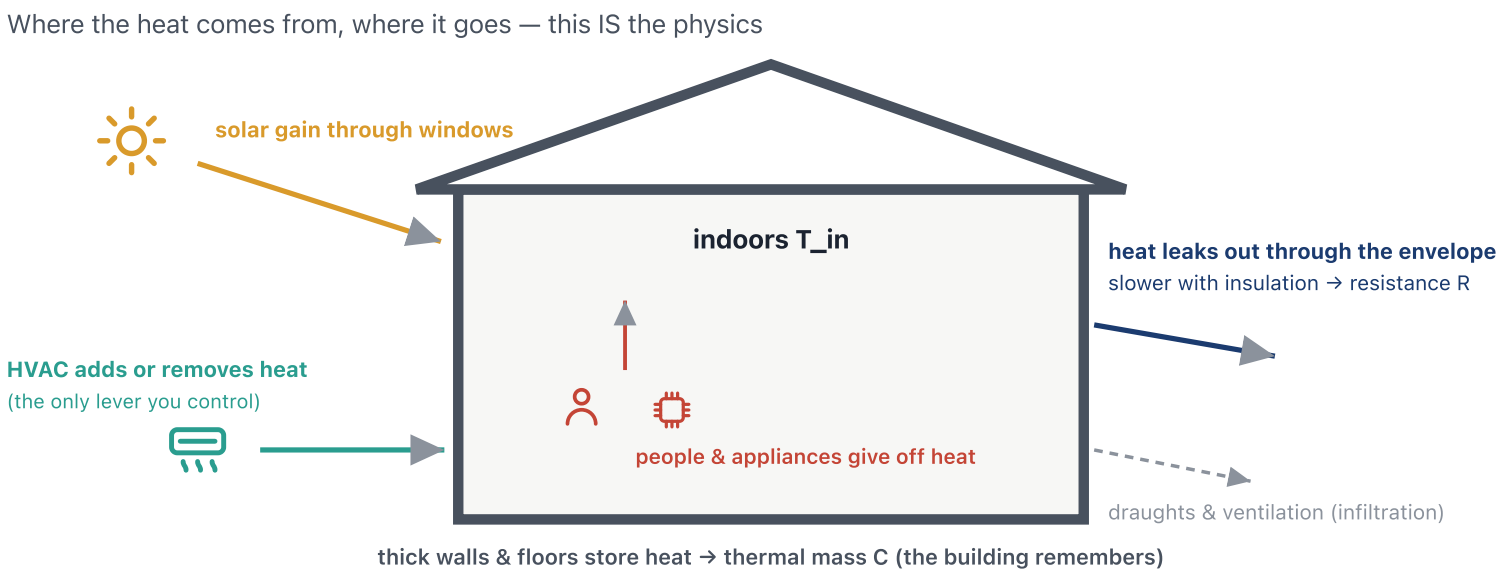
Outcomes

Energy/cost/carbon, thermal and visual discomfort, and peak demand.

TAKEAWAY

Weights expose value judgments; **hard constraints** protect IAQ, equipment, and minimum comfort. The main talk uses the same structure without the notation.

A building is a leaky, sun-lit box with memory



TAKEAWAY

Sun enters, heat leaks through the envelope, draughts cross gaps, people and appliances add heat, HVAC pushes back, and the structure stores heat. That stored heat is why pre-cooling can move demand through time.

BACKUP · LOCAL FEEDBACK

A supervisory rule chooses the target; PI/PID tracks it

```
# supervisor: choose the zone target
T_sp = 24.0 if occupied else 28.0

# installed local loop: move the cooling valve
error = T_room - T_sp
integral += error * dt
derivative = (error - previous_error) / dt

cooling = clip(Kp * error
             + Ki * integral
             + Kd * derivative, 0.0, 1.0)
```

P

Responds to error now.

PI · common in HVAC

Also removes persistent offset.

PID

Also reacts to the slope; derivative can amplify noise.[45]

TAKEAWAY

The local loop reacts in seconds. It does not know that a price peak is coming; the slower supervisor changes its target in advance.

Three plausible actions — three different answers

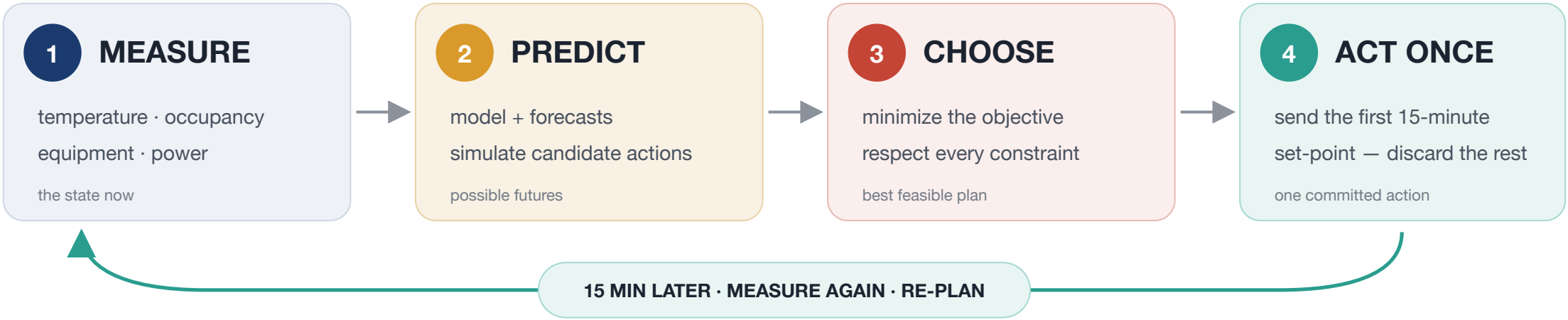
A · HOLD 24 °C	B · SWITCH OFF	C · PRE-COOL, COAST
TOTAL ENERGY: 15 kWh	TOTAL ENERGY: 7 kWh	TOTAL ENERGY: 16 kWh
BILL: ₹186	BILL: ₹42	BILL: ₹132
EVENT PEAK: 5.0 kW	EVENT PEAK: 0 kW	EVENT PEAK: 2.5 kW
DISCOMFORT: 0 min	DISCOMFORT: 75 min	DISCOMFORT: 0 min
Comfortable, but peak-heavy.	Cheapest because people pay.	1 kWh more, shifted earlier.

Illustrative arithmetic · tariff ₹6 before 17:00, ₹18 during the event

TAKEAWAY

No action wins every column. Switching off minimizes energy and cost but violates comfort. Pre-cooling uses 1 kWh more than the ordinary schedule, yet halves event peak and lowers the bill because energy moved earlier. The example is illustrative.

MPC is a disposable plan plus permanent feedback

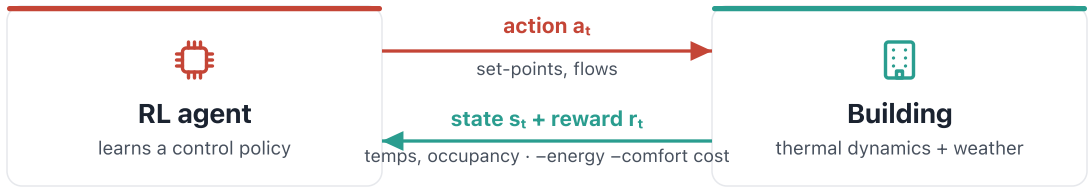


Window opens? Occupancy changes? The next loop changes the plan.

TAKEAWAY

MPC predicts candidate trajectories, chooses the cheapest feasible plan, applies only its **first** action, then measures and re-plans. Feedback repairs the plan when a window opens or the forecast is wrong.[39]

RL learns the same state ↠ target ↠ outcome interface



why real buildings are hard

safety | sample efficiency | sim-to-real gap | non-stationarity

TRAINING

repeat safely in simulator or offline data → improve long-run reward

> DEPLOYMENT
>
> **live state → learned policy → safe set-points; local loops execute them**

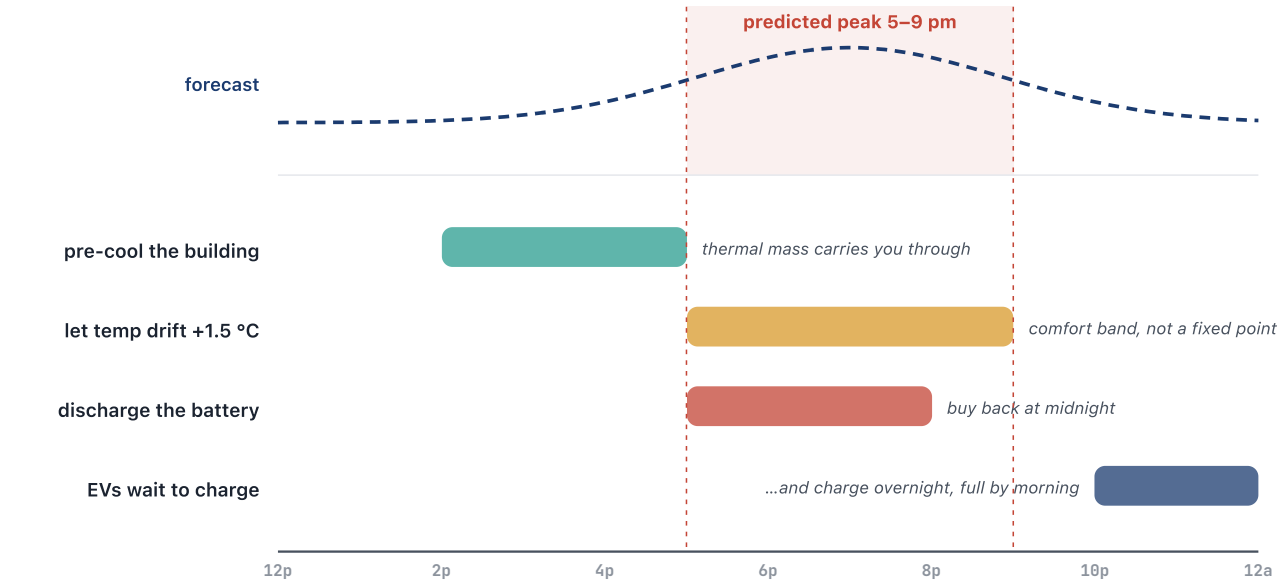
TAKEAWAY

The state, actions, reward, baselines, and safe training environment define the problem before the learning algorithm. Never learn by unsafe trial-and-error on occupants or equipment.[41]

BACKUP · FORECAST → SCHEDULE

Predict the 5–9 pm peak at noon; act before it arrives

Noon forecast → feasible actions before, during, and after the peak



TAKEAWAY

The forecast saves nothing by itself. It makes a sequence feasible: **pre-cool before** the peak, relax temperature and discharge a battery **during** it, then charge EVs **after** it.

BACKUP · MEASURED EV FLEXIBILITY

At the same 60 kW limit, scheduling serves every day

60 kW
the same Caltech charging-capacity limit in both cases

22%
of days feasible if every EV charges immediately

100%
of days feasible when charging is scheduled

TAKEAWAY

The practical value of flexibility is **hosting more EVs on existing wiring**. These results use 30,000+ real workplace sessions and respect each driver's energy need and departure time.[46]

Demand forecast says when; carbon forecast says when to act

Forecast

Estimate grid carbon intensity over the next 96 hours.

Choose

Find the cleanest feasible window.

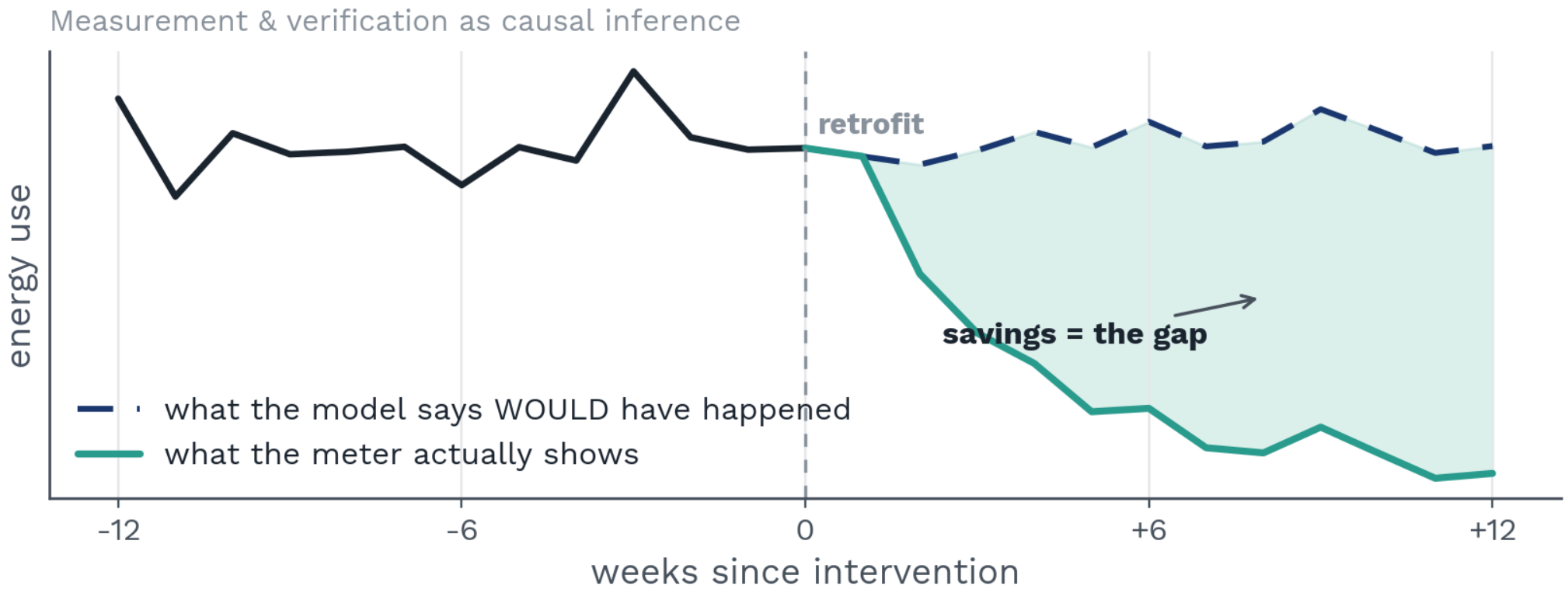
Act

Move EV, thermal, battery, or compute load.

TAKEAWAY

A shift can reduce price or peak yet **increase emissions**. CarbonCast achieved 4.8–13.9% MAPE across six US/EU grids at 96 hours.[47]

Forecasting creates the missing “no-retrofit” meter



TAKEAWAY

Decision enabled: verify ROI before scaling. Meter actual use, predict the no-retrofit counterfactual under comparable weather and occupancy, and call the gap realized savings.[48]

BACKUP · DR SETTLEMENT

DR payment = estimated baseline − metered load

100 kW
estimated no-event baseline

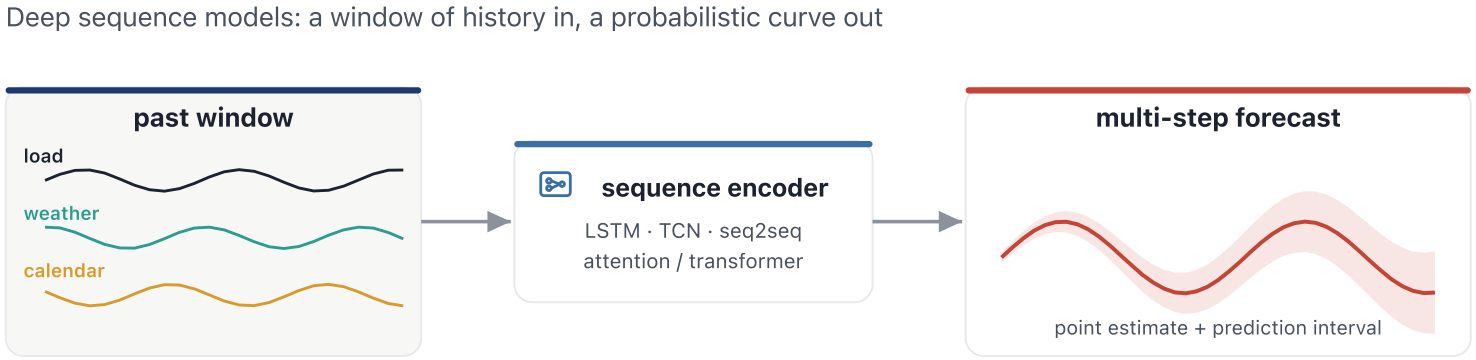
80 kW
metered event load

20 kW
reduction used for settlement

TAKEAWAY

If the true no-event load was 90 kW, only 10 kW was real: a 10 kW baseline error doubled the payment. **Baseline design is market design.**[49]

Deep sequence models: history in, a distribution out

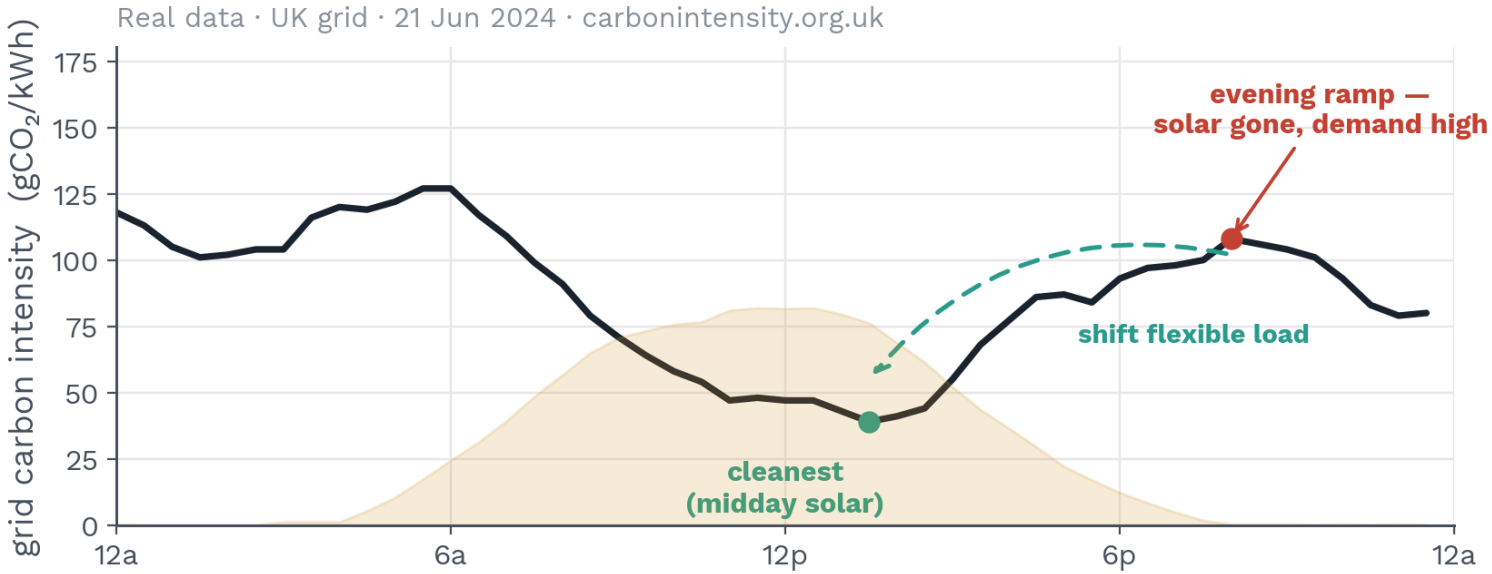


TAKEAWAY

LSTMs, temporal convolutions, and attention pay off when you have **many correlated series and long history**, or need calibrated intervals rather than a point estimate. They are data-hungry, so they rarely win on a single sparse building.[25]

A kWh at 6pm is not a kWh at 2pm

GIVEN	FIND	SUBJECT TO
hourly price / carbon signal	**when to shift flexible load**	**comfort + verifiable savings**



TAKEAWAY

You know these moves from the peak playbook — here they are **priced in carbon, on real grid data**: midday solar is ~2.6× cleaner than the evening ramp. Same playbook, second currency.[32]

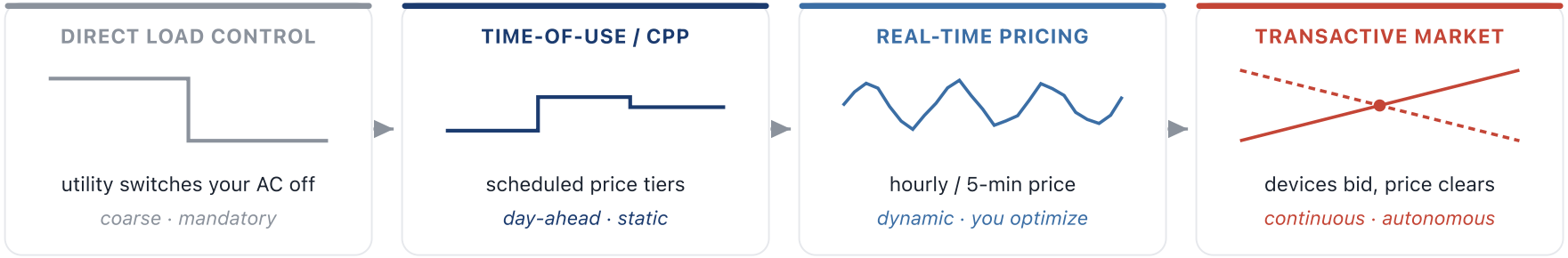
POLL

Your utility pays you to nudge your AC 1 °C during peaks — enroll, enroll only with an override button, or no?

HOW YOU DRIVE IT

From a switch, to prices, to a live market

The signal that drives flexibility got richer — from a switch to a live market



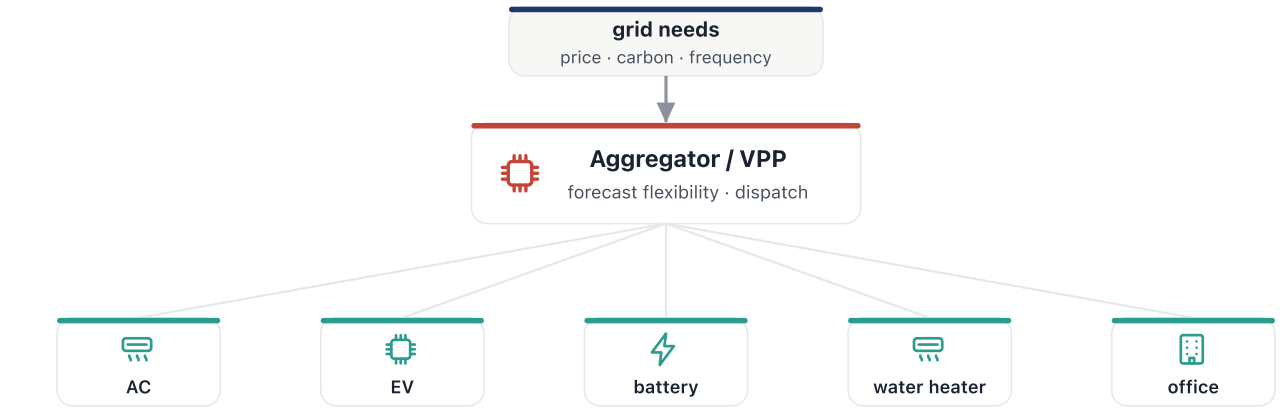
richer signal → more autonomy for the building, less top-down control

TAKEAWAY

Each step to the right hands the building **more autonomy** — and asks more of its ML: a schedule needs nothing, a live market needs forecasting, bidding, and control.[23, 50]

Coordinating a fleet without melting comfort

One aggregator, thousands of loads — ML dispatches flexibility within hard limits



TAKEAWAY

A population of thermostatically controlled loads behaves like **grid-scale storage** — set-point-control it for load-following and regulation.[51] The ML job: forecast each load's flexibility, then dispatch it to a grid signal while respecting comfort, feeder limits, and fairness — via device-level RL or multi-agent control.[52, 53]

Nine buildings learn to flatten a district

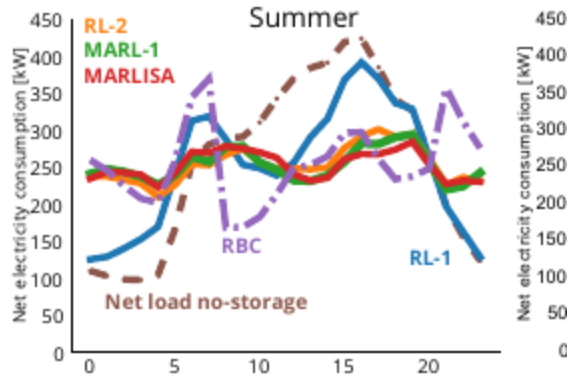


Fig. 7 (summer) · **Vázquez-Canteli et al., BuildSys 2020** · doi:10.1145/3408308.3427604

TAKEAWAY

MARLISA: nine building agents, each sharing only its **predicted next-step consumption** with the next agent, learn within a simulated year to flatten the district curve — **35% less ramping, 15% lower daily peaks** than a hand-tuned rule-based controller (which just shifts the peak to night).[54]

REAL-WORLD EVIDENCE · E-ENERGY 2013

SPOT: control comfort itself, not temperature

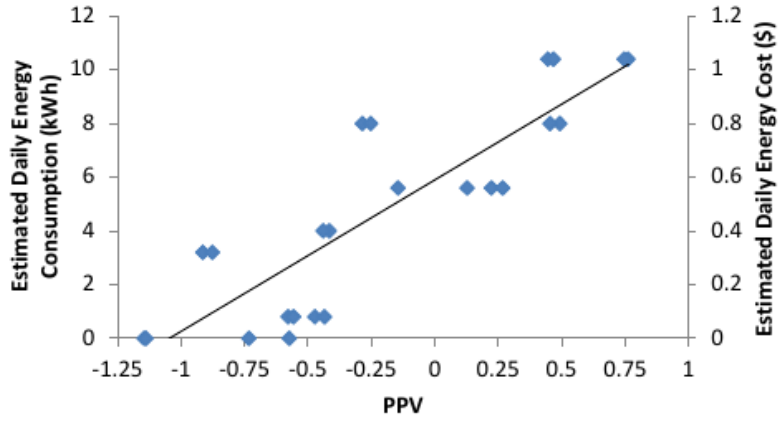


Fig. 10 · **Gao & Keshav, e-Energy 2013** · doi:10.1145/2487166.2487193

TAKEAWAY

Senses all six PMV inputs live and holds a personalized comfort level (SPOT's “PPV”), not a temperature — relaxing the target by 0.25 comfort units saves ~1.5 kWh of a ~6 kWh/day total.[55]

Comfort becomes a **control variable with a price** — the quantified version of "every degree counts."

THE PROBLEM

One sensor, three names — your app breaks

The same sensor — a zone temperature — in three vendors' languages



a large building has 1,500–2,500 sensor points; labels are typed by installers, per project

an app written for one building silently breaks on the next — this is why building software doesn't scale

TAKEAWAY

Point labels are typed by installers, per project, per vendor. Tested on three real buildings, the best existing schema could express only **54% of the tags** vendors actually use.[56]

Brick: describe relationships, not vendor names

The fix: describe the building as a graph of types and relationships (Brick)



THE SAME RULE, WRITTEN ONCE AGAINST THE GRAPH

```
for every Zone: temp = its point of type Zone_Temperature
status = the status point of the AHU feeding it (follow isFedBy)
IF temp > 26 °C AND status == OFF → alert
```

✓ no building-specific names — types and relationships only, so it runs on any Brick building

Mortar then lets you run exactly such queries across 100+ real instrumented buildings

TAKEAWAY

Brick replaces names with **types and relationships** — the rule asks for "the AHU feeding this zone," so it can run across Brick-described buildings.[37] Mortar hosts 100+ such buildings for testing.[57]

The most-cited BuildSys result is an occupancy sensor

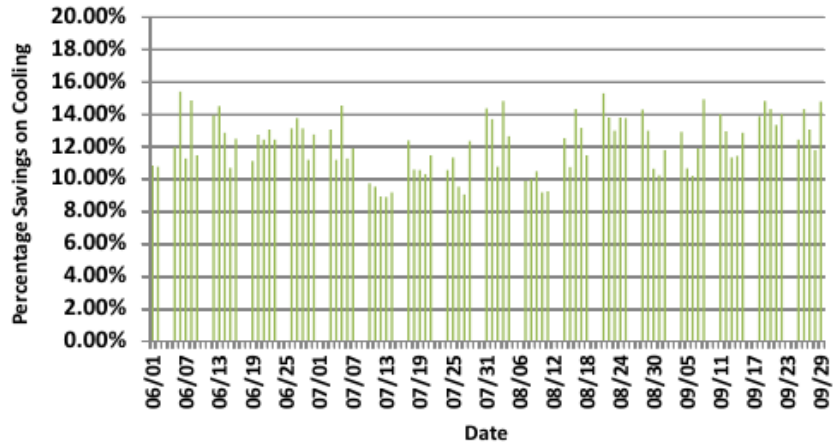


Fig. 6 · **Agarwal et al., BuildSys 2010** · doi:10.1145/1878431.1878433

TAKEAWAY

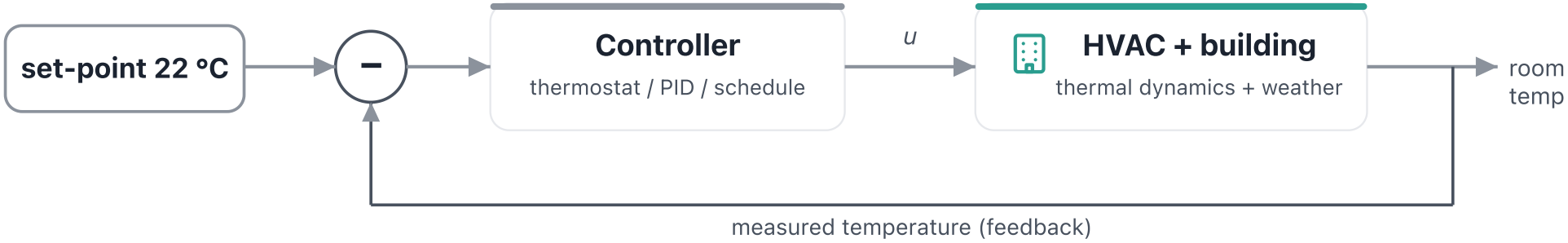
A summer of daily savings from occupancy-driven setback in UCSD's CSE building model: **10–15% every single day**. At >400 kW of HVAC load, 15% ≈ 480 kWh per 8-hour day.[58]

Occupancy-sensing thermostats in homes: ~28% HVAC savings from ~$25 of sensors.[59, 60]

THE STARTING POINT

For decades: schedules, thermostats, PID

Conventional control: react to the error between measured and target temperature



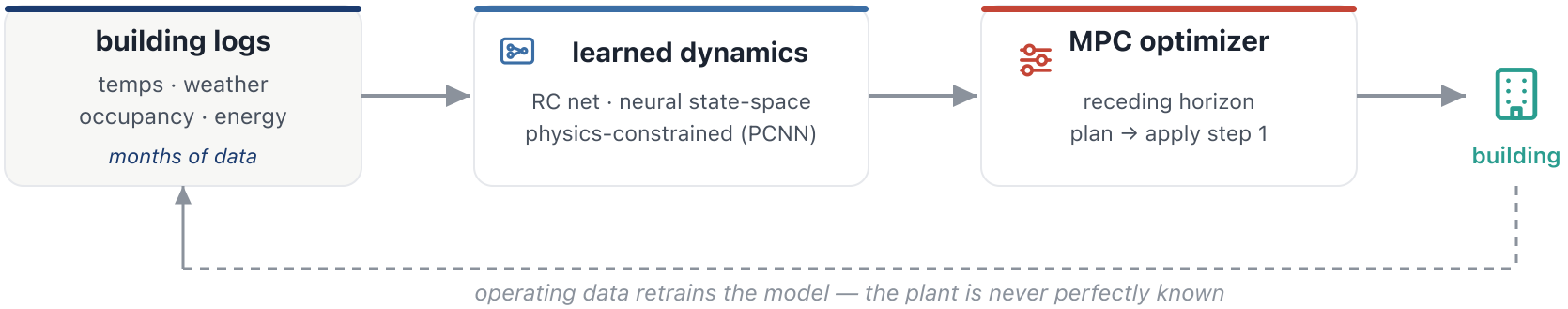
Acts only on the current error — no weather forecast, no look-ahead, no price signal.

TAKEAWAY

Simple, robust, and in almost every building. But it reacts to the **current** error only — it cannot pre-cool for a forecast heatwave or shift load off a price peak. Adding occupancy sensing to a fixed schedule was the first data-driven win.[59]

Learn the dynamics, then plan

Learn the model MPC needs from data, instead of hand-calibrating physics



TAKEAWAY

Replace hand-built physics with a model learned from logs — an RC network, or a physics-constrained neural net that stays stable and interpretable — then run MPC on top. A cheaper variant trains a network to imitate a slow MPC and run in milliseconds. Both attack the calibration cost that keeps MPC from scaling.[38, 61, 62, 63]

WHICH BOX FOR WHICH JOB?

Match the model to the job — or mix them

White → design time

The building doesn't exist yet, so only physics can answer — but it's slow and needs **calibration** once built.[64]

Grey → control

MPC re-solves every 15 minutes; a 2-parameter circuit is fast, stable, and interpretable.[38]

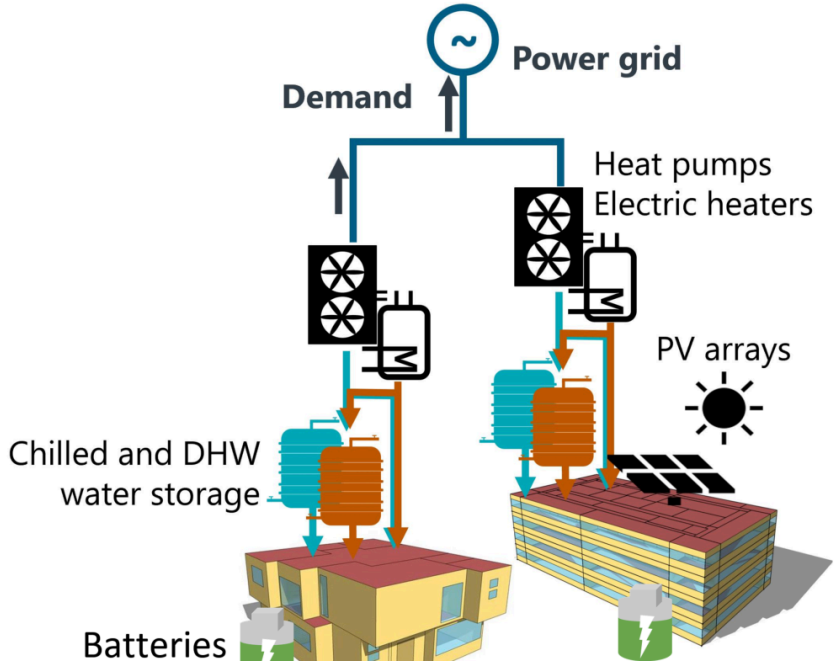
Black → forecasting

With years of meter history, let the data speak — as in the forecasting arc we just saw.

TAKEAWAY

And the frontier **mixes them**: put the physics **inside** the network — physics-informed ML (PIML), physics-constrained nets (PCNN) → ~15% lower error, longer horizons, less data — and it can't predict the physically impossible.[61, 62]

Benchmarks for building control



What CityLearn simulates: a grid-interactive building (PV, storage, heat pumps)[53]

CityLearn

Gym env for multi-building demand response; annual challenge. [53]

BOPTEST

Modelica emulators + KPI suite; MPC vs RL.[65]

Sinergym

EnergyPlus wrapped as a Gym environment.[66]

TAKEAWAY

As NILMTK did for disaggregation, shared environments let controllers be compared on equal terms.

Gnu-RL: competent on day one, no simulator needed

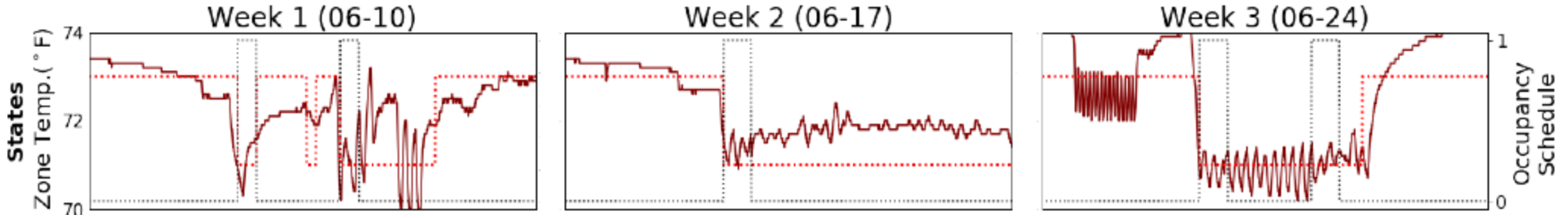
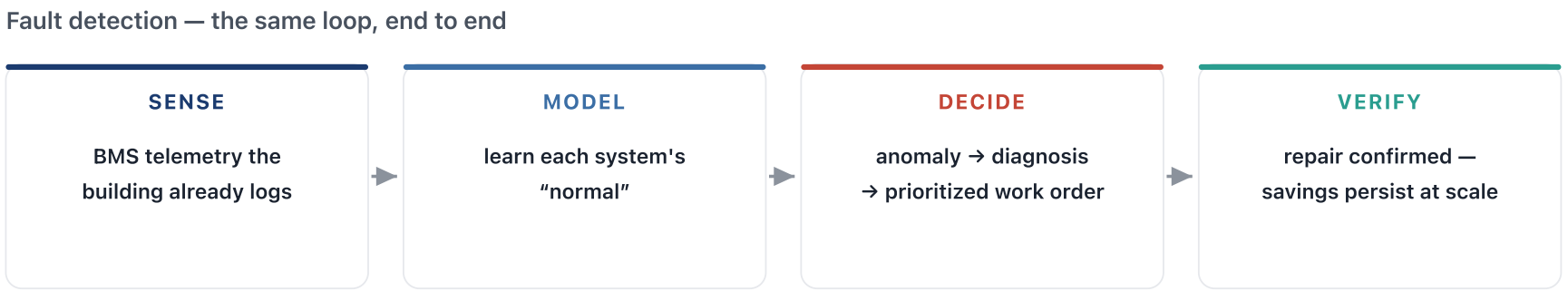


Fig. 8 (zone temperature, weeks 1→3) · **Chen et al., BuildSys 2019** · doi:10.1145/3360322.3360849

TAKEAWAY

The policy **is** a differentiable MPC — pre-trained by **imitating the existing controller's logs**, then improved online with PPO. In a real CMU conference room it cut cooling demand **16.7%** in three weeks while tracking the set-point far better (RMSE 1.02 °F vs 2.4 °F) — no simulator was ever built.[67]

Fault detection, end to end



swap the application, keep the loop — this is the transferable skill of the field

TAKEAWAY

Fourth pass — no new sensors, no control authority, and still the biggest verified savings pool in commercial buildings. The loop is the constant; the application is the variable.

STANDING ON SHOULDERS

Landmark work the field is built on

NILMTK — e-Energy Test-of-Time '25
open NILM benchmarking
Batra+ '14

SmartCharge — e-Energy Test-of-Time '24
home storage cuts the bill
Mishra, Shenoy+ '12

Synergy occupancy
the most-cited BuildSys paper
Agarwal, Gupta+ '10

Brick
a unified building metadata schema
Balaji, Srivastava+ '16

Neural NILM
deep nets for disaggregation
Kelly & Knottenbelt '15

Private Memoirs
what a smart meter reveals
Molina-Markham, Shenoy+ '10

Smart Thermostat
occupancy-driven HVAC, ~28% saved
Lu, Whitehouse+ '10

sMAP → BOSS → Mortar
portable building software
Culler, Fierro+ '10–'18

Gnu-RL
RL that controls a real room, day one
Chen, Bergés+ '19

REDD · UK-DALE · ACN · BDG2
the datasets the field runs on
2011–2020

TAKEAWAY

Open datasets, benchmarks, and systems — largely from **BuildSys** and **e-Energy**.

IF YOU READ FIVE THINGS

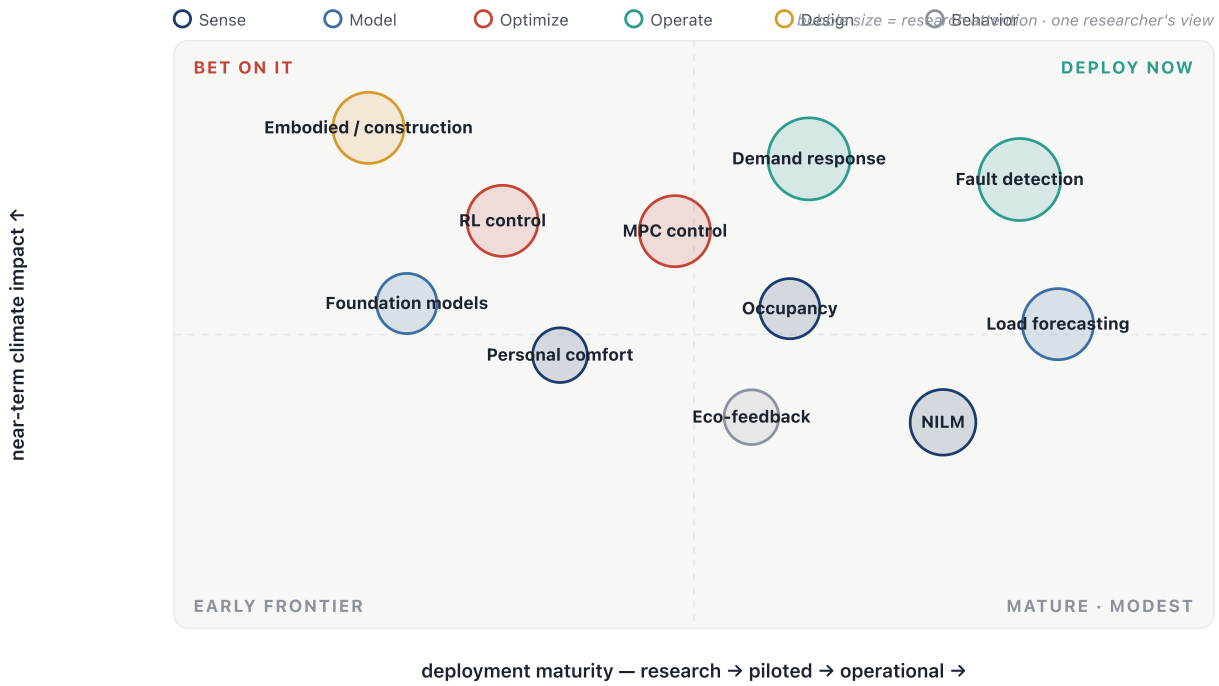
Surveys worth reading

- **The field in one paper** — Tackling Climate Change with ML, Buildings & Cities[68]; end-to-end ML-for-buildings survey[69]
- **Control** — RL reviews[40, 70], why RL isn't deployed yet[41], the MPC tutorial-review[39]
- **NILM** — deep-learning review[17], evaluation pitfalls[18]
- **Data scarcity** — transfer learning for smart buildings[71]
- **Operations** — AI for fault detection[72]

Hands-on: CCAI tutorial notebooks — CityLearn, BOPTEST, load forecasting, smart-meter analytics — github.com/climatechange-ai-tutorials

SUMMARY VIEW

Which problems matter most



TAKEAWAY

Fault detection, control, and demand response can be deployed now and save real energy. Embodied carbon is high-potential but less mature. Disaggregation and eco-feedback are well studied but show smaller measured savings.

References (1/6)

[1] International Energy Agency. **Buildings — energy system overview**. IEA, Paris, 2024. link

[2] Copernicus Climate Change Service (C3S/ECMWF). **Surface air temperature anomaly, June 2025**. Copernicus C3S, 2025. link

[3] International Energy Agency. **The Future of Cooling: Opportunities for energy-efficient air conditioning**. IEA, Paris, 2018. link

[4] Ember. **Heat and power: impacts of the 2025 heatwave in Europe**. Ember, 2025. link

[5] U.S. Energy Information Administration. **Use of energy in homes (RECS 2020)**. EIA, 2020. link

[6] Eurostat. **Energy use in EU households down second year in a row**. Eurostat, 2025. link

[7] G. o. I. Bureau of Energy Efficiency. **Residential Energy Consumption Patterns and Appliance Ownership in India: Insights from a 2024 Household Survey**. 2025. link

[8] S. National Climate Change Secretariat. **Buildings: Singapore's Climate Action**. 2026. link

[9] K. C. Armel, A. Gupta, G. Shrimali, A. Albert. **Is disaggregation the holy grail of energy efficiency? The case of electricity**. Energy Policy, 2013. doi:10.1016/j.enpol.2012.08.062

[10] S. Darby. **The effectiveness of feedback on energy consumption**. 2006. link

[11] T. W. Hnat et al. **The hitchhiker's guide to successful residential sensing deployments**. ACM SenSys, 2011. doi:10.1145/2070942.2070966

[12] N. Batra, M. Gulati, A. Singh, M. B. Srivastava. **It's different: Insights into home energy consumption in India**. ACM BuildSys, 2013. doi:10.1145/2528282.2528293

References (2/6)

[13] G. W. Hart. **Nonintrusive appliance load monitoring**. Proceedings of the IEEE, 1992. doi:10.1109/5.192069

[14] J. Z. Kolter, T. Jaakkola. **Approximate inference in additive factorial HMMs with application to energy disaggregation**. AISTATS, 2012.

[15] J. Kelly, W. Knottenbelt. **Neural NILM: Deep neural networks applied to energy disaggregation**. ACM BuildSys, 2015. doi:10.1145/2821650.2821672

[16] C. Zhang et al. **Sequence-to-point learning with neural networks for non-intrusive load monitoring**. AAAI, 2018. doi:10.1609/aaai.v32i1.11873

[17] P. Huber, A. Calatroni, A. Rumsch, A. Paice. **Review on deep neural networks applied to low-frequency NILM**. Energies, 2021. doi:10.3390/en14092390

[18] L. Pereira, N. Nunes. **Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools — a review**. WIREs Data Mining and Knowledge Discovery, 2018. doi:10.1002/widm.1265

[19] N. Batra et al. **NILMTK: An open source toolkit for non-intrusive load monitoring**. ACM e-Energy, 2014. doi:10.1145/2602044.2602051

[20] N. Batra et al. **Towards reproducible state-of-the-art energy disaggregation**. ACM BuildSys, 2019. doi:10.1145/3360322.3360844

[21] N. Batra, A. Singh, K. Whitehouse. **If you measure it, can you improve it? Exploring the value of energy disaggregation**. ACM BuildSys, 2015. doi:10.1145/2821650.2821660

[22] A. Molina-Markham et al. **Private memoirs of a smart meter**. ACM BuildSys, 2010. doi:10.1145/1878431.1878446

[23] P. Palensky, D. Dietrich. **Demand side management: Demand response, intelligent energy systems, and smart loads**. IEEE Transactions on Industrial Informatics, 2011. doi:10.1109/TII.2011.2158841

[24] M. Fleschutz et al. **The effect of price-based demand response on carbon emissions in European electricity markets: The importance of adequate carbon prices**. Applied Energy, 2021. doi:10.1016/j.apenergy.2021.117040

References (3/6)

[25] K. Amasyali, N. M. El-Gohary. **A review of data-driven building energy consumption prediction studies**. Renewable and Sustainable Energy Reviews, 2018. doi:10.1016/j.rser.2017.04.095

[26] C. Miller et al. **The ASHRAE Great Energy Predictor III competition: Overview and results**. Science and Technology for the Built Environment, 2020. doi:10.1080/23744731.2020.1795514

[27] Andreas Veit et al. **Household electricity demand forecasting: benchmarking state-of-the-art methods**. ACM e-Energy, 2014. doi:10.1145/2602044.2602082

[28] Guy R. Newsham, Benjamin J. Birt. **Building-level occupancy data to improve ARIMA-based electricity use forecasts**. ACM BuildSys, 2010. doi:10.1145/1878431.1878435

[29] P. Emami, A. Sahu, P. Graf. **BuildingsBench: A large-scale dataset of 900K buildings and benchmark for short-term load forecasting**. NeurIPS Datasets and Benchmarks, 2023. link

[30] A. Mishra et al. **SmartCharge: Cutting the electricity bill in smart homes with energy storage**. ACM e-Energy, 2012. doi:10.1145/2208828.2208857

[31] O. Ardakanian, C. Rosenberg, S. Keshav. **Distributed control of electric vehicle charging**. ACM e-Energy, 2013. doi:10.1145/2487166.2487178

[32] E. C. Kara et al. **Quantifying flexibility of residential thermostatically controlled loads for demand response**. ACM BuildSys, 2014. doi:10.1145/2674061.2674082

[33] U.S. Department of Energy. **DOE's National Roadmap for Grid-Interactive Efficient Buildings**. U.S. Department of Energy, 2021. link

[34] P. O. Fanger. **Thermal Comfort: Analysis and Applications in Environmental Engineering**. 1970.

[35] T. Cheung et al. **Analysis of the accuracy on PMV–PPD model using the ASHRAE Global Thermal Comfort Database II**. Building and Environment, 2019. doi:10.1016/j.buildenv.2019.01.055

[36] R. J. de Dear, G. S. Brager. **Thermal comfort in naturally ventilated buildings: Revisions to ASHRAE Standard 55**. Energy and Buildings, 2002. doi:10.1016/S0378-7788(02)00005-1

References (4/6)

[37] B. Balaji et al. **Brick: Towards a unified metadata schema for buildings**. ACM BuildSys, 2016. doi:10.1145/2993422.2993577

[38] P. Bacher, H. Madsen. **Identifying suitable models for the heat dynamics of buildings**. Energy and Buildings, 2011. doi:10.1016/j.enbuild.2011.02.005

[39] J. Drgoňa et al. **All you need to know about model predictive control for buildings**. Annual Reviews in Control, 2020. doi:10.1016/j.arcontrol.2020.09.001

[40] J. R. Vázquez-Canteli, Z. Nagy. **Reinforcement learning for demand response: A review of algorithms and modeling techniques**. Applied Energy, 2019. doi:10.1016/j.apenergy.2018.11.002

[41] Z. Nagy et al. **Ten questions concerning reinforcement learning for building energy management**. Building and Environment, 2023. doi:10.1016/j.buildenv.2023.110435

[42] M. Röck et al. **Embodied GHG emissions of buildings — The hidden challenge for effective climate change mitigation**. Applied Energy, 2020. doi:10.1016/j.apenergy.2019.114107

[43] P. Westermann, R. Evins. **Surrogate modelling for sustainable building design — A review**. Energy and Buildings, 2019. doi:10.1016/j.enbuild.2019.05.057

[44] B. Ekici, C. Cubukcuoglu, M. Turrin, I. S. Sariyildiz. **Performative computational architecture using swarm and evolutionary optimisation: A review**. Building and Environment, 2019. doi:10.1016/j.buildenv.2018.10.023

[45] A. Pertzborn, D. A. Veronica. **Baseline Control Systems in the Intelligent Building Agents Laboratory**. 2021. doi:10.6028/NIST.TN.2178

[46] Zachary J. Lee, Tongxin Li, Steven H. Low. **ACN-Data: Analysis and Applications of an Open EV Charging Dataset**. ACM e-Energy, 2019. doi:10.1145/3307772.3328313

[47] Diptyaroop Maji, Prashant J. Shenoy, Ramesh K. Sitaraman. **CarbonCast: multi-day forecasting of grid carbon intensity**. ACM BuildSys, 2022. doi:10.1145/3563357.3564079

[48] F. Burlig et al. **Machine learning from schools about energy efficiency**. Journal of the Association of Environmental and Resource Economists, 2020. doi:10.1086/710606

References (5/6)

[49] H. Chao. **Demand response in wholesale electricity markets: the choice of customer baseline**. Journal of Regulatory Economics, 2011. doi:10.1007/s11149-010-9135-y

[50] K. Kok, S. Widergren. **A society of devices: Integrating intelligent distributed resources with transactive energy**. IEEE Power and Energy Magazine, 2016. doi:10.1109/MPE.2016.2524962

[51] D. S. Callaway. **Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy**. Energy Conversion and Management, 2009. doi:10.1016/j.enconman.2008.12.012

[52] Z. Wen, D. O'Neill, H. Maei. **Optimal demand response using device-based reinforcement learning**. IEEE Transactions on Smart Grid, 2015. doi:10.1109/TSG.2015.2396993

[53] J. R. Vázquez-Canteli, J. Kämpf, G. Henze, Z. Nagy. **CityLearn v1.0: An OpenAI Gym environment for demand response with deep reinforcement learning**. ACM BuildSys, 2019. doi:10.1145/3360322.3360998

[54] José R. Vázquez-Canteli, Gregor Henze, Zoltán Nagy. **MARLISA: Multi-Agent Reinforcement Learning with Iterative Sequential Action Selection for Load Shaping of Grid-Interactive Connected Buildings**. ACM BuildSys, 2020. doi:10.1145/3408308.3427604

[55] Peter Xiang Gao, Srinivasan Keshav. **SPOT: a smart personalized office thermal control system**. ACM e-Energy, 2013. doi:10.1145/2487166.2487193

[56] A. Bhattacharya, J. Ploennigs, D. Culler. **Short Paper: Analyzing Metadata Schemas for Buildings: The Good, the Bad, and the Ugly**. ACM BuildSys, 2015. doi:10.1145/2821650.2821669

[57] G. Fierro et al. **Mortar: An open testbed for portable building analytics**. ACM BuildSys, 2018. doi:10.1145/3276774.3276793

[58] Yuvraj Agarwal et al. **Occupancy-driven energy management for smart building automation**. ACM BuildSys, 2010. doi:10.1145/1878431.1878433

[59] J. Lu et al. **The smart thermostat: Using occupancy sensors to save energy in homes**. ACM SenSys, 2010. doi:10.1145/1869983.1870005

[60] Z. Chen, C. Jiang, L. Xie. **Building occupancy estimation and detection: A review**. Energy and Buildings, 2018. doi:10.1016/j.enbuild.2018.03.084

References (6/6)

[61] L. Di Natale, B. Svetozarevic, P. Heer, C. N. Jones. **Physically consistent neural networks for building thermal modeling: Theory and analysis**. Applied Energy, 2022. doi:10.1016/j.apenergy.2022.119806

[62] J. Drgoňa, A. R. Tuor, V. Chandan, D. L. Vrabie. **Physics-constrained deep learning of multi-zone building thermal dynamics**. Energy and Buildings, 2021. doi:10.1016/j.enbuild.2021.110992

[63] J. Drgoňa, D. Picard, M. Kvasnica, L. Helsen. **Approximate model predictive building control via machine learning**. Applied Energy, 2018. doi:10.1016/j.apenergy.2018.02.156

[64] M. C. Kennedy, A. O'Hagan. **Bayesian calibration of computer models**. Journal of the Royal Statistical Society: Series B, 2001. doi:10.1111/1467-9868.00294

[65] D. Blum et al. **Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings**. Journal of Building Performance Simulation, 2021. doi:10.1080/19401493.2021.1986574

[66] J. Jiménez-Raboso et al. **Sinergym: A building simulation and control framework for training reinforcement learning agents**. ACM BuildSys, 2021. doi:10.1145/3486611.3488729

[67] Bingqing Chen, Zicheng Cai, Mario Bergés. **Gnu-RL: A Precocial Reinforcement Learning Solution for Building HVAC Control Using a Differentiable MPC Policy**. ACM BuildSys, 2019. doi:10.1145/3360322.3360849

[68] D. Rolnick et al. **Tackling climate change with machine learning**. ACM Computing Surveys, 2022. doi:10.1145/3485128

[69] H. P. Das et al. **Machine learning for smart and energy-efficient buildings**. Environmental Data Science, 2024. doi:10.1017/eds.2023.43

[70] Z. Wang, T. Hong. **Reinforcement learning for building controls: The opportunities and challenges**. Applied Energy, 2020. doi:10.1016/j.apenergy.2020.115036

[71] G. Pinto et al. **Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives**. Advances in Applied Energy, 2022. doi:10.1016/j.adapen.2022.100084

[72] Y. Zhao, T. Li, X. Zhang, C. Zhang. **Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future**. Renewable and Sustainable Energy Reviews, 2019. doi:10.1016/j.rser.2019.04.021