ACM e-Energy 2026, Banff, Canada

Energy Injection Identification enabled Disaggregation with Deep Multi-Task Learning

Xudong Wang[1,*], Guoming Tang[2,†], Junyu Xue[3], Srinivasan Keshav[4], Tongxin Li[1], Chris Ding[1]

[1]The Chinese University of Hong Kong, Shenzhen

[2]The Hong Kong University of Science and Technology (Guangzhou)

[3]Southern University of Science and Technology

[4]University of Cambridge

Jun 24, 2026

[†] Corresponding author [*] Work done while visiting at SusCom Lab, HKUST(GZ).

What is Non-Intrusive Load Monitoring (NILM)?

Definition

Given only the **whole-home smart-meter sequence**, NILM aims to infer the power consumption and/or operating state of individual appliances, **without installing appliance-level sub-meters**.

Classical observation model

[Hart'1992; Kolter & Jaakkola'2012]

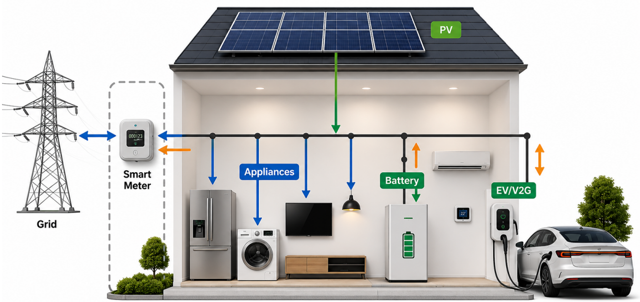
$$\underbrace{y_t}_{\text{aggregate meter}} = \sum_{k=1}^{K} \underbrace{s_{k,t} p_{k,t}}_{\text{appliance } k} + \underbrace{\epsilon_t}_{\text{background/noise}}$$

$$s_{k,t} \in \{0,1\}, \qquad p_{k,t} \geq 0.$$

- y_t: total household power measured by the smart meter;
- $s_{k,t}$: ON/OFF state of appliance k;
- $p_{k,t}$: appliance-level power contribution.

Input: aggregate meter data
$\Downarrow$
Output: appliance-level visibility

Modern Smart Homes Contain Behind-the-Meter Resources



Generation
PV and micro-inverters reduce apparent demand.

Storage
Batteries switch between load and source behavior.

Flexible mobility
EV charging and V2G create strong bidirectional flows.

Smart Homes Now Both Consume and Inject Energy

Why NILM still matters

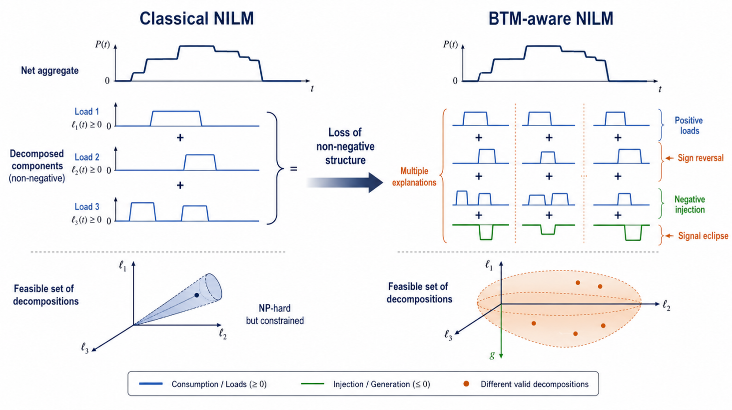
- Appliance-level visibility is still important for demand response, efficiency, and planning.
- Installing sub-meters everywhere is costly and hard to scale.

What changed

- Rooftop PV, batteries, and EV/V2G now share the same **behind-the-meter** bus.
- These resources alter the meter reading even when appliance states do not change. However, the meter sees consumption minus local injection.

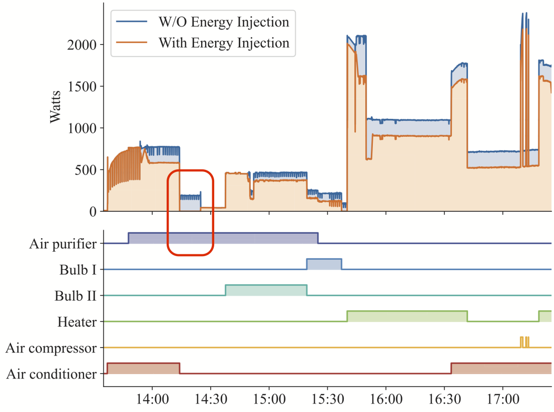
Key shift: NILM is no longer recovering appliances from aggregate demand alone; it is recovering them from **net load**.

From Constrained NILM to an Under-Determined Inverse Problem



Interpretation: classical NILM is already a NP-hard combinatorial inverse problem, but non-negativity and one-directional flow constrain the feasible set. BTM injection weakens identifiability by expanding valid consumption-generation decompositions.

Controlled Evidence: Injection Eclipses Appliance Events



Observed in lab data

- Real controllable micro-inverter: **0–500 W**.
- Net power departs from the consumption-only baseline.
- Appliance states remain active even when signatures are hidden.

This is not just sensor noise; the meter equation itself changes.

Motivation: From Classical NILM to Injection-Aware NILM

ACM e-Energy 2025, Banff, Canada

Classical NILM

$$y_t = \sum_{k=1}^{K} x_{k,t} + \epsilon_t, \qquad x_{k,t} \geq 0$$

The meter is modeled as a **non-negative mixture** of appliance loads and background consumption.

$\Longrightarrow$

BTM-Augmented NILM

$$y_t = \sum_{k=1}^{K} x_{k,t} - \underbrace{\pi_t}_{\text{BTM injection}} + \epsilon_t$$

The meter records **net power flow**: consumption can be partially or completely eclipsed by local injection.

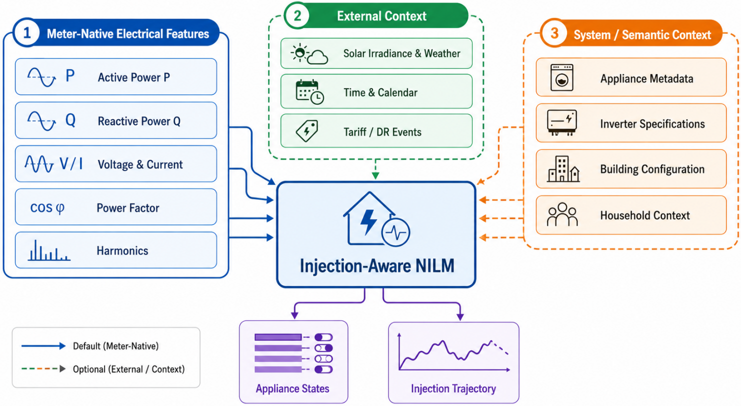
Why does the problem become harder?

- **Sign reversal:** net power may become zero or negative.
- **Signal eclipse:** injection can hide appliance switching events.
- **Underdetermination:** several decompositions can explain the same meter trace.

Our answer: DualNILM

1. **Explicit injection modeling**
   jointly predict appliance states and injection trajectories;
2. **Richer electrical evidence or features**
   combine active power P with reactive power Q or other features;
3. **Multi-scale temporal learning**
   use CNNs for local events and Transformers for long-range context.

BTM-Aware NILM Needs Richer Evidence



Design principle: BTM weakens single-channel identifiability, so the model must support richer side information when available.

Dual-Task NILM Formulation

Input

$$\mathbf{X}_t \in \mathbb{R}^{T \times F}$$

where T is the sequence length, and for the F feature channels:

- P: active power, directly perturbed by BTM injection.
- Q: reactive power, a complementary appliance-signature channel.
- Extensible channels:
  - voltage, current, PF, harmonics;
  - irradiance, weather, calendar;
  - appliance metadata, inverter specs, household context.

Output 1: appliance states

$$\hat{\mathbf{s}}(t) = \{\hat{s}_1(t), \ldots, \hat{s}_K(t)\} \in [0,1]^K.$$

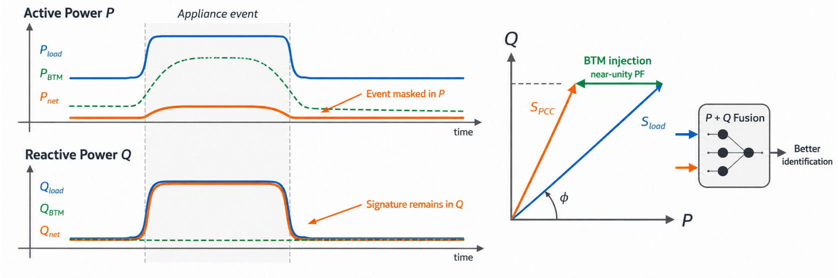
Output 2: injected energy

$$\hat{\boldsymbol{\pi}}_{t-T+1:t} \in \mathbb{R}_+^T.$$

Instead of forcing injection into the appliance set, the model explicitly estimates it as a continuous generation sequence.

$$\boxed{(\hat{\mathbf{s}}_t, \hat{\boldsymbol{\pi}}_{t-T+1:t}) = f_\theta(\mathbf{X}_{t-T+1:t}), \quad \mathbf{X} \in \mathbb{R}^{T \times F}.}$$

Why Reactive Power Is a Strong First Additional Channel

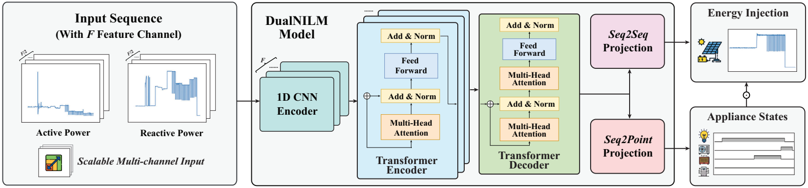


$$P_{\mathrm{net}} = P_{\mathrm{load}} - P_{\mathrm{BTM}}, \qquad Q_{\mathrm{net}} = Q_{\mathrm{load}} - Q_{\mathrm{BTM}} \approx Q_{\mathrm{load}} \quad \text{when } \mathrm{PF}_{\mathrm{BTM}} \approx 1.$$

Physical intuition. Near-unity-PF BTM injection primarily perturbs P; Appliance-specific Q signatures are often less affected.

Empirical evidence.
Q helps recover events masked in P. This is why our model uses $P + Q$ **fusion** rather than just active power. Our laboratory battery injection scenario ablation: 72.31% F1 with P only $\rightarrow$ 84.54% with $P + Q$ Fusion.

DualNILM Architecture



Feature-wise CNN

local temporal signatures from F feature channels

Transformer context

cross-channel and long-range dependencies

Dual outputs

state recognition + injection identification

$T = 300$, $F = 2$; three 1D CNN layers per channel; kernel size 5; hidden 64; Transformer Blocks: 8 heads, $d = 128$.

Cross-Task Coupling Enforces Physical Consistency

ACM e-Energy 2025, Banff, Canada

State recognition: sequence-to-point

$$\hat{s}_i = \sigma\left(\mathbf{W}_{s_i}\mathbf{z}_T + \mathbf{b}_{s_i}\right).$$

The last encoded context summarizes a local window and predicts the current ON/OFF state.

State-filtered injection

$$\hat{\mathbf{x}}_K^{\mathrm{final}} = \hat{\mathbf{x}}_K \odot \hat{\mathbf{s}}_K.$$

If the injection source is predicted inactive, its estimated injection is forced toward zero.

Injection identification: sequence-to-sequence

$$\hat{\mathbf{x}}_K = \sigma\left(\mathbf{W}_x\, \mathrm{Decoder}(\mathbf{Z}) + \mathbf{b}_x\right).$$

The decoder predicts a continuous injection trajectory.

Why this matters

- State predictions regularize injection estimation.
- Injection awareness prevents false appliance detections.
- Shared representations preserve distinct physical meanings while improving both tasks.

PV-Augmented NILM Datasets: Constructing BTM Ground Truth

Simulation pipeline

Public NILM traces: REDD / UK-DALE

↓

NSRDB solar/weather: GHI, DNI, DHI, T

↓

PV + inverter model: temperature correction

↓

Reactive power via appliance / inverter PF

↓

Benchmark: (P_{net}, Q) + states + PV label

Net meter construction:
$$P_{\text{net}}(t) = P_{\text{agg}}(t) - \min\{P_{\text{PV}}(t), P_{\text{agg}}(t)\}.$$

Simulated PV-Augmented Datasets

- **Public bases:** REDD H1–H3; UK-DALE H1–H2.
- **Locations:** Boston and London solar profiles.
- **PV scenario:** residential-scale BTM PV.
- **Targets:** microwave, fridge, dishwasher, washer, kettle.
- **Labels:** appliance states + PV injection.
- **Splits:** REDD 1-week/3-day; UK-DALE 2-week/1-week.

Why this matters: the lab setup validates hardware realism; the PV-augmented release adds household diversity, reproducibility, and a public benchmark for BTM-aware NILM.

Evaluation Datasets & Regimes, Tasks, and Metrics

ACM e-Energy 2025, Banff, Canada

Datasets: Two complementary evaluation regimes

- **Lab real injection:** Dec. 6–15, 2023; 2 s sampling; six appliances + micro-inverter; leave-one-day-out testing.
- **Synthetic PV injection:** REDD and UK-DALE with weather-driven PV profiles and public appliance labels.

Tasks

Appliances State Recognition and **BTM Injection Disaggregation**

Evaluation Metrics

Appliance states: Accuracy, Recall, Precision, **F1**.
Injection: **RMSE** and MAE.

Table 1: Statistics of the Laboratory Dataset

Appliance	Mean Power (W)	Std Dev (W)	On-time (%)
Air purifier	34.25	2.35	40.57
Heater	1345.73	476.90	28.69
Light bulb Type I	204.84	13.64	10.54
Light bulb Type II	412.99	49.39	10.59
Air compressor	1408.04	287.72	0.4885
Air conditioner	695.59	82.92	38.61
Micro-inverter	219.31	115.77	63.56

Micro-inverter's value means the power injection

Evaluation Baselines

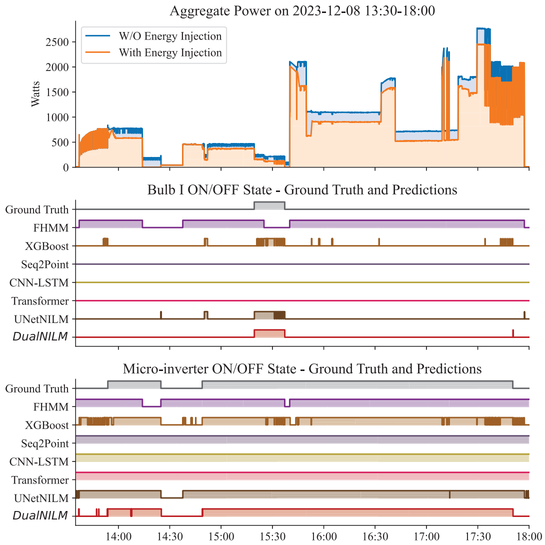
To comprehensively evaluate DualNILM, we compared representative and advanced methods across both appliance state recognition and energy injection disaggregation tasks.

Table 2: Summary of Benchmark Methods

Category	Method	Key Characteristics	Primary Strengths	Tasks*
Traditional	FHMM [Kim et al.'11]	Parallel hidden Markov chains	Interpretable state modeling	Both
	XGBoost [Chen & Guestrin'16]	Tree-based ensemble	Non-linear pattern recognition	State Recognition
	SunDance [Chen & Irwin'17]	Physical solar modeling	Black-box estimation	Disaggregation
Deep Learning	Seq2Point [Zhang et al.'18]	CNN-based architecture	Local temporal features	State Recognition
	CNN-LSTM [Kaselimi et al.'20]	Hybrid CNN-LSTM	Local-global temporal modeling	State Recognition
	Transformer [Vaswani et al.'17]	Self-attention mechanism	Long-range dependencies	State Recognition
	Seq2Seq [Du et al.'16]	Encoder-decoder LSTM	Sequence mapping	Disaggregation
	DAE [Jia et al.'19]	Denoising autoencoder	Signal reconstruction	Disaggregation
	BERT+ [Li et al.'24]	BERT blocks	Utilize extra user embeddings	Disaggregation
	UNetNILM [Faustine et al.'20]	Skip-connected encoder-decoder	Multi-task learning	Both

Tasks include State Recognition* *(appliance state recognition),* ***Disaggregation* *(energy injection disaggregation), and* ***Both*** *(both tasks).*

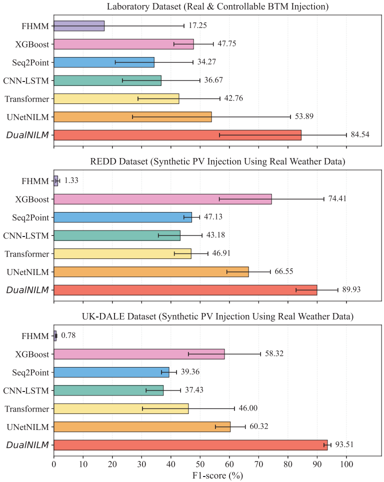
Qualitative Result: State Recognition under Real Micro-inverter Injection



What to look at

- Top: net aggregate power with injection.
- Middle/bottom: state ground truth and predictions.
- Baselines fire on injection-induced variations; DualNILM tracks states more consistently.

State Recognition Results: Dual-Task Modeling Improves F1



Laboratory real injection
Avg. 84.54%
F1, vs. best baseline 53.89%

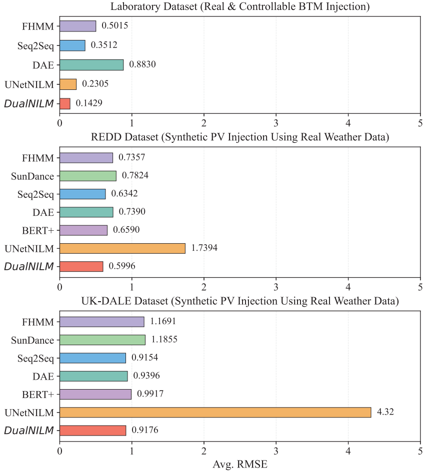
REDD synthetic PV
Avg. 89.93%
Average F1 across houses

UK-DALE synthetic PV
Avg. 93.51%
Average F1 across houses

Insight

The gain is largest exactly when injection violates the assumptions of state-only NILM.

Injection Identification Results: Competitive and Stable RMSE



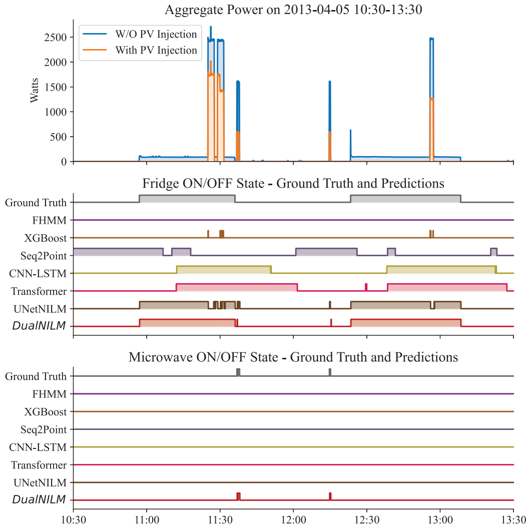
Laboratory real injection
0.1429
RMSE, best among compared methods

REDD synthetic PV
0.5996
Average RMSE across houses

UK-DALE synthetic PV
0.9176
Average RMSE, stable vs. UNetNILM instability

State recognition isolates consumption changes, which stabilizes injection estimation.

Synthetic PV Case Study: Generalization Beyond the Lab



Scenario

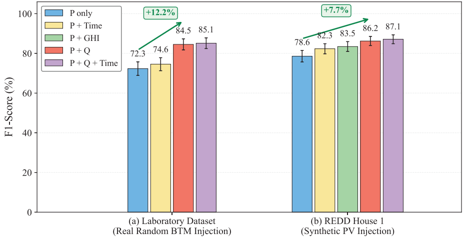
- UK-DALE House 1.
- Midday weather-driven PV injection.
- Fridge and microwave states are evaluated under distorted net power.

Observation

DualNILM remains robust when PV dynamics overlap with sparse appliance switching events.

Ablations: Joint Training and Reactive Power are Both Needed

Figure 1: Input Feature Ablation Study (Time: Time positional encoding, GHI: Irradiance related feature)



Training on UK-DALE H1	F1	RMSE
Separate state / injection	88.27%	0.9521
Sequential pipeline	90.24%	0.9453
DualNILM joint	**94.71%**	**0.9008**

Table 3: Ablation Study: Joint vs. Separate Training (UK-DALE House 1)

Feature ablation

- Lab: adding *Q* improves F1 from 72.31% to 84.54%.
- The randomized lab setup reduces time-of-day leakage.
- Therefore the gain is consistent with electrical physics, not schedule memorization.

Core lesson

Reactive power addresses feature ambiguity; joint training addresses task ambiguity.

Summary and Future Work

Main takeaways

1. BTM injection turns NILM into a bidirectional inverse problem.
2. Signal eclipse changes identifiability, not just noise level.
3. Explicit injection modeling lets consumption and generation disambiguate each other.
4. $P + Q$ is a practical core configuration, while DualNILM supports richer F-channel extensions.

Future directions

- **Broader BTM resources:** batteries, EV/V2G, smart inverters with VAR support, and hybrid DER portfolios.
- **LLM-context integration:** device metadata, tariffs, weather summaries, calendars, and inverter logs as structured priors.
- **Deployment:** uncertainty estimation, household personalization, privacy, and physics-constrained calibration.

Main message: robust NILM in renewable-rich homes requires identifying injected energy, not merely filtering it out.

Acknowledgements






Xudong Wang CUHK-Shenzhen

Guoming Tang HKUST(GZ)

Junyu Xue SUSTech

S. Keshav Cambridge

Tongxin Li CUHK-Shenzhen

Chris Ding CUHK-Shenzhen

Thank you to all collaborators.

We thank the SusCom Lab of HKUST(GZ), School of Data Science (SDS) of CUHK-Shenzhen, and all collaborators for valuable discussions, experiment support, and feedback on BTM-aware NILM.

Open Resources and Questions

GitHub

PV-augmented NILM datasets and simulation suites

https://github.com/MathAdventurer/PV-Augmented-NILM-Datasets

Hugging Face

PV-augmented NILM datasets

https://huggingface.co/datasets/wangxudong/PV-Augmented-NILM-Datasets

Take Away

DualNILM =
(1) Injection-aware task design for NILM with BTM,
(2) $P + Q$ complementary sensing with the scalability for rich features,
(3) Public PV-augmented NILM datasets and the simulation suites.

Thank you!

Questions?

Email: xudongwang@link.cuhk.edu.cn & guomingtang@hkust-gz.edu.cn