Graph-based reinforcement learning → transferable HVAC control

HVAC-GRACE: Transferable Building Control via Graph RNNs

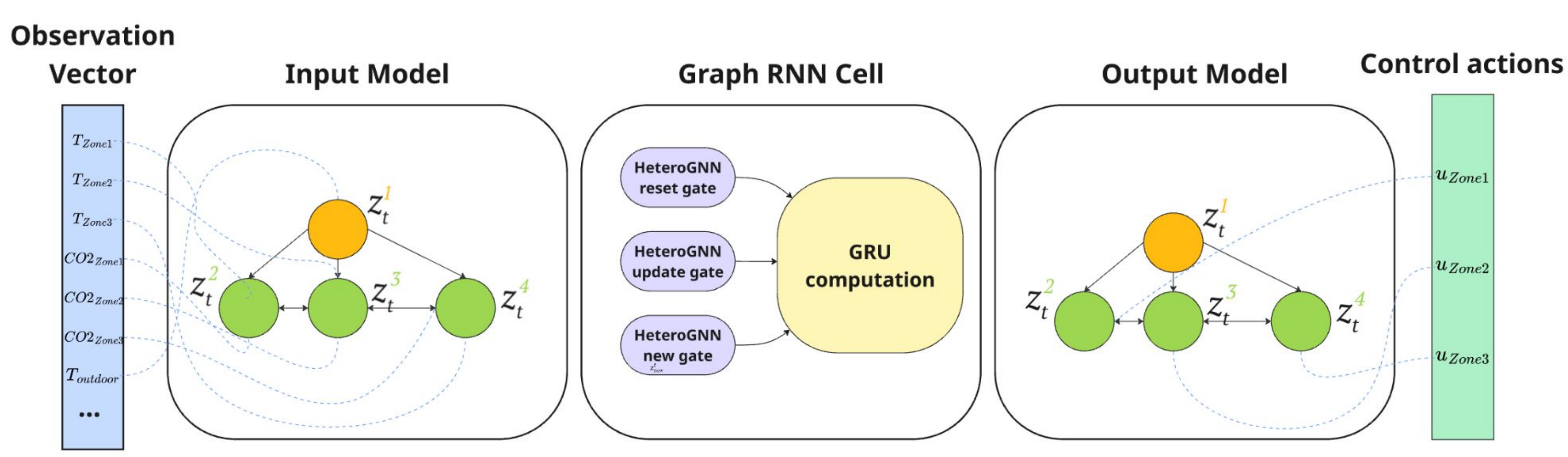
Anaïs Berkes Donna Vakalis David Rolnick Yoshua Bengio

- Buildings consume 40% of global energy, with heating/cooling responsible for ½
- better controls are not transferable → every building needs custom control

- **Solution:** Structure data from buildings as heterogeneous graphs!
- **Why?** this learns topology-agnostic functions, for zero-shot transfer
- **HVAC-GRACE:** first graph-based RL framework, learns for zero-shot transfer with spatial message-passing integrated into temporal GRUs + use RL (PPO) to improve the policy
- **Work-in-progress:** first steps established framework for building-building transfer; currently scaling to multi-building (~50k) training datasets

Algorithm 1 HVAC-GRACE Training Algorithm

Input: Building epJSON file, training episodes N
$\mathcal{G} = (\mathcal{V}, \mathcal{E}) \leftarrow$ ConstructGraph(epJSON)
Initialise policy π_θ, critic V_ϕ, Graph RNN states $\{h_v^0\}_{v\in\mathcal{V}}$
for episode $= 1$ to N **do**
  Reset environment, initialise s_0, $\{h_v^0\}_{v\in\mathcal{V}}$
  for timestep $t = 0$ to $T-1$ **do**
    Stage 1: Input Processing
    parsed_obs $\leftarrow$ ParseObservation(s_t)
    $x^t \leftarrow$ InputMLP(parsed_obs)
    Stage 2: Graph RNN Processing
    $\{h_v^t\}_{v\in\mathcal{V}} \leftarrow$ GraphRNNCell(x^t, $\mathcal{E}$, $\{h_v^{t-1}\}_{v\in\mathcal{V}}$)
    Stage 3: Action Generation
    for each conditioned zone $v \in \mathcal{V}_c$ **do**
      $\mu_v, \log\sigma_v \leftarrow f_{\text{policy}}(h_v^t)$
      Sample $a_v \sim \mathcal{N}(\mu_v, \exp(\log\sigma_v))$
    end for
    Execute actions $\{a_v\}_{v\in\mathcal{V}_c}$, observe s_{t+1}, r_t
  end for
  Update policy π_θ and critic V_ϕ using PPO
end for



HVAC-GRACE METHOD

- **Stage 1**: Parse building observations into heterogeneous graph (conditioned zones, unconditioned zones, outdoor environment)
- **Stage 2**: Use a modified GRU where the standard gate operations are replaced with graph neural networks, allowing zones to coordinate spatially while maintaining temporal memory of building dynamics.
- **Stage 3**: During training, a reinforcement learning agent learns type-specific control policies by interacting with building simulations (in EnergyPlus), optimizing for energy efficiency and comfort.
- **Stage 4:** Generate control actions for conditioned zones

WORK IN PROGRESS

- Tested on DOE Small Hotel (6 zones, 1 conditioned = 14% connectivity)
- Training/testing: Alabama → Montreal weather
- Key insight: this method requires sufficient connectivity between conditioned zones (~50%) to maintain gradient flow during learning—sparse building topologies cause transfer learning to fail.
- Roadmap for our current work: curating well-connected and domain-randomized building datasets (~50k buildings) to validate the full potential of this approach

THE GRAPH ADVANTAGE

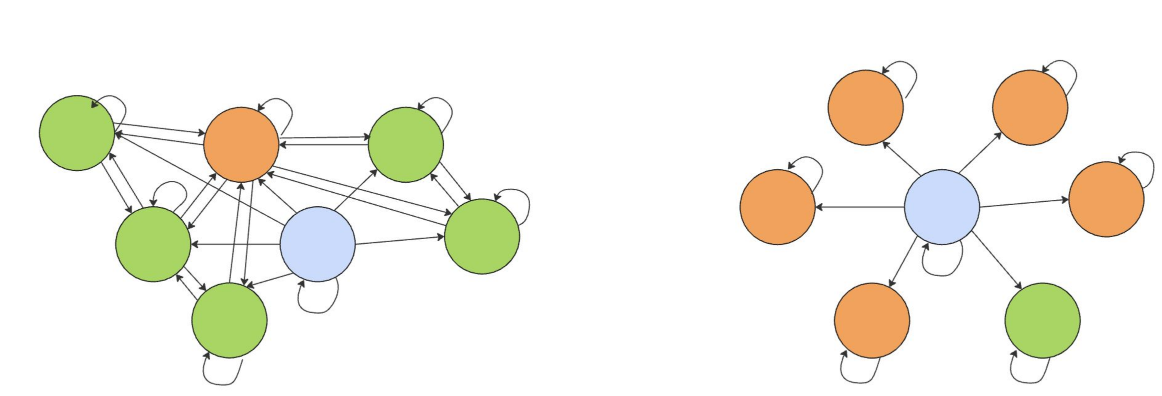
Topology-Agnostic Design: HVAC-GRACE learns type-specific functions that work on semantic node types instead of requiring fixed-dimensional input vectors, enabling zero-shot transfer to new buildings without retraining.

Traditional fixed-vector approaches:

- Training on more buildings doesn't help much b/c each building needs its own model
- Can't effectively share knowledge between different building sizes

HVAC-GRACE:

- Every new building contributes to the same type-specific functions
- Knowledge accumulates across all buildings in the training set



Gradient flow in well-connected (left) vs. sparse (right) building topologies. Green nodes (HVAC conditioned zones) receive direct policy gradients; orange nodes (unconditioned zones) depend on message passing. Sparse connectivity disrupts learning.