Structure-Preserving Model Reduction of Nonlinear Building Thermal Models ⋆

Kun Deng [a], Siddharth Goyal [b], Prabir Barooah [b], Prashant G. Mehta [a]

[a] *Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

[b] *Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611 USA*

Abstract

This paper proposes an aggregation-based model reduction method for nonlinear models of multi-zone building thermal dynamics. The full-order model, which is already a lumped-parameter approximation, quickly grows in state space dimension as the number of zones increases. An advantage of the proposed method, apart from being applicable to the nonlinear thermal models, is that the reduced model obtained has the same structure and physical intuition as the original model. This makes the reduced model useful not only for control design and analysis but also for building design iterations. The key to the methodology is an analogy between a continuous-time Markov chain and the linear part of the thermal dynamics. A recently developed aggregation-based method of Markov chains is employed to aggregate the large state space of the full-order model into a smaller one. Simulations are provided to illustrate tradeoffs between prediction error and computation time.

Key words: Model reduction; Structure preserving; Model-based control; Control oriented models; Markov models.

1 Introduction

In 2009, commercial and residential buildings accounted for 42% of the total energy usage and 75% of total electricity consumption in the United States [43, Table A.2]. Among all energy consumers of buildings, Heating, Ventilation, and Air Conditioning (HVAC) account for a large share. A large fraction of the energy delivered to buildings is wasted because of inefficient building technologies [3,39]. Interests in methods for controlling building HVAC systems to reduce their energy usage or cost have been on the increase in recent years; particularly in model-based approaches such as Model Predictive Control (MPC) [20, 29, 30, 33, 47].

Accurate models of temperature evolution in a building are required for real-time prediction and control, especially in model-based control methods. This paper focuses on model order reduction of buildings with multiple zones. A zone in this paper refers to a single space (room, hallway, etc.) that is serviced by a single "terminal box" with supply air diffuser and return air grilles. Fig. 1 shows a four-zone building HVAC system, where each zone refers to a single room here; terminologies and more details appear in Section 2.

The physical processes that determine thermal dynamics in buildings, which are governed by a set of coupled partial differential equations, are complex. In principle, Computational Fluid Dynamics (CFD) can be used to solve these equations. CFD models are, however, computationally intensive [10] and sensitive to boundary condition specifications [25]. Complexity issues of CFD models have led to development of simplified models in the past few decades. In this framework, the air in each zone is assumed to be well mixed with a uniform temperature. The thermal response of a zone and conduction between zones that are separated by solid surfaces (walls, floors, ceiling, windows, partitions, etc.) are modeled by capacitances and resistances, respectively. Such resistor-capacitor models have been extensively used to construct dynamic models of zone temperatures in the HVAC and building modeling literature; see, for instance [9, 17, 38, 45]. The resistances and capacitances are carefully chosen to model the combined effect of conduction between the air masses separated by the surface, as well as long wave radiation and convection between the sur-

⋆ The material in this paper was partially presented at the American Control Conference on June 30-July 2, 2010, Baltimore, Maryland, USA. The research is supported by the National Science Foundation under Grants CNS-0931416, CNS-0931885, and ECCS-0955023. Corresponding author Kun Deng. Tel.: +1 217 390 9567.

Email addresses: kundeng2@illinois.edu (Kun Deng), siddgoya@ufl.edu (Siddharth Goyal), pbarooah@ufl.edu (Prabir Barooah), mehtapg@illinois.edu (Prashant G. Mehta).

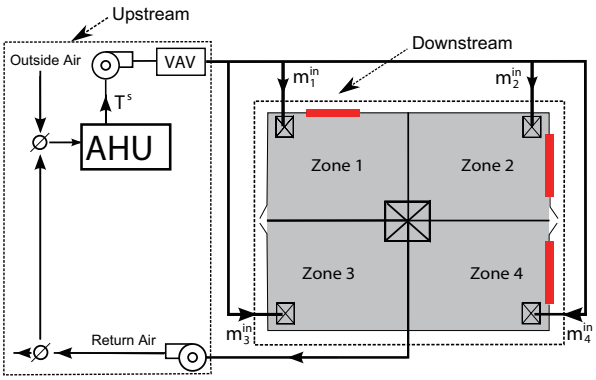


Fig. 1. The configuration of a four-zone building HVAC system.

face and the air mass in contact with it [18,32], [7, Chapters 4, 15, & 25]. In addition, there is thermal interaction between each zone and the outside due to the ventilation air that is supplied to and extracted from the zone that has to be accounted for.

A complete model of the entire multi-zone building's thermal response can then be constructed by using (i) resistor-capacitor networks for combined conduction-convection-radiation through surfaces, and (ii) heat balance equations to account for the enthalpy exchange between a zone and the outside due to the ventilation air. The resulting lumped parameter model is called the *full-order building thermal model* in this paper. This approach of constructing multi-zone thermal models have been pursued previously in [19,26,31,44,46]. The full-order model we consider here is from [19]. The ventilation heat exchange terms make the thermal dynamic model nonlinear; more modeling details appear in Section 2.

A fundamental problem with the full-order models is that they quickly explode in complexity as the number of zones increases. For example, the full-order model of the downstream part of a four-zone building, shown in Fig. 1, has 37 nodes (more details appear in Section 5). For a large commercial building with hundreds of zones, the number of nodes are of the order of several hundreds, or even thousands. This is a cause of concern for optimization-based control schemes such as MPC, particularly if the optimization is to be performed with a day-long prediction horizon to take advantage of slow thermal responses of buildings as well as daily variations in environment and energy prices. Thus, model reduction methods are required for successful implementation of advanced control schemes in realistic buildings by reducing the computational complexity. Large model complexity is also an issue even for off-line predictions during the building design phase, when a large number of parametric studies are to be performed using building energy prediction software such as EnergyPlus or DOE-2. These design iterations require yearly energy consumption predictions, which need whole-year building simulations. As a result, speedy simulation that comes from low model complexity is important in design iterations. Using a small number of zones to reduce computation time is common practice, and in fact recommended for EnergyPlus [24]. Thus, model reduction techniques can aid in the building design phase as well.

Due to the nonlinear nature of the building thermal model, the number of available techniques for model reduction is limited. Balanced truncation method for nonlinear systems has been introduced by Scherpen in [37], which uses controllability and observability energy functions of a system to balance the realization. Related methods [4, 11, 22] has also been developed for bilinear systems. These energy functions however are difficult to compute in practice. Lall *et al.* in [28] use empirical Gramians to determine the importance of a particular subspace in terms of its contribution to the input-output behavior. These empirical Gramians are calculated by simulation or experimental data generated within the system's expected operating region, in which some of the nonlinear behavior is captured by resulting Gramians. Hahn and Edgar [23] propose a hybrid method by introducing controllability and observability covariance matrices, which can be computed from data along system trajectories. The balanced truncation method is further extended in [36] by introducing the so-called extended Gramians to improve error bounds and enforce structural constraints. The recent paper [19] proposes a method for reduction of multi-zone building thermal models of the type considered here. The method in [19] is also based on balanced truncation; it applies a balanced transformation computed from the linear part of the dynamics to the nonlinear full-order model and then performs truncation. The sparsity pattern of the nonlinear terms are exploited to reduce loss of predictive ability in performing the truncation. However, the resulting states of the reduced model have no physical meaning, unlike the states of the original model that relate to temperatures of the zones and internal nodes of walls. The same is true for all model reduction methods mentioned above: the reduced order models do not retain the structure and the physical intuition of the original full-order model.

In this paper, we propose an *aggregation-based approach* that preserves the structure of the original model, that is, the reduced building thermal model is still a nonlinear RC-network. This is achieved by obtaining super-nodes via aggregation, and determining the super-capacitance for each super-node and super-resistance for each edge between two adjacent super-nodes. The aggregation-based approach proposed in this paper is based on model reduction method of Markov chains that has recently been developed in [15]. The Kullback-Leibler (KL) divergence rate is used as a "metric" to reduce Markov chains via aggregation of states in [15]. The idea of this paper is to connect the linear portion of the multi-zone thermal model to a continuous-time Markov chain, and extend the model reduction procedure for Markov chains to the nonlinear full-order building thermal model. The degree of reduction can be specified by the user, and the full-order model with n nodes can be reduced to a model with m super-nodes, with m being any integer between 1 and n. Simulations reported in this paper show that the proposed

method produces reduced-order models that well approximate the time-domain predictions of the original model. As one would expect, the prediction accuracy decreases as the specified degree of reduction increases.

There are several advantages of the model reduction method of multi-zone building thermal models proposed here compared to the existing general model-reduction methods mentioned above. Unlike the empirical Gramian based methods, we do not need simulation data from full order models to construct the reduced order model; it is obtained directly from the model description. The proposed method does not suffer from the computational difficulty of the energy function based methods. The building thermal model reduction method of [19] has the limitation that the minimum number of states in the reduced model is equal to the number of zones in the building. In contrast, any user specified reduction in the model order is possible with the proposed method. This makes the proposed method more attractive for MPC-type control schemes, and for performing off-line control design and analysis studies for a building with a large number of zones, when a large reduction in model order is called for. However, for the same (reduced) model order, it turns out that the reduced-order model obtained by the truncation-based method has a slightly smaller prediction error than that by the aggregation-based method.

The key difference of the proposed method over existing work comes from the fact that, unlike all the previously mentioned methods, the method proposed here is *structure-preserving* in the sense that that the reduced model of a nonlinear RC-network is still a nonlinear RC-network. Thus, the parameters and nodes of the reduced model retain the same physical meaning of the parameters and nodes of the original RC-network model. A number of zones can be reduced to a smaller number of "super-zones" with the proposed method. This makes the model reduction method proposed in this paper is of potential use in the design of buildings as well.

To see the use in architectural design, note that it is common in the building design stage to combine a number of zones into a large "super-zones". This is done to reduce simulation time [24]. For instance, a building with 15 zones was reduced to 7 zones for reduction in computation complexity in [5]. However, the process of combining multiple zones into single super zone is done manually; we are not aware of any formal method to perform such aggregation. Frequently, such aggregation is done in an an-doc fashion. For instance, all office spaces scattered throughout a building are combined into one zone in [27]. The method proposed in this paper provides a formal method to perform such aggregation, which benefits building design studies.

The rest of the paper is organized as follows. In Section 2, the full-order model is described and the model reduction problem is stated. In Section 3, the Markov chain analogy of the building thermal dynamics is presented. In Section 4, the aggregation-based methodology is applied to reduce the building thermal model. In Section 5, theoretical results are illustrated by numerical simulations. The conclusions appear in Section 6.

2 Full-order Building Thermal Model

A typical HVAC system consists of AHUs, supply ducts, and terminal boxes; see Fig. 1 for an example. The AHU (Air Handling Unit) supplies conditioned air (usually cold and dry) to terminal boxes at so-called leaving-air temperature and humidity. Each terminal box delivers air to one or more zones. When the box is equipped with a reheat coil (a common configuration), the supply air temperature downstream of the box can be increased beyond the AHU leaving temperature. In a VAV (Variable-Air-Volume) system, the terminal box may vary the supply air mass flow rate through dampers, but not in a CAV (constant air volume) system. A controller at each terminal box can be used to maintain the temperature of a zone at a specified value by controlling the mass flow rate of air supplied to the zone.

The dynamics of the building with its HVAC system can be divided into upstream and downstream parts (see Fig. 1). The upstream part includes the AHU dynamics and the downstream part includes the thermal dynamics of the zones. The focus of this paper is on modeling the downstream thermal dynamics. The reasons for ignoring the AHU dynamics are twofold. First, the dimension of the downstream model increases quickly with the number of zones and internal thermal nodes, while the dimension of the upstream model increases only with the number of AHUs. The later is typically small even for a large building. Second, the AHU has fast dynamics in the HVAC system, with a time constant of about a minute [8], whereas the thermal dynamics of the zones are relatively slow with time constants in tens of minutes [42] to hours [16]. As a result, the dynamics of the AHUs are replaced by static gains in this paper without significant loss of accuracy. From now on, "building thermal dynamics" would mean dynamics of the downstream part. Variations of temperature within a zone are neglected; each zone is characterized by a single temperature variable.

2.1 RC-network model of building thermal dynamics

A building thermal model is constructed by combining lumped parameter models of thermal interaction between two zones separated by a solid surface (e.g., walls, windows, ceilings, and floors). For the sake of simplicity, here we ignore the inter-zone convective heat transfer that occurs through the open doors and hallways. A lumped parameter model of combined heat flow across a surface is modeled as a simple RC-network, with current and voltage being analogous of heat flow and temperature. In this modeling framework, the capacitances are used to model the total thermal capacity of the wall, and the resistances are used to represent the total resistance that the wall offers to the flow of heat from one side to the other. In [17], Gouda *et. al.* showed that a second-order RC-network model with 3

resistors and 2 capacitors, which we will call 3R2C model, is sufficient to capture the conductive dynamic interaction between two spaces through a single wall; see Example 1 at the end of this subsection.

For a building consisting of a number of surface elements (e.g., walls, windows, ceilings, and floors), 3R2C models for surface elements can be inter-connected to obtain a RC-network model of the entire building. The resulting model can be represented as an *undirected graph* $G = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} := \{1, \ldots, n+1\}$ denotes the set of *nodes* of the graph. A node may represent a physical zone (e.g., a room, a hallway, or "the outside"), or some point inside a wall. For the sake of simplicity of the description, the nodes are assumed to be re-indexed so that the first N nodes correspond to the temperatures of zones $1, \ldots, N$; these are called the *zone nodes*. The next $(n-N)$ nodes correspond to the temperatures internal to the surfaces that appear due to the 3R2C networks; these are called the *internal nodes*. The last, i.e. $(n+1)$th node, corresponds to the outside. Each node $i \in \mathcal{V}$ corresponds to a temperature T_i and each node $i \in \mathcal{V}/\{n+1\}$ has an associated capacitance C_i. The set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ denotes the set of all *edges*. Edges represent pathways for conductive heat transports. More specifically, each edge $(i,j) \in \mathcal{E}$ represents the conductive thermal interaction between the nodes i and j and there exists a resistance $R_{ij} \in \mathbb{R}_+$ connecting them directly. Since the graph is undirected, $R_{ji} = R_{ij}$ by convention.

The states and inputs of the building thermal model are summarized below:

$$\begin{aligned}&\text{States} : T_1, \ldots, T_N, T_{N+1}, \ldots, T_n, T_{n+1}\\ &\text{Inputs} : T_o, T^s; \dot{m}_i^{in}, \dot{Q}_i^r, \dot{Q}_i^{int}, \dot{Q}_i^{ext}, i = 1, \ldots, N\end{aligned}$$

where $T_1, \ldots, T_N$ denote the space temperature of the zones, and $T_{N+1}, \ldots, T_n$ denote temperature of the points internal to the surface elements, and T_{n+1} & T_o denote the same quantity – the outside temperature (it is denoted as T_{n+1} if considered as a state, and T_o if considered as an input), T^s denotes the *supply temperature*, that is the temperature of the air supplied by the AHU, $\dot{m}_i^{in}$ denotes the mass flow rate of the supply air entering the ith zone, $\dot{Q}_i^r$ denotes the heat gain due to reheating that may occur at the VAV box of the ith zone, $\dot{Q}_i^{int}$ denotes the *internal heat gain*, i.e., the rate of heat generated by occupants, equipments, lights, etc. in the ith zone, and $\dot{Q}_i^{ext}$ denotes the *external heat gain*, i.e., the rate of solar radiation entering the ith zone.

The supplied air temperature T^s is usually constant for a VAV system, at least over short intervals of time [7]. All other inputs are time varying. In this paper, it is assumed that (i) the supply air temperature T^s is given as a constant, (ii) the (estimation of) the outside temperature T_0 and the heat gains $\dot{Q}^r, \dot{Q}^{int}, \dot{Q}^{ext}$ are available based on historical data, weather predictions, and various sensors.

The thermal dynamics of a multi-zone building, described by a graph G, are modeled by the following coupled nonlinear differential equations: For $i = 1, \ldots, n$,

$$C_i \frac{dT_i}{dt}(t) = \sum_{j \in \mathcal{N}_i} (T_j(t) - T_i(t))/R_{ij} + \dot{Q}_i(t) + \Delta H_i(t) \quad (1)$$

where $\mathcal{N}_i := \{j \in \mathcal{V} : j \neq i, (i,j) \in \mathcal{E}\}$ denotes the set of *neighbors* connecting to the node i (note that the outside node $n+1$ may belong to the set $\mathcal{N}_i$ for some node i), and the terms $\dot{Q}_i, \Delta H_i$ are described below:

- The *heat gain* term $\dot{Q}_i$ is the rate of thermal energy entering the node i from all sources other than ventilation air and conduction from neighboring nodes. It is non-zero only for zone nodes:

$$\begin{cases} \dot{Q}_i(t) = \dot{Q}_i^r(t) + \dot{Q}_i^{int}(t) + \dot{Q}_i^{ext}(t), & i = 1, \ldots, N \\ \dot{Q}_i(t) = 0, & i = N+1, \ldots, n. \end{cases}$$

- The *ventilation heat exchange* term ΔH_i is the rate of thermal energy entering the node i due to ventilation. It is non-zero only for the zone nodes:

$$\begin{cases} \Delta H_i(t) = C_{pa} \dot{m}_i^{in}(t)(T^s - T_i(t)), & i = 1, \ldots, N \\ \Delta H_i(t) = 0, & i = N+1, \ldots, n \end{cases}$$

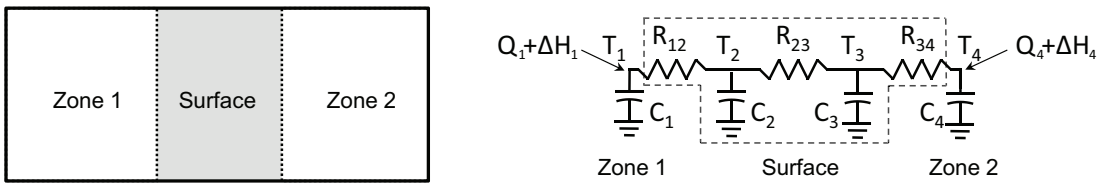
  where C_{pa} is the specific heat capacitance of the supplied air at constant pressure. Recall that T^s is the supply air temperature.

The coupled ordinary differential equation model (1) so obtained is nonlinear because of the presence of the bilinear term $\dot{m}_i^{in} T_i$ in defining the ventilation heat exchange ΔH_i. Note that the term $\dot{Q}_i + \Delta H_i$ can be together interpreted as a current source injected into (or extracted from) the node i of the RC-network, except that the source strength depends on the "voltage" T_i of node i as well: The full-order model can be thought as a RC-network model with additional current sources, where the source strengths are dependent on the voltage of the nodes they are connected to.

In the following sections, a compact state-space representation of the building thermal model (1) is used. The outside temperature is taken as a "virtual state" T_{n+1} to the system. We assign a very large "virtual capacitance" to the outside node: $C_{n+1} \gg C_i$, for $i = 1, \ldots, n$. Letting $C_{n+1} \to \infty$, the system of equations (1) is expressed as a state-space representation:

$$\frac{dT}{dt} = AT + L(T, U, \dot{Q}) \quad (2)$$

where the state vector $T := [T_1, \ldots, T_{n+1}]^T$, the control vector $U := [\dot{m}_1^{in}, \ldots, \dot{m}_N^{in}, 0, \ldots, 0]^T$, and the heat gain vector $\dot{Q} := [\dot{Q}_1, \ldots, \dot{Q}_N, 0, \ldots, 0]^T$. The *transition rate*



(a) (b)

Fig. 2. (a) Two zones separated by a single surface, and (b) its lumped RC-network model.

matrix A is an $(n+1)\times(n+1)$ matrix with entries given by

$$A_{ij} := \begin{cases} 0, & \text{if } j \neq i,\ (i,j) \notin \mathcal{E} \\ 1/(C_i R_{ij}), & \text{if } j \neq i,\ (i,j) \in \mathcal{E} \\ -\sum_{k\neq i} A_{ik}, & \text{if } j = i,\ (i,j) \in \mathcal{E} \end{cases} \tag{3}$$

and the nonlinear function:

$$\begin{cases} L_i(T,U,\dot{Q}) = \dfrac{C_{pa}U_i(T^s - T_i) + \dot{Q}_i}{C_i}, \quad i = 1,\dots,N \\ L_i(T,U,\dot{Q}) = 0, \quad i = N+1,\dots,n \\ L_i(T,U,\dot{Q}) = \eta, \quad i = n+1 \end{cases}$$

where $\eta(t) \in \mathbb{R}$ is chosen such that $\eta(t) = \dot{T}_o(t)$. To see the equivalence between (1) and (2), note that the entries in the last row of A approach 0 as $C_{n+1} \to \infty$ (since they are of the form $1/(C_{n+1}R_{n+1,j})$). In the limit, $\dot{T}_{n+1} = \eta(t)$, which gives $T_{n+1}(t) = T_o(t)$ for all $t \geq 0$.

Example 1 (A simple two-zone building) Consider the simplest example where two zones are separated by a single wall/surface as shown in Fig. 2 (a). Here it is assumed that two zones have no thermal interaction with other zones or the outside. A 3R2C network model is used to model the surface as shown in Fig. 2 (b). Ventilation air enters each zone at temperature T^s, and leaves the zone at the same temperature as that of the zone. There are 4 building nodes, two zone nodes plus two internal nodes. T_1 and T_4 denote the space temperatures of zone 1 and zone 2, respectively. T_2 and T_3 denote temperatures of the points internal to the surface that arise due to the 3R2C model of the surface. The parameters C_1 and C_4 are the thermal capacitances of the two zones, while C_2, C_3 are thermal capacitances for the 3R2C model of the surface. The dynamics for the RC-network model are described by the following differential equations obtained by using the heat balance:

$$\begin{aligned} C_1\frac{dT_1}{dt} &= -\frac{1}{R_{12}}T_1 + \frac{1}{R_{12}}T_2 + \dot{Q}_1 + \Delta H_1 \\ C_2\frac{dT_2}{dt} &= -(\frac{1}{R_{21}} + \frac{1}{R_{23}})T_2 + \frac{1}{R_{21}}T_1 + \frac{1}{R_{23}}T_3 \\ C_3\frac{dT_3}{dt} &= -(\frac{1}{R_{32}} + \frac{1}{R_{34}})T_3 + \frac{1}{R_{32}}T_2 + \frac{1}{R_{34}}T_4 \\ C_4\frac{dT_4}{dt} &= -\frac{1}{R_{43}}T_4 + \frac{1}{R_{43}}T_3 + \dot{Q}_4 + \Delta H_4 \end{aligned} \tag{4}$$

where R_{12}, R_{23}, R_{34} are thermal resistances, the $\dot{Q}_i = \dot{Q}_i^r + \dot{Q}_i^{int} + \dot{Q}_i^{ext}$ is the heat gain, and $\Delta H_i = C_{pa}\dot{m}_i^{in}(T^s - T_i)$ is the heat exchange due to ventilation for $i = 1, 4$, with C_{pa} being the specific heat capacitance of the supply air. ■

2.2 Problem statement of model reduction

For a building with N zones, the number of states in the full-order model (2) described above is of the order of $7N$, usually more. A medium size commercial building has close to 100 zones and a larger building can have several hundreds. The dimension of the full-order model thus can be quite large. The goal of this paper is to obtain a reduced-order model of smaller dimension such that model reduction produces the RC-network physical structure.

To achieve this goal, an aggregation methodology is considered: Mathematically, suppose the goal is to reduce the state dimension from n to m, where $m \leq n$ is the (user-specified) number of super-nodes. The first step is to choose a *partition function* $\phi : \mathcal{V} \to \bar{\mathcal{V}}$, where $\bar{\mathcal{V}} = \{1,\dots,m+1\}$ denotes the set of "super-nodes" for the reduced-order model, and recall that $\mathcal{V} = \{1,\dots,n+1\}$ denotes the set of nodes for the full-order model. A partition function is an onto function but possibly many-to-one. The elements of $\bar{\mathcal{V}}$ are the super-nodes, and for every $k \in \bar{\mathcal{V}}$, the inverse mapping $\phi^{-1}(k) \subset \mathcal{V}$ denotes the group of nodes in the full-order model that are aggregated into the kth super-node using the partition function ϕ. The second step is to define a graph for $\mathcal{V}$ by defining "super-edges" between super-nodes. The third step is to define appropriate super-capacitances and super-resistances.

To perform such a structure-preserving model reduction, we need to answer the following questions:

Q 1: *How to choose the partition function?*
Q 2: *Given a partition function, how to find the super-capacitances and super-resistances, and how to aggregate the nonlinear terms of the full-order model?*
Q 3: *How to compare the full and reduced-order models?*

The rest of the paper is about answering these questions. A brief outline of the approach is provided below.

A 1: *Partition by Markov chain aggregation:* It is shown in Section 3 that the linear thermal dynamics is analogous to a continuous-time Markov chain. A recently developed aggregation method for Markov chains is then

employed to obtain a (sub)-optimal partition function. In this method, a measure of optimality of the aggregation is defined in terms of the Kullback-Leibler divergence rate. Solving the optimal partition problem is shown to be hard, and a recursive bi-partition algorithm is proposed here to obtain sub-optimal partitions. Details of the algorithm are summarized in Appendix B.

A 2: *Finding super-quantities and the reduced-order model:* The super-capacitances and super-resistances are obtained directly based on the Markov chain analogy. Details appears in Section 4.1. Due to the current source interpretation of the nonlinear part, the current sources connecting to the same group of the aggregated nodes are directly added up to form a super-current source for the corresponding super-node. It is shown that the reduced-order model is a RC-network model defined with super-quantities. Details appear in Section 4.2.

A 3: *Comparison between full and reduced-order models:* We obtain the reduced-order model by aggregating the nodes into a smaller number of super-nodes. Then we employ the *lifting technique* to lift the reduced-order model to the one with the same number of nodes as the full-order model. Details of the lifting technique is summarized in Section 4.1. We can now directly compare the temperatures of zone nodes for the full and reduced-order models.

3 Markov Chain Analogy and Aggregation

In this section, it is shown that the linear part of the building thermal model (2) is analogous to a *continuous-time Markov chain*. The *linear dynamics* of the building thermal model (2) are given by:

$$\frac{dT}{dt} = AT. \tag{5}$$

Due to the special structure of the matrix A (see (3)), the linear thermal model (5) is *conservative*. Specifically, a scalar-valued function $V(t) := \sum_{i\in\mathcal{V}} C_i T_i(t)$ is conserved for all time:

$$\begin{aligned}\frac{dV}{dt}(t) &= \sum_{i\in\mathcal{V}} C_i \frac{dT_i}{dt}(t)\\ &= \sum_{j\in\mathcal{V}} T_j(t) \sum_{i\in\mathcal{V}} C_i A_{ij} &(6)\\ &= \sum_{j\in\mathcal{V}} C_j T_j(t) \sum_{i\in\mathcal{V}} A_{ji} &(7)\\ &= 0 &(8)\end{aligned}$$

where the equality (6) follows from (5), the equality (7) is due to the fact that $C_i A_{ij} = C_j A_{ji}$ for all $i, j \in \mathcal{V}$ (see (3)), and the equality (8) uses the fact that each row sum of the matrix A is zero. We denote $V_0 := V(0) = \sum_{i\in\mathcal{V}} C_i T_i(0)$ as the *invariant quantity* of the linear thermal model (5).

3.1 Analogy to a Markov chain

Based upon the conservative property of the linear thermal model (5), define the *thermal distribution* as a row vector, denoted by f, where

$$f_i = \frac{C_i}{V_0} T_i, \quad i \in \mathcal{V}.$$

Note that $\sum_{i\in\mathcal{V}} f_i(t) \equiv 1$ for all $t \geq 0$.

On differentiating f_i with respect to t, and using (5),

$$\frac{df_i}{dt} = \frac{C_i}{V_0}\frac{dT_i}{dt} = \sum_{j\in\mathcal{V}} A_{ij}\frac{C_i}{V_0} T_j. \tag{9}$$

By substituting (3) in (9), we have

$$\begin{aligned}\frac{df_i}{dt} &= A_{ii}\frac{C_i}{V_0}T_i + \sum_{j\neq i}\frac{1}{C_i R_{ij}}\frac{C_i}{V_0}T_j\\ &= A_{ii}\frac{C_i}{V_0}T_i + \sum_{j\neq i}\frac{1}{C_j R_{ji}}\frac{C_j}{V_0}T_j &(10)\\ &= \sum_{j\in\mathcal{V}} f_j A_{ji} &(11)\end{aligned}$$

where the fact that $R_{ij} = R_{ji}$ is used in deriving the equality (10). Using matrix notation for representing (11), we obtain the dynamics of the thermal distribution and its solution

$$\frac{df}{dt} = fA \qquad \Rightarrow \qquad f(t) = f(0)e^{At}, \quad \forall t \geq 0.$$

The Markov chain analogy is now clear. Note that each row sum of A is zero, its diagonal entries are negative, and its non-diagonal entries are non-negative (see (3)). Thus, the transition rate matrix A is the *infinitesimal generator* of a *transition semigroup* $\{e^{At}\}_{t\geq 0}$: For any $t, s \geq 0$, (i) $e^{A0} = I$, (ii) e^{At} is a stochastic matrix (That is e^{At} is a nonnegative matrix whose row sums are equal to one), and (iii) $e^{A(t+s)} = e^{At}e^{As}$.

Consider now a continuous-time Markov chain $\{X(t)\}_{t\geq 0}$ on the state space $\mathcal{V}$ with the transition semigroup $\{e^{At}\}_{t\geq 0}$ [40]. Let $g(t)$ denote the *probability distribution* at time t, i.e.,

$$g_i(t) = \Pr(X(t) = i), \quad i \in \mathcal{V}.$$

Using the transition semigroup property, we have

$$\begin{aligned}g_i(t) &= \sum_{j\in\mathcal{V}} \Pr(X(0) = j)\Pr(X(t) = i \mid X(0 = j))\\ &= \sum_{j\in\mathcal{V}} g_j(0)(e^{At})_{ji}.\end{aligned}$$

If $f(0)$ is the initial distribution of $\{X(t)\}_{t\geq 0}$, i.e., $g(0) = f(0)$, then

$$g(t) = g(0)e^{At} = f(0)e^{At} = f(t). \tag{12}$$

Thus, starting from the same initial distribution, the probability distribution of the continuous-time Markov chain $\{X(t)\}_{t\geq 0}$ is equal to the thermal distribution of the linear thermal model (5). For more details on continuous-time Markov chains, we refer the reader to [35,40] and references therein.

For any ergodic Markov chain, there exists a unique *stationary distribution* π (obtained as a solution to $\pi A = 0$), whereby starting from any initial distribution

$$\lim_{t\to\infty} g(t) = \pi.$$

For linear thermal model (5), the associated Markov chain is shown to be ergodic in [14], and the stationary distribution is given by:

$$\pi_i = \frac{C_i}{\sum_{j\in\mathcal{V}} C_j}, \quad i \in \mathcal{V}. \tag{13}$$

3.2 Discretization of the continuous-time Markov chain

In practice, it is more convenient to work with discrete-time Markov chains (DTMC) instead of continuous-time Markov chains (CTMC). The DTMC $\{X(k\Delta t)\}_{k\geq 0}$ is obtained by discretizing the CTMC $\{X(t)\}_{t\geq 0}$ with a step-size Δt. Let $\xi(k)$ denote the probability distribution of the DTMC at kth time-step, i.e., $\xi_i(k) = \Pr(X(k\Delta t) = i)$ for $i \in \mathcal{V}$. Using (12),

$$\xi(k) = \xi(0)P^k(\Delta t), \quad k \geq 0$$

where the *transition matrix* is defined as

$$P(\Delta t) := e^{A\Delta t}. \tag{14}$$

For any $t \geq 0$, there exists an integer $k \geq 0$ such that $k\Delta t \leq t < (k+1)\Delta t$ and $e^{At} \approx e^{Ak\Delta t}$ for small enough Δt. Thus the CTMC $\{X(t)\}_{t\geq 0}$ with the transition semigroup $\{e^{At}\}_{t\geq 0}$ is approximated by the DTMC $\{X(k\Delta t)\}_{k\geq 0}$ with the transition matrix $P(\Delta t)$ as $\Delta t \to 0$. One can verify that π given in (13) is also the stationary distribution of the DTMC, i.e., $\lim_{k\to\infty} \xi(k) = \pi$.

3.3 Kullback-Leibler metric

For the model reduction problem, one needs a metric to quantify the differences between the full and the reduced-order models. In this paper, the *Kullback-Leibler (KL) divergence* is proposed as a "probability metric" to quantify the difference between distributions. Let ξ and ζ denote two probability distributions defined on the same space $\mathcal{V} = \{1, \ldots, n+1\}$. The KL divergence between ξ and ζ is given by:

$$D(\xi\|\zeta) := \begin{cases} \sum_{i\in\mathcal{V}} \xi_i \log(\xi_i/\zeta_i), & \text{if} \quad \xi \prec \zeta \\ +\infty, & \text{otherwise} \end{cases} \tag{15}$$

where $\xi \prec \zeta$ means that ξ is *absolutely continuous* with respect to ζ, i.e., $(\zeta_i = 0) \Rightarrow (\xi_i = 0)$ for all $i \in \mathcal{V}$. Note that KL divergence is not a true metric since it is not symmetric and it does not satisfy the triangle inequality [12]. However, KL divergence is a natural pseudo-metric for comparing probability distributions and is widely used in statistics, information theory, and control theory [15, 41]. It has two useful properties:

(1) KL divergence is a *pre-metric*:

$$D(\xi\|\zeta) \geq 0, \text{ with equality if and only if } \xi = \zeta.$$

(2) If the KL divergence between two distributions is small then the two distributions are also close in the sense of the standard L_1 metric [12]:

$$D(\xi\|\zeta) \geq \frac{1}{2}\|\xi - \zeta\|_1^2, \text{with } \|\xi - \zeta\|_1 := \sum_{i\in\mathcal{V}} |\xi_i - \zeta_i|.$$

For the model reduction problem, it is of interest to compare two probability distributions defined on different state spaces. Let ξ and $\bar{\xi}$ denote two probability distributions defined on $\mathcal{V} = \{1, \ldots, n+1\}$ and $\bar{\mathcal{V}} = \{1, \ldots, m+1\}$ ($m \leq n$), respectively. The relationship between $\mathcal{V}$ and $\bar{\mathcal{V}}$ is described by a given partition function $\phi : \mathcal{V} \mapsto \bar{\mathcal{V}}$. Since ξ and $\bar{\xi}$ are not defined on the same space, one can not directly use the formula (15) to compute the KL divergence between ξ and $\bar{\xi}$. One strategy, proposed in [15], is to *lift* $\bar{\xi}$ from the space $\bar{\mathcal{V}}$ to the space $\mathcal{V}$ according a *lifting distribution* μ, where $\mu_i \geq 0$ and $\sum_{i\in\mathcal{V}} \mu_i = 1$. The lifted probability distribution $\widehat{\xi}$ is defined as

$$\widehat{\xi}_i^{(\mu)}(\phi) := \frac{\mu_i}{\sum_{k\in\psi(i)} \mu_k} \bar{\xi}_{\phi(i)}, \quad i \in \mathcal{V}$$

where $\psi(i) = \phi^{-1} \circ \phi(i) \subset \mathcal{V}$ denotes the set of states belonging to the same group as the ith state. The KL divergence is then extended for two distributions defined on different state spaces:

$$D_\phi(\xi\|\bar{\xi}) := \min_{\mu} D(\xi\|\widehat{\xi}^{(\mu)}(\phi)).$$

The *Kullback-Leibler divergence rate* is a generalization of the KL divergence. It is a measure of distance between the probability laws of two stochastic processes. The formula for KL divergence rate, denoted by $R_\phi(P\|\bar{P})$, between two Markov chains (π, P) defined on $\mathcal{V}$ and $(\bar{\pi}, \bar{P})$ defined on $\bar{\mathcal{V}}$, appears in Appendix A with additional details in [15].

3.4 Aggregation of Markov chain

Let (π, P) denote a discrete-time Markov chain defined on the state space $\mathcal{V} = \{1, \ldots, n+1\}$ with the transition matrix P and the stationary distribution π. The model reduction problem is to find an optimal aggregated Markov chain, denoted by $(\bar{\pi}, \bar{P})$, defined on the state space $\bar{\mathcal{V}} = \{1, \ldots, m+1\}$, where $m \leq n$, such that the KL divergence rate $R_\phi(P\|\bar{P})$ between two Markov chains is minimized.

The *m-partition problem* is to find a partition function $\phi : \mathcal{V} \mapsto \bar{\mathcal{V}}$ and an aggregated transition matrix $\bar{P}$ that solves the following optimization problem:

$$\begin{aligned} &\min_{\phi, \bar{P}} \; R_\phi(P\|\bar{P}) \\ &\text{s.t.} \quad \bar{P}\mathsf{1} = \mathsf{1},\, \bar{P} \geq 0. \end{aligned}$$

As shown in Theorem 2 of [15], for a fixed (whether optimal or not) partition function ϕ, the optimal aggregated Markov chain $(\bar{\pi}(\phi), \bar{P}(\phi))$ is given by:

$$\bar{P}_{kl}(\phi) = \frac{\sum_{i \in \phi^{-1}(k)} \sum_{j \in \phi^{-1}(l)} \pi_i P_{ij}}{\sum_{i \in \phi^{-1}(k)} \pi_i}, \quad k, l \in \bar{\mathcal{V}} \quad (16)$$

where the stationary distribution of $\bar{P}(\phi)$ is given by

$$\bar{\pi}_k(\phi) = \sum_{i \in \phi^{-1}(k)} \pi_i, \quad k \in \bar{\mathcal{V}}. \quad (17)$$

As a result, the m-partition problem reduces to finding only an *optimal partition function* $\phi^* : \mathcal{V} \to \bar{\mathcal{V}}$ such that

$$\phi^* \in \arg\min_{\phi} \; R_\phi(P\|\bar{P}(\phi)). \quad (18)$$

It is shown in [15] that solving the optimization problem (18) exactly is difficult for $m > 2$, but a sub-optimal solution for $m = 2$ can be easily computed. This leads to a sub-optimal solution for arbitrary $m \geq 2$ through the *recursive bi-partition algorithm* AlgoBIPA, which is described in Appendix B. Here we summarize the basic ideas: First, AlgoBIPA is used to obtain $\mathcal{V}_1, \mathcal{V}_2$ so that $\mathcal{V}_1 \cup \mathcal{V}_2 = \mathcal{V}$ and $\mathcal{V}_1 \cap \mathcal{V}_2 = \Phi$, where Φ denotes the empty set. Then, AlgoBIPA is used on the set $\mathcal{V}_1$ to obtain $\mathcal{V}_{11}$ and $\mathcal{V}_{12}$ such that $\mathcal{V}_{11} \cup \mathcal{V}_{12} = \mathcal{V}_1$ and $\mathcal{V}_{11} \cap \mathcal{V}_{12} = \Phi$. If the model order is less than m, AlgoBIPA is used on the set $\mathcal{V}_2$ to obtain partitions $\mathcal{V}_{21}$ and $\mathcal{V}_{22}$ such that $\mathcal{V}_{21} \cup \mathcal{V}_{22} = \mathcal{V}_2$ and $\mathcal{V}_{21} \cap \mathcal{V}_{22} = \Phi$. This procedure is repeated until m partitions are obtained. At each step the number of partition increases by one.

3.5 Analogy to thermal dynamics

Based on the Markov chain analogy for the linear thermal dynamics (see Section 3.1), the model reduction framework for Markov chains is extended to building thermal models:

- *Metric for comparing thermal distributions*: The KL divergence is employed as a metric to compare two thermal distributions f and g defined on the same building node set $\mathcal{V}$:
$$D(f\|g) = \sum_{i \in \mathcal{V}} f_i \log(f_i/g_i).$$
For the model reduction problem, it is of interest to compare two thermal distributions defined on building graphs of different cardinalities. Let f and $\bar{f}$ denote two thermal distributions defined on $\mathcal{V}$ and $\bar{\mathcal{V}}$, respectively. The low-dimensional distribution $\bar{f}$ is lifted to a high-dimensional distribution $\widehat{f}$ defined on $\mathcal{V}$ by using partition function ϕ and lifting distribution μ:
$$\widehat{f}_i^{(\mu)}(\phi) = \frac{\mu_i}{\sum_{k \in \psi(i)} \mu_k} \bar{f}_{\phi(i)}, \quad i \in \mathcal{V}. \quad (19)$$
The lifting may be viewed as a linear transformation that conserves the total heat. The KL metric is then used to compare the two thermal distributions f and $\widehat{f}$ on the same node set $\mathcal{V}$.
- *Metric for comparing thermal models*: The KL divergence rate is used as a measure to compare two building thermal models. In particular, suppose full-order model is simulated starting from an initial distribution $f(0)$. Denote the resulting trajectory of the thermal distribution as $\{f(k\Delta t)\}_{0 \leq k \leq N}$. Now, suppose the reduced-order model is also simulated starting from the initial distribution
$$\bar{f}_l(0) = \sum_{i \in \phi^{-1}(l)} f_i(0), \quad l \in \mathcal{V}.$$
Denote the resulting trajectory of thermal distribution as $\{\bar{f}(k\Delta t)\}_{0 \leq k \leq N}$, which evolves over reduced graph $\bar{\mathcal{V}}$. The trajectory $\{\bar{f}(k\Delta t)\}_{0 \leq k \leq N}$ is lifted to the full building graph by using (19), and denoted by $\{\widehat{f}(k\Delta t)\}_{0 \leq k \leq N}$. The KL divergence rate between full and reduced-order models is given by,
$$\frac{1}{N} \sum_{k=1}^{N} D(f(k\Delta t)\|\widehat{f}(k\Delta t)). \quad (20)$$
Thus, the KL divergence rate is a measure of average distance between trajectories generated from simulating two thermal models.
- *Bi-partition*: An optimal bi-partition of a given model produces a 2-state reduced-order model that is closest to the full-order model in the sense of distance (20). Since the distance is a time average, discrepancies between two models at the slow(est) time-scales contribute more to the error compared to the fast transients. The choice of metric thus leads to a 2-state model that approximates the full-order model on the slowest time-scale.
- *Recursive bi-partition*: The recursive application of bi-partition algorithm produces a reduced-order model that progressively captures multiple time-scales in the problem. The first bi-partition results in splitting of the graph

into two clusters, and a 2-state model that captures the slowest time-scale. The next bi-partition further splits one of the two clusters so as to capture the slowest time-scale in that cluster and so on. In effect after m-applications of the algorithm, the reduced-order model describes the m slowest time-scales of the full-order model.

4 Aggregated Building Thermal Model

In this section, the aggregation methodology is applied to obtain a reduced-order model for building thermal model (2). We first describe the reduced-order model for the linear part of the building thermal model (2), and then the reduced-order model for the nonlinear part of (2).

4.1 Aggregated linear thermal dynamics

For the linear thermal model (5), the goal is to aggregate the node set $\mathcal{V} = \{1, \ldots, n+1\}$ into a smaller super-node set $\bar{\mathcal{V}} = \{1, \ldots, m+1\}$ where $m \leq n$. For each super-node $k \in \bar{\mathcal{V}}$, we introduce the super-temperature $\bar{T}_k$, super-capacitance $\bar{C}_k$, and super-resistance $\bar{R}_{kl}$. For a given partition function ϕ, the reduced-order model for (5) has the form:

$$\frac{d\bar{T}}{dt} = \bar{A}(\phi)\bar{T}, \tag{21}$$

where $\bar{T} = [\bar{T}_1, \ldots, \bar{T}_{m+1}]^T$ denotes the super-temperature vector, and $\bar{A}(\phi)$ denotes the $(m+1) \times (m+1)$ super-transition-rate matrix. The Markov chain analogy also works for the reduced-order model with the associated transition semigroup $\{e^{\bar{A}(\phi)t}\}_{t\geq 0}$. Discretizing with a small step-size Δt, one obtains the transition matrix for the aggregated Markov chain defined on $\bar{\mathcal{V}}$:

$$\bar{P}(\Delta t) := e^{\bar{A}(\phi)\Delta t}.$$

Recall that the transition matrix for the discrete-time Markov chain associated with the full-order linear thermal dynamics is denoted by $P(\Delta t)$ (see (14)). The goal is to choose $\bar{A}(\phi)$ so that the aggregated Markov chain with the transition matrix $\bar{P}(\Delta t)$ optimally approximates the original Markov chain with the transition matrix $P(\Delta t)$. The aggregation method described in Section 3.4 is employed to determine the formula for the optimal aggregated transition matrix $\bar{A}(\phi)$. According to (16), the formula for the optimal aggregated Markov transition matrix is given by:

$$\bar{P}_{kl}(\Delta t) = \frac{\sum_{i\in\phi^{-1}(k)} \sum_{j\in\phi^{-1}(l)} \pi_i P_{ij}(\Delta t)}{\sum_{i\in\phi^{-1}(k)} \pi_i}, \quad k, l \in \bar{\mathcal{V}}. \tag{22}$$

By expressing $P(\Delta t)$ and $\bar{P}(\Delta t)$ in the form

$$P(\Delta t) = I + A\Delta t + O(\Delta t^2),$$
$$\bar{P}(\Delta t) = I + \bar{A}(\phi)\Delta t + O(\Delta t^2),$$

the equation (22) becomes

$$\begin{aligned} &\mathbb{1}_{\{k=l\}} + \bar{A}_{kl}(\phi)\Delta t + O(\Delta t^2) \\ &= \frac{\sum_{i\in\phi^{-1}(k)} \sum_{j\in\phi^{-1}(l)} \pi_i (\mathbb{1}_{\{i=j\}} + A_{ij}\Delta t + O(\Delta t^2))}{\sum_{i\in\phi^{-1}(k)} \pi_i} \\ &= \mathbb{1}_{\{k=l\}} + \frac{\sum_{i\in\phi^{-1}(k)} \sum_{j\in\phi^{-1}(l)} \pi_i A_{ij}}{\sum_{i\in\phi^{-1}(k)} \pi_i}\Delta t + O(\Delta t^2). \end{aligned} \tag{23}$$

By matching terms on both sides of (23), we obtain the formula for the optimal super-transition-rate matrix

$$\bar{A}_{kl}(\phi) = \frac{\sum_{i\in\phi^{-1}(k)} \sum_{j\in\phi^{-1}(l)} \pi_i A_{ij}}{\sum_{i\in\phi^{-1}(k)} \pi_i}, \quad k, l \in \bar{\mathcal{V}}. \tag{24}$$

By substituting (3) and (13) into (24), one can verify that $\bar{A}(\phi)$ is indeed a *transition-rate matrix* for any partition function ϕ, i.e., the row sums of $\bar{A}(\phi)$ are zeros, diagonal entries are negative, and non-diagonal entries are non-negative:

$$\begin{cases} \bar{A}_{kl}(\phi) = \dfrac{\sum_{i\in\phi^{-1}(k)} \sum_{j\in\phi^{-1}(l)} 1/R_{ij}}{\sum_{i\in\phi^{-1}(k)} C_i}, \quad k \neq l \in \bar{\mathcal{V}} \\ \bar{A}_{kk}(\phi) = -\displaystyle\sum_{l\neq k} \bar{A}_{kl}(\phi), \quad k \in \bar{\mathcal{V}} \end{cases} \tag{25}$$

The super-capacitances and super-resistances can also be expressed in terms of C_i and R_{ij}:

- According to (13), the stationary distribution of the aggregated Markov chain has the form:

$$\bar{\pi}_k(\phi) = \frac{\bar{C}_k(\phi)}{\sum_{l\in\bar{\mathcal{V}}} \bar{C}_l(\phi)}, \quad k \in \bar{\mathcal{V}} \tag{26}$$

  where $\bar{C}_k(\phi)$ denotes the *super-capacitances* for the kth node. By substituting (13) into (17), we obtain formula for the the optimal stationary distribution:

$$\bar{\pi}_k(\phi) = \sum_{i\in\phi^{-1}(k)} \pi_i = \frac{\sum_{i\in\phi^{-1}(k)} C_i}{\sum_{l\in\bar{\mathcal{V}}} \sum_{j\in\phi^{-1}(l)} C_j}, \quad k \in \bar{\mathcal{V}}. \tag{27}$$

  By comparing (26) and (27), we obtain the formulae for the super-capacitances:

$$\bar{C}_k(\phi) = \sum_{i\in\phi^{-1}(k)} C_i, \quad k \in \bar{\mathcal{V}}. \tag{28}$$

- By using (25) and (28), we obtain the formulae for the

super-resistances:

$$\bar{R}_{kl}(\phi) = \frac{1}{\bar{C}_k(\phi)\bar{A}_{kl}(\phi)} = \frac{1}{\sum_{i\in\phi^{-1}(k)}\sum_{j\in\phi^{-1}(l)} 1/R_{ij}}, \quad k \neq l \in \bar{\mathcal{V}}. \tag{29}$$

Thus, the reduced-order linear model (21) corresponds to a reduced RC-network with super-capacitances and super-resistances given by (28) and (29), respectively. The super-capacitance $\bar{C}_k(\phi)$, given in (28), is the equivalent capacitance of parallel configuration of all capacitors in the kth partition. Similarly, the super-resistance $\bar{R}_{kl}(\phi)$ given in (29) is the equivalent resistance of parallel configuration of all resistors connecting the kth partition and the lth partition. These observations also serve to provide an intuitive justification of the aggregation approach.

Similar to the full-order model (2), the reduced-order model (21) is also *conservative* because of the fact that $\bar{A}(\phi)$ is a super-transition-rate matrix. The invariant quantity for the reduced-order model is given by

$$\bar{V}_0 := \sum_{k\in\bar{\mathcal{V}}} \bar{C}_k\bar{T}_k(0).$$

If one chooses the initial condition for the reduced-order model (21) as

$$\bar{T}_k(0) = \sum_{i\in\phi^{-1}(k)} (C_i/\bar{C}_k(\phi))T_i(0), \quad k \in \bar{\mathcal{V}}, \tag{30}$$

then

$$\bar{V}_0 = \sum_{k\in\bar{\mathcal{V}}}\sum_{i\in\phi^{-1}(k)} C_iT_i(0) = \sum_{i\in\mathcal{V}} C_iT_i(0) = V_0.$$

This implies that the invariant quantity of the linear thermal dynamics is unchanged after the aggregation. The *aggregated thermal distribution* is defined as

$$\bar{f}_k = \frac{\bar{C}_k}{\bar{V}_0}\bar{T}_k, \quad k \in \bar{\mathcal{V}}. \tag{31}$$

Recall that we introduce the lifting technique to compare the low and high-dimensional distributions. The *lifted thermal distribution* is defined as

$$\widehat{f}_i = \frac{C_i}{V_0}\widehat{T}_i, \quad i \in \mathcal{V} \tag{32}$$

where $\widehat{T}_i$ is called the *lifted temperature* for the node i. Using (19) and choosing the lifting distribution as $\mu = \pi$, we obtain

$$\widehat{f}_i = \frac{\pi_i}{\sum_{j\in\psi(i)}\pi_j}\bar{f}_{\phi(i)}. \tag{33}$$

Substituting (13) and (31) into (33), we have

$$\widehat{f}_i = \frac{C_i}{\sum_{j\in\psi(i)} C_j}\frac{\bar{C}_{\phi(i)}}{\bar{V}_0}\bar{T}_{\phi(i)} = \frac{C_i}{V_0}\bar{T}_{\phi(i)} \tag{34}$$

where we use the fact that $\bar{C}_{\phi(i)} = \sum_{j\in\psi(i)} C_j$ and $\bar{V}_0 = V_0$. By comparing (32) and (34), we have the explicit expression for the lifted temperature

$$\widehat{T}_i = \bar{T}_{\phi(i)}, \quad i \in \mathcal{V}.$$

Note that the lifted temperature $\widehat{T}_i$ of the node i is indeed a temperature quantity for the corresponding aggregated node $\phi(i)$. Thus, we can compare the full and reduced-order models by directly comparing T_i and $\bar{T}_{\phi(i)}$ for each node i.

4.2 Aggregated building thermal model

Recall that the outside node is taken as a virtual $(n+1)$th node in the full-order building thermal model (2), and the outside temperature is denoted as T_{n+1}. We also take the outside node as a virtual $(m+1)$th node in the reduced-order model and we denote its temperature as $\bar{T}_{m+1}$. That is, for any given partition function ϕ, the building node set $\{1,\dots,n\}$ is aggregated into the super-node set $\{1,\dots,m\}$, and the $(n+1)$th outside node has a one-to-one correspondence to the $(m+1)$th super-node.

Due to the current source interpretation of nonlinear thermal dynamics $L(T,U,Q)$ (see Section 2.1), the current sources connecting to the same group of the aggregated nodes are directly added up to form a super-current source for the corresponding super-node:

- For $k = 1,\dots,m$, the aggregated nonlinear thermal dynamics is given by:

$$\begin{aligned}\widetilde{L}_k(T,U,\dot{Q}) &= \sum_{i\in\phi^{-1}(k)} C_iL_i(T,U,Q)/\bar{C}_k(\phi) \\ &= (C_{pa}(T^s\bar{U}_k(\phi) - \widetilde{W}_k(\phi)) + \dot{\bar{Q}}_k(\phi))/\bar{C}_k(\phi)\end{aligned} \tag{35}$$

  where

$$\begin{aligned}\bar{U}_k(\phi) &:= \sum_{i\in\phi^{-1}(k)} U_i \\ \dot{\bar{Q}}_k(\phi) &:= \sum_{i\in\phi^{-1}(k)} \dot{Q}_i \\ \widetilde{W}_k(\phi) &:= \sum_{i\in\phi^{-1}(k)} U_iT_i.\end{aligned} \tag{36}$$

- For $k = m+1$, the aggregated nonlinear thermal dynamics is given by:

$$\widetilde{L}_{m+1}(T,U,\dot{Q}) = \eta$$

  where $\eta(t) = \dot{T}_o(t)$ and $T_o(t)$ denotes the outside temperature. The construction here is to make sure $\bar{T}_{m+1}(t) = T_o(t)$ for all $t \geq 0$ in the reduced-order model (38) described later.

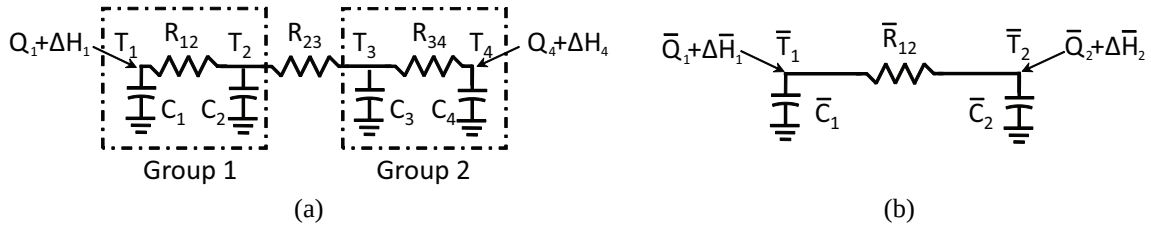


Fig. 3. The RC-network representation of (a) the full-order thermal model and (b) the reduced-order thermal model for the single surface separating two zones.

Since U and Q are external inputs to the full-order model, we can also take $\bar{U}(\phi)$ and $\dot{\bar{Q}}(\phi)$ defined in (36) as the super-inputs to the reduced-order model. One problem is that the term $\widetilde{W}(\phi)$ defined in (36) depends on T, which is the state vector of the full-order model. We use $\bar{T}_k$ (the temperature of the kth super-node) to approximate T_i (the temperature of the ith node that belongs to the kth group) in $\widetilde{W}_k(\phi)$:

$$\bar{W}_k(\phi) := \sum_{i\in\phi^{-1}(k)} U_i \bar{T}_k = \bar{U}_k(\phi)\bar{T}_k, \quad k = 1,\ldots,m.$$

Replacing $\widetilde{W}_k(\phi)$ by $\bar{W}_k(\phi)$ in (35), we approximate $\widetilde{L}_k(T,U,\dot{Q})$ by

$$\begin{aligned}\bar{L}_k(\bar{T},&\bar{U}(\phi),\dot{\bar{Q}}(\phi))\\ &= \left(C_{pa}\bar{U}_k(\phi)(T^s-\bar{T}_k) + \dot{\bar{Q}}_k(\phi)\right)/\bar{C}_k(\phi)\end{aligned} \tag{37}$$

for $k = 1,\ldots,m$ and $\bar{L}_{m+1}(\bar{T},\bar{U}(\phi),\dot{\bar{Q}}(\phi)) = \eta$. Note that the aggregated nonlinear thermal dynamics in (37) only depends on super-quantities for the reduced-order model.

By combining the aggregated linear thermal dynamics (21) with the aggregated nonlinear thermal (37), we obtain the state-space representation of the reduced-order building thermal model:

$$\frac{d\bar{T}}{dt} = \bar{A}(\phi)\bar{T} + \bar{L}(\bar{T},\bar{U}(\phi),\dot{\bar{Q}}(\phi)). \tag{38}$$

The model reduction method proposed in this paper preserves the RC-network structure of the original building model, that is, the reduced-order model (38) is still a RC-network defined with super-nodes with super-edges connecting these super-nodes. According to state-space representation (38), the aggregated building thermal dynamics can be also expressed by the following coupled differential equations: For each $k = 1,\ldots,m$,

$$\begin{aligned}\bar{C}_k(\phi)\frac{d\bar{T}_k}{dt}(t) = &\sum_{l\in\bar{\mathcal{N}}_k}(\bar{T}_l(t)-\bar{T}_k(t))/\bar{R}_{kl}(\phi)\\ &+\dot{\bar{Q}}_k(\phi)(t) + \Delta\bar{H}_k(\phi)(t)\end{aligned} \tag{39}$$

where $\bar{T}_k$ is the temperature of the kth super-node, $\bar{\mathcal{N}}_k \subset \bar{\mathcal{V}}$ denotes the set of neighbors of the kth super-node, $\dot{\bar{Q}}_k(\phi)$ denotes the heat gain for the kth super-node, and the ventilation heat exchange $\Delta\bar{H}_k(\phi)$ for the kth super-node is given by

$$\Delta\bar{H}_k(\phi)(t) = C_{pa}\dot{\bar{m}}_k^{in}(\phi)(t)(T^s - \bar{T}_k(t))$$

with the mass flow rate entering the kth super-node given by $\dot{\bar{m}}_k^{in}(\phi) = \sum_{i\in\phi^{-1}(k)}\dot{m}_i^{in}$. The initial condition of the reduced-order model (39) is chosen as (30).

The reduced-order model so far depends on the choice of the partition function ϕ. The sub-optimal partition function ϕ^* is obtained by using the recursive bi-partition algorithm AlgoBIPA. However, one can also directly choose a sub-optimal ϕ^* based on physical intuition (e.g., floor plans in a multi-zone building), or some other kinds of expert-based heuristics.

Example 2 (Bi-partition of the two-zone building) Consider the model reduction problem for the two-zone building shown in Example 1. The goal is to find a reduced-order model with two super-states. The thermal dynamics of this two-zone building is described by (4). To perform model reduction, we consider the linear dynamics of (2) first. The linear model is given by (5) with the transition rate matrix

$$A = \begin{bmatrix} -\frac{1}{C_1R_{12}} & \frac{1}{C_1R_{12}} & 0 & 0\\ \frac{1}{C_2R_{21}} & -\frac{R_{21}+R_{23}}{C_2R_{21}R_{23}} & \frac{1}{C_2R_{23}} & 0\\ 0 & \frac{1}{C_3R_{32}} & -\frac{R_{32}+R_{34}}{C_3R_{32}R_{34}} & \frac{1}{C_3R_{34}}\\ 0 & 0 & \frac{1}{C_4R_{43}} & -\frac{1}{C_4R_{43}}\end{bmatrix}$$

where the capacitances and resistances are given by $C_1 = 0.1$, $C_2 = 0.15$, $C_3 = 0.2$, $C_4 = 0.25$, $R_{12} = 0.15$, $R_{34} = 0.15$, and $R_{23} = 1.5$.

Using the bi-partition algorithm AlgoBIPA described in Appendix B, the sub-optimal bi-partition is obtained according to the sign structure of the second largest eigenvalue of the Markov transition matrix $P(\Delta t) = e^{A\Delta t}$, where the step-size $\Delta t = 0.01$. The second eigenvector of $P(\Delta t)$ is given by

$$u^{(2)} = [-0.639, -0.601, +0.312, +0.366].$$

The sign-structure of $u^{(2)}$ suggests the optimal bi-partition function $\phi^* = [1,1,2,2]$. That is, the nodes $\{1\}$ and $\{2\}$ are aggregated to form one super-node, and the states $\{3\}$

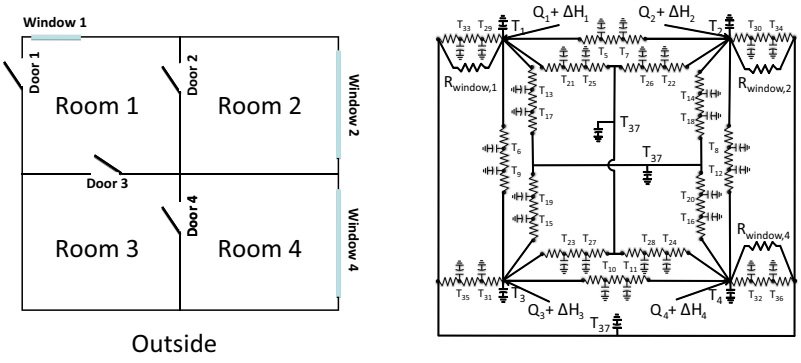


Fig. 4. (a) The layout of the four-zone building of the HVAC system shown in Fig. 1, and (b) its RC-network representation.

and $\{4\}$ are aggregated to form the other super-node. The optimal bi-partition is shown in Fig. 3 (a).

Using (24) with the optimal bi-partition function ϕ^*, we obtain the super-transition-rate matrix

$$\bar{A} = \begin{bmatrix} -\frac{1}{\bar{C}_1 \bar{R}_{12}} & \frac{1}{\bar{C}_1 \bar{R}_{12}} \\ \frac{1}{\bar{C}_2 \bar{R}_{21}} & -\frac{1}{\bar{C}_2 \bar{R}_{21}} \end{bmatrix}$$

where the super-capacitances and super-resistances are obtained according to (28) and (29)

$$\begin{aligned} \bar{C}_1 &= C_1 + C_2 = 0.25, \\ \bar{C}_2 &= C_3 + C_4 = 0.45, \\ \bar{R}_{12} &= \bar{R}_{21} = R_{23} = 1.5. \end{aligned}$$

The RC-network representation of the reduced-order model is shown in Fig. 3 (b). Let $\bar{T}_k$ denote the temperatures of super-nodes k for $k = 1, 2$. Using (39), we represent the reduced-order building thermal model by the following differential equations:

$$\begin{aligned} \bar{C}_1 \frac{d\bar{T}_1}{dt} &= -\frac{1}{\bar{R}_{12}} \bar{T}_1 + \frac{1}{\bar{R}_{12}} \bar{T}_2 + \dot{\bar{Q}}_1 + \Delta \bar{H}_1 \\ \bar{C}_2 \frac{d\bar{T}_2}{dt} &= -\frac{1}{\bar{R}_{21}} \bar{T}_2 + \frac{1}{\bar{R}_{21}} \bar{T}_1 + \dot{\bar{Q}}_2 + \Delta \bar{H}_2 \end{aligned}$$

where the heat gains $\dot{\bar{Q}}_1 = \dot{Q}_1$ and $\dot{\bar{Q}}_2 = \dot{Q}_4$, the heat exchanges $\Delta \bar{H}_1 = \Delta H_1$ and $\Delta \bar{H}_2 = \Delta H_4$.

The main observations from this model reduction example are as follows:

- The thermal interaction between the group of nodes $\{1, 2\}$ and the group $\{3, 4\}$ is much weaker compared to the interactions within each group since the resistance R_{23} separating the two groups is 10 times larger than the resistances R_{12} and R_{34} within each group. The optimal bi-partition produced by the algorithm AlgoBIPA thus is consistent with the heuristic that would partition the building based on the thermal interactions among the nodes.
- The reduced-order model is also a RC-network, just as the original model is. ■

5 Simulation and Discussion

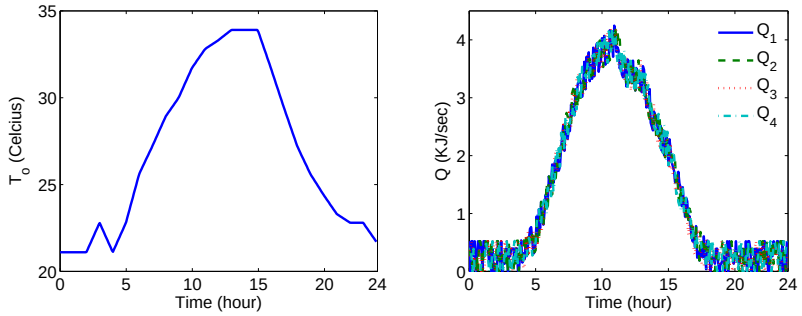
5.1 Simulation setup

In this section, we apply the aggregation-based model reduction method to the downstream part of the four-zone building HVAC system shown in Fig. 1, where each zone is serviced by a single terminal box. The layout of the four-zone building is shown in Fig. 4 (a). Each of the four rooms/zones has an equal floor area of $5\text{m} \times 5\text{m}$ and each wall is 3m tall, which provides a volumetric area of 75m^3 for each room. Room 1 has a small window (5m^2) on the north facing wall, whereas rooms 2 and 4 have larger windows (7m^2 each) on the east facing wall. Room 3 does not have a window.

The HVAC system used for simulation is designed to supply maximal mass flow rate of 0.25 kg/s per room. The supplied air temperature is fixed at $T^s = 12.8°\text{C}$. Here we assume that there is no return air and 100% of the outside air is sent to the AHU. The number of occupants in each room is uniformly generated as a random integer between 0 and 4. Outside temperature and outside solar radiation data is obtained for a summer day ($05/24/1996$) of Gainesville, FL [2]. Numerical results presented here are simulated using ode45 function in Matlab for 24 hours with the time step size chosen as 10 minutes. All temperatures are initialized at $24°\text{C}$, respectively. The mass flow rates entering four zones are given by $\dot{m}_1^{in} = 0.15$ kg/sec, and $\dot{m}_2^{in} = \dot{m}_3^{in} = \dot{m}_4^{in} = 0$ kg/sec. Fig. 5 shows the other two inputs: outside temperature T_0 and the heat gains $\dot{Q}_i$.

5.2 Recursive bi-partition of building graph

The RC-network representation of the four-zone building is shown in Fig. 4 (b). There are total 36 building nodes plus 1 outside node for the model of this four-zone building: 4



(a) (b)

Fig. 5. (a) The outside temperature T_0, and (b) the total heat gains Q_i, for room i ($i = 1, \ldots, 4$).

zone nodes $\{1, \ldots, 4\}$, 8 internal-wall nodes $\{5, \ldots, 12\}$, 8 internal-floor nodes $\{13, \ldots, 20\}$, 8 internal-ceiling nodes $\{21, \ldots, 28\}$, 8 external wall nodes $\{29, \ldots, 36\}$, and 1 outside node $\{37\}$. Each node is assigned with a thermal capacitance, and two adjacent nodes are connected with a thermal resistance. The windows are modeled as single resistors since they have relatively little capacitance. The values of capacitances and resistances used for simulation are obtained from commercially available software Carrier Hourly Analysis Program [1]. The outside node is assumed to have a very large capacitance $C_{37} = 10^{10}$ KJ/(m^{2}K).

The recursive bi-partition algorithm AlgoBIPA, described in Appendix B, is used to find sub-optimal partitions of the building graph based on the analysis of the linear thermal dynamics. The first iteration of the algorithm AlgoBIPA divides the node set into two groups: the first group contains all building nodes: $\{1, 2, \ldots, 36\}$, and the second group contains only the outside node: $\{37\}$. Such a 2-partition result makes sense since it captures the slowest time-scale of building thermal dynamics.

The second iteration of the algorithm AlgoBIPA leads to a 3-partition, which divides the nodes into three groups: the first group consists of all (zone, wall, ceiling, window, and floor) nodes associated with the room 3, the second group contains all other building nodes associated with the rooms $1, 2, 4$, and the third group contains only the outside node: $\{37\}$. Compared with the 3-partition results, the 4-partition identifies a new group containing all nodes associated with room 1. For the 5-partition, the algorithm returns five groups of nodes with clear physical intuition: group i contains all nodes corresponding to room i, for $i = 1, \ldots, 4$, and group 5 consists of the single outside node $\{37\}$! For $m > 5$, the m-partition further partitions the nodes associated with individual rooms. The largest possible m is 37, which corresponds to no reduction in model order.

Recall that the KL divergence rate (20) is used as a measure of the modeling error for aggregating the linear thermal dynamics. Fig. 6 depicts the KL divergence rate with respect to the number of partitions m for $2 \leq m \leq 37$. We observe from Fig. 6 that the modeling error monotonically decreases

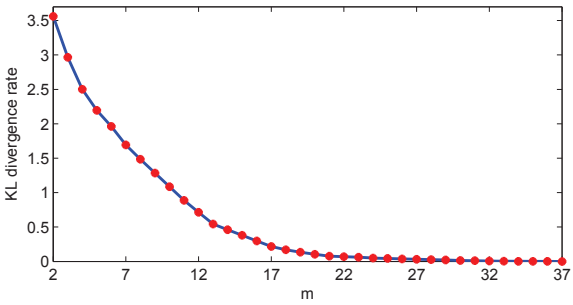


Fig. 6. Modeling error (KL divergence rate) in aggregating the linear thermal dynamics vs. number of partitions.

to the zero as the number of partitions increases to the dimension of the full-order model, and that there is little additional improvement beyond a model order of around 18. Although the KL divergence rate is only applicable to the linear part of the model, one can still use it as a conservative guideline for the reduction of the nonlinear model. In that case, we can guess that for good prediction accuracy, the reduced order model should have about 18 states. This is verified by simulations we report next.

5.3 Simulation of full and reduced-order models

The full-order model (2) is used to describe the full building thermal dynamics, with 36 building nodes plus 1 outside node. The multiple partition results obtained in Section 5.2 are used to construct the reduced-order models through aggregation of building nodes into groups, where each group of nodes is represented by a super-node. For $k = 1, \ldots, 36$, the kth-order reduced model (38) is used to describe the reduced building thermal dynamics with k super-nodes plus 1 outside node. For comparison, we lift the reduced model to one with 36 building nodes plus 1 outside node (see Section 4.1 for more details). That allows direct comparison between the temperature of a zone predicted by the full and reduced-order models.

All simulations reported here are open-loop simulations: the same mass flow rates ($\dot{m}_1^{in} = 0.15$ kg/sec and $\dot{m}_2^{in} =$

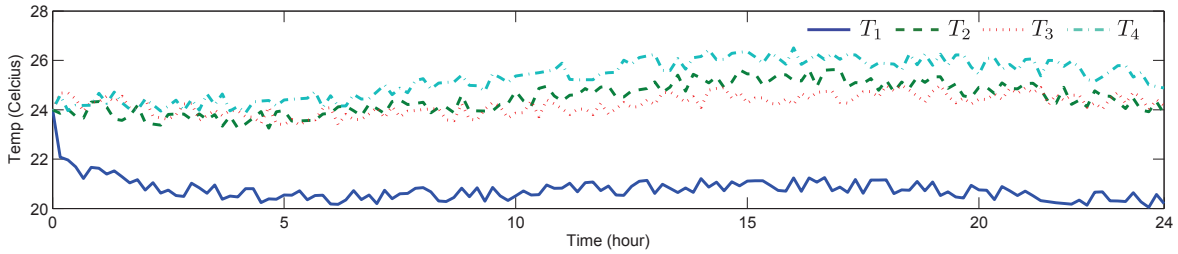


Fig. 7. Four zone temperatures $T_1, \ldots, T_4$ simulated by the full 36th-order model.

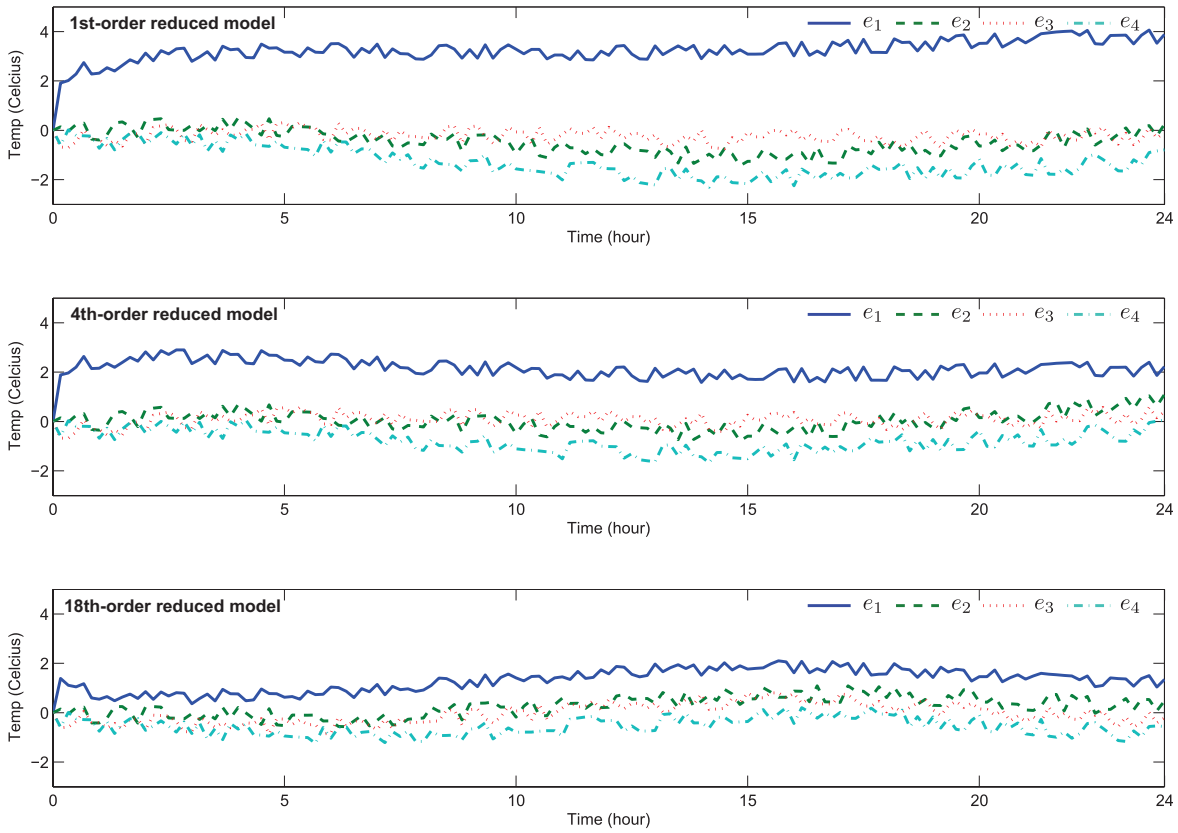


Fig. 8. Four zone temperature simulation errors are given by $e_i = \widehat{T}_i - T_i$ for $i = 1, \ldots, 4$, where T_i is the temperature simulated by the full-order model and $\widehat{T}_i$ is the lifted temperature simulated by the (1st-order, 4th-order, and 18th-order) reduced models.

$\dot{m}_3^{in} = \dot{m}_4^{in} = 0$ kg/sec) are used as inputs in conducting simulations for both full and reduced-order models; the inputs are shown in Fig. 5. Note that the inputs are aggregated accordingly to obtain the super-inputs for the reduced-order model (see Section 4.2 for more details). To test the goodness of the reduced-order models, we compare the four zone temperatures simulated by the full and reduced-order models. When simulated by the full-order model, the temperature of room i is denoted by T_i. When simulated by the reduced-order model, the lifted temperature of room i is denoted by $\widehat{T}_i$ (see Section 4.1 for more details). The ith zone temperature prediction error is denoted by $e_i = \widehat{T}_i - T_i$.

Fig. 7 shows the temperatures of the four zones predicted by the full-order model. Fig. 8 shows the temperature prediction errors corresponding to reduced-order models with varying degree of reduction: (i) 1st-order reduced model (1 super-node corresponding to all building nodes), (ii) 4th-order reduced model (4 super-nodes corresponding to 4 groups of nodes associated with 4 zones), and (iii) 18th-order reduced model. Note that a kth-order reduced model corresponds to the $(k+1)$-partition described in Section 5.2 with k super-building nodes and 1 outside node.

We observe from Fig. 8 that, as expected, prediction errors decrease as the order of the reduced model increases. In addition, the conjecture based on KL divergence rate that the 18th-order model will have predictions close to that of the full-order model turns out to be true. In the 18th-order model, the prediction error for the zone with the maximum error (zone 1 here) has a mean of 1.30°C and standard de-

viation of $0.46^{\circ}\mathrm{C}$. Note that even in a building that meets ASHRAE thermal comfort standards, the temperature inside a zone may vary by up to $3^{\circ}\mathrm{C}$ [6]. A lumped model that uses the well-mixed air assumption therefore is fundamentally limited to about a $1.5^{\circ}\mathrm{C}$ prediction error.

We also observe from the Fig. 8 that *except for zone 1*, temperature prediction with even the 4th-order model (middle plot), which represents a nine-fold reduction in model order, is quite accurate, where the prediction error for the zone with the maximum error (zone 4 here) has a mean of $-0.77^{\circ}\mathrm{C}$ and the standard deviation of $0.44^{\circ}\mathrm{C}$. However, the mean and standard deviation of prediction error for zone 1 with the 4th-order model are $2.14^{\circ}\mathrm{C}$ and $0.37^{\circ}\mathrm{C}$, which is much larger. Thus, large reduction in the model order is not likely to be useful for control design and analysis studies. However, we expect such low order models to be still useful in preliminary building and HVAC system design studies.

The higher error in the temperature prediction of zone 1 could be due to the the method's inability to accurately approximate enthalpy dynamics due to the ventilation (note that zone 1 is the only one with the ventilation), or due to the error introduced in lifting the reduced model to a full order model for purposes of comparison. Future work will examine these factors in greater detail.

The computation time for executing the Matlab simulation code increases as the order of the reduced model increases. For the case considered in this paper, the computation times for simulation are $6.829s$, $6.988s$, $7.623s$, $349.86s$ for the 1st-order, 4th-order, 18th-order, and full 36th-order models, respectively. In practice, one can make a tradeoff between the accuracy and complexity of the reduced order model by choosing an appropriate order of the reduced model.

6 Conclusions and Future Directions

We proposed a method to reduce the order of a building thermal model via aggregation of states. The original model is a large number of coupled nonlinear differential equations. Structurally, it is an RC-network with nonlinear terms due to ventilation air enthalpy. The heat conservation property of the system is used to obtain an analogy between the linear portion of the thermal dynamics and the continuous-time Markov chains. A recently developed aggregation-based model reduction technique for Markov chains can now be applied to the linear portion, with the associated KL divergence rate serving as a metric for the modeling error. Extension of the aggregation method to the nonlinear building thermal model is then carried out by aggregating inputs accordingly into the super-inputs.

A key advantage of the proposed aggregation technique over existing model reduction techniques is that it is structure-preserving by design. The reduced model is also a RC-network model with nonlinearities with the same structure as those in the original model. Thus, the reduced model directly corresponds to a building with smaller number of zones: groups of zones in the original building are now aggregated to form super-zones. The degree of reduction can be controlled by the user. This property makes the method applicable to not only real-time optimization-based control techniques and off-line control design and analysis, but also to parametric studies during building design phase.

Application of the proposed method to a four-zone building show that with with a reduction in the model order up to 50%, the maximum error in the time-domain predictions of temperature are quite accurate. With more reduction in the model order, time domain predictions are less accurate. However, for preliminary design studies such levels of error might still be acceptable, especially considering fundamental limits of prediction accuracy of any lumped model of temperature that relies on the "well-mixed air" assumption.

In the current work, the full-order model we consider does not have inter-zone convection effects. Inter-zone convection is difficult to model due to the complex physics that govern this phenomena. Recently, a data-driven identification scheme was proposed to obtain a RC-network model of convection among zones [21]. The model reduction method proposed here is immediately applicable if the full-scale model is augmented by such convection models. We are also developing the innovative control schemes based on the reduced building model proposed in this paper [13].

References

[1] *HAP Quick Reference Guide*. Carrier Corporation, Farmington, CT, 2003.

[2] National solar radiation data base (NSRDB), 2005.

[3] Energy future: Think efficiency. Technical report, the American Physical Society, September 2008.

[4] S. A. Al-Baiyat, M. Bettayeb, and U. M. Al-Saggaf. New model reduction scheme for bilinear systems. *International Journal of Systems Science*, 25:1631–1642, 1994.

[5] F. Alghimlas. Building load simulation and validation of an office building. In *Second International Conference for Enhanced Building Operations*, October 2002.

[6] American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. ANSI/ASHRAE standard 55-2010, thermal environmental conditions for human occupancy, 2010.

[7] ASHRAE. The ASHRAE handbook fundamentals (SI Edition), 2009.

[8] J. P. Bourdouxhe, M. Grodent, and J. Lebrun. *Reference Guide for Dynamic Models of HVAC Equipment*. American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc, Atlanta, 1998.

[9] J. E. Braun, K. W. Montgomery, and N. Chaturvedi. Evaluating the performance of building thermal mass control strategies. *Heating Ventillation Air Conditioning and Refrigeration Research*, 7(4):403–428, 2004.

[10] Q. Chen. Ventilation performance prediction for buildings: A method overview and recent applications. *Building and Environment*, 44(4):848–858, 2009.

[11] I. J. Couchman, E. C. Kerrigan, and C. Bohm. New model reduction scheme for bilinear systems. *Model reduction of homogeneous-in-the-state bilinear systems with input constraints*, 47(4):761–768, 2011.

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, NY, first edition, 1991.

[13] K. Deng, P. Barooah, and P. G. Mehta. Mean-field control for energy efficient buildings. In *Proceedings of American Control Conference*, Montréal, Canada, Jun 2012. submitted.

[14] K. Deng, P. Barooah, P. G. Mehta, and S. Meyn. Building thermal model reduction via aggregation of states. In *Proceeding of American Control Conference*, pages 5118–5123, Baltimore, MD, Jun 2010.

[15] K. Deng, P. G. Mehta, and S. P. Meyn. Optimal Kullback-Leibler aggregation via the spectral theory of Markov chains. *IEEE Trans. Automat. Contr.*, 56(12):2793–2808, 2011.

[16] M. M. Gouda, S. Danaher, and C. P. Underwood. Low-order model for the simulation of a building and its heating system. *Building services energy research technology*, 21(3):199–208, 2000.

[17] M. M. Gouda, S. Danaher, and C.P. Underwood. Building thermal model reduction using nonlinear constrained optimization. *Building and Environment*, 37(12):1255–1265, 2002.

[18] MM Gouda, S Danaher, and CP Underwood. Building thermal model reduction using nonlinear constrained optimization. *Building and Environment*, 37:1255–1265, 2002.

[19] Siddharth Goyal and Prabir Barooah. A method for model-reduction of nonlinear building thermal dynamics of multi-zone buildings. *Energy and Buildings*, 47:332–340, April 2012.

[20] Siddharth Goyal, Herbert Ingley, and Prabir Barooah. Zone-level control algorithms based on occupancy information for energy efficient buildings. In *American Control Conference*, June 2012. accepted.

[21] Siddharth Goyal, Chenda Liao, and Prabir Barooah. Identification of multi-zone building thermal interaction model from data. In *50th IEEE Conference on Decision and Control and European Control Conference*, pages 181–186, December 2011.

[22] M. S. Gray and J. Mesko. Energy functions and algebraic gramians for bilinear systems. In *Proceedings of the 4th IFAC Nonlinear Control Systems Design Symposium*, The Netherlands, 1998.

[23] J. Hahn and T. F. Edgar. An improved method for nonlinear model reduction using balancing of empirical gramians. *Computers & Chemical Engineering*, 26(10):1379–1397, 2002.

[24] Tianzhen Hong, Fred Buhl, and Philip Haves. EnergyPlus run time analysis, 2008. California PIER Project # 500-07-008.

[25] J.M. Horan and D.P. Finn. CFD reliability issues in the prediction of airflows in a naturally ventilated building. *International Journal of Ventilation*, 4:255–268, 2005.

[26] JH Kampf and D Robinson. A simplified thermal model to support analysis of urban resource flows. *Energy and buildings*, 39:445–453, April 2007.

[27] Mukesh K. Khattar and Hugh I. Henderson. Experiences with modeling supermarket energy use and performance. In *IEA Supermarket Refrigeration Workshop*, October 2000.

[28] S Lall, JE Marsden, and S Glavaski. A subspace approach to balanced truncation for model reduction of nonlinear control systems. *Journal of Robust and Nonlinear Control*, 12:519–526, 2002.

[29] K. H. Lee and J. E. Braun. Model-based demand-limiting control of building thermal mass. *Building and Environment*, 43:1755–1768, 2008.

[30] Y. Ma, G. Anderson, and F. Borrelli. A distributed predictive control approach to building temperature regulation. In *Proceedings of American Control Conference*, pages 2089 – 2094, San Francisco, CA, 2011.

[31] M Mossolly, K Ghalib, and N Ghaddar. Optimal control strategy for a multi-zone air conditioning system using a genetic algorithm. *Energy*, 34(1):58–66, Jan 2009.

[32] TR Nielsen. Simple tool to evaluate energy demand and indoor environment in the early stages of building design. *Solar Energy*, 78:73–83, January 2005.

[33] F. Oldewurtel, A Parisio, CN Jones, M Morari, D Gyalistras, M Gwerder, V Stauch, B Lehmann, and K Wirth. Energy efficient building climate control using stochastic model predictive control and weather predictions. In *American Control Conference*, pages 5100–5105, July 2010.

[34] Z. Rached, F. Alalaji, and L. L. Campbell. The Kullback-Leibler divergence rate between Markov sources. *IEEE Trans. Info. Thy.*, 50(5):917–921, 2004.

[35] D. Revuz. *Markov Chains*. Elsevier, New York, NY, second edition, 1984. North-Holland Mathematical Library, vol. 11, North-Holland, Amsterdam.

[36] H. Sandberg. An extension to balanced truncation with application to structured model reduction. *IEEE Trans. Automat. Contr.*, 55(4):1038–1043, 2010.

[37] J. Scherpen. Balancing for nonlinear systems. *Systems & Control Letters*, 21(2):143–153, 1993.

[38] J. E. Seem, S. A. Klein, W. A. Beckman, and J. W. Mitchell. Transfer functions for efficient calculation of multidimensional transient heat transfer. *Journal of Heat Transfer*, 111(1):5–12, 1989.

[39] M. Sourbrona, R. De Herdtb, T. Van Reetb, W. Van Passelb, M. Baelmansa, and L. Helsena. Efficiently produced heat and cold is squandered by inappropriate control strategies: A case study. *Energy and Buildings*, In Press, Accepted Manuscript:–, 2009.

[40] D. W. Stroock. *An Introduction to Markov Processes*. Springer, New York, NY, 2005. Volume 230 of Graduate texts in mathematics.

[41] Y. Sun and P. G. Mehta. The Kullback-Leiber rate pseudo-metric for comparing dynamical systems. *IEEE Transactions on Automatic Control*, 55(7):1585–1598, 2010.

[42] B. Tashtoush, M. Molhim, and M. Al-Rousan. Dynamic model of an hvac system for control analysis. *Energy*, 30(10):1729–1745, 2005.

[43] United States Energy Information Administration. Annual energy outlook, April 2011.

[44] S Wang. Dynamic simulation of building VAV air-conditioning system and evaluation of emcs on-line control strategies. *Building and Environment*, 36:681–705, November 1999.

[45] Xinhua Xu and Shengwei Wang. A simplified dynamic model for existing buildings using CTF and thermal network models. *International Journal of Thermal Sciences*, 47(9):1249–1262, 2008.

[46] Xinhua Xu, Shengwei Wang, Zhongwei Sun, and Fu Xiao. A model-based optimal ventilation control strategy of multi-zone VAV air-conditioning systems. *Applied Thermal Engineering*, 29(1):91–104, Jan 2009.

[47] M. Zaheer-uddin and G.R. Zheng. Optimal control of time-scheduled heating, ventilating and air conditioning processes in buildings. *Energy Conversion and Management*, 41(1):49 – 60, 2000.

A KL divergence rate for comparing Markov chains

Let (π, P) denote a discrete-time Markov chain defined on the state space $\mathcal{V} = \{1, \dots, n+1\}$ with the transition matrix P and the stationary distribution π. For two Markov chain models (π, P) and (ϖ, Q) defined on the *same state space* $\mathcal{V}$, the KL divergence rate is defined as:

$$R(P\|Q) := \lim_{N\to\infty} \frac{1}{N} D(\mathsf{P}_N\|\mathsf{Q}_N) \tag{A.1}$$

where $D(\cdot\|\cdot)$ denote the KL divergence between two distributions, and P_N (or Q_N) denotes the *joint probability distribution* defined on the cartesian product space $\mathcal{V}^N$ associated with the Markov chain model (π, P) (or (ϖ, Q)). We can write $R(P\|Q)$ more explicitly as

$$R(P\|Q) = \lim_{N\to\infty} \frac{1}{N} \sum_{i_0^{N-1}\in\mathcal{V}^N} \mathsf{P}_N(i_0^{N-1}) \log\left(\frac{\mathsf{P}_N(i_0^{N-1})}{\mathsf{Q}_N(i_0^{N-1})}\right).$$

In particular, we have a closed-form formula for the KL divergence rate defined in (A.1) (see [34]):

$$R(P\|Q) = \sum_{i,j\in\mathcal{V}} \pi_i P_{ij} \log(P_{ij}/Q_{ij}).$$

For model reduction problems, it is of interest to compare two Markov chains (π, P) and $(\bar{\pi}, \bar{P})$ defined on different state spaces $\mathcal{V} = \{1, \ldots, n+1\}$ and $\bar{\mathcal{V}} = \{1, \ldots, m+1\}$, respectively. Without loss of generality, we let $m \le n$. The relationship between $\mathcal{V}$ and $\bar{\mathcal{V}}$ is described by a partition function $\phi : \mathcal{V} \mapsto \bar{\mathcal{V}}$. The formula (A.2) can not be directly used to compute the KL divergence rate between (π, P) and $(\bar{\pi}, \bar{P})$. The strategy is to *lift* the Markov transition matrix $\bar{P}$ to another one defined on the state space $\mathcal{V}$ using a lifting distribution. As shown in [15], π is the optimal lifting distribution and the optimal lifted Markov transition matrix is given by

$$\widehat{P}_{ij}^{(\pi)}(\phi) = \frac{\pi_j}{\sum_{k\in\psi(j)} \pi_k} \bar{P}_{\phi(i)\phi(j)}, \quad i, j \in \mathcal{V}.$$

The KL divergence rate is then extended for two Markov chains defined on *different state spaces*:

$$R_\phi(P\|\bar{P}) := R(P\|\widehat{P}^{(\pi)}(\phi)).$$

B Algorithm AlgoBIPA for finding ϕ^*

The optimization problem (18) is an *integer nonlinear program* due to the fact that ϕ is integer valued and $R_\phi(P\|\bar{P}(\phi))$ is a nonlinear function of ϕ. The problem is non-convex on account of the non-convex constraints on ϕ. In general, it is prohibitively time-consuming to obtain the optimal solution of (18) for Markov chains with large state space. In this section, we summarize some heuristics to approach the optimal solution to (18).

In [15], we first consider the bi-partition ($m = 2$) problem for (18). After relaxing the integer constraints on the partition function ϕ, the optimization problem is shown to lead to a spectral partition associated with the following eigenvalue problem for reversible Markov chain:

$$Pu = \lambda u. \tag{B.1}$$

Let $u^{(2)}$ denote the eigenvector corresponding to the second largest eigenvalue of P. A sub-optimal bi-partition function ϕ^* is obtained by considering the sign-structure of $u^{(2)}$ (see [15] for more details):

$$\phi_i^* = \begin{cases} 1, \text{ if } u_i^{(2)} \ge 0 \\ 2, \text{ otherwise.} \end{cases}$$

A recursive bi-partition algorithm (AlgoBIPA) is also described in [15] to obtain the $m \ge 2$ partitions in a sub-optimal way: In the mth iteration of the algorithm, we assume that a partition with m groups (or super-states) is given. The objective of the mth-iteration is to obtain a refinement that has $(m+1)$ groups. For $i = 1, \ldots, m$, we denote by $P^{(i)}$ the sub-Markov transition matrix that describes the transition probabilities within the ith group. The ith group is split into two sub-groups according to the sign-structure of the second eigenvector for the eigenvalue problem associated with $P^{(i)}$. The spectral split of the ith group alone provides a partition of the states into $(m+1)$ groups. We denote this partition as $\phi^{(i)}$, and use it to evaluate the optimal aggregated transition matrix $\bar{P}(\phi^{(i)})$ according to (16). From the resulting m possible choices of $(m+1)$-partitions, we select the one that minimizes $R_{\phi^{(i)}}(P\|\bar{P}(\phi^{(i)}))$, i.e.,

$$i_{\min} = \arg\min_{i\in\{1,\ldots,m\}} R_{\phi^{(i)}}(P\|\bar{P}(\phi^{(i)})).$$

The associated aggregated transition matrix is chosen as $\bar{P}(\phi^{(i_{\min})})$.

The recursive algorithm is a heuristic based on the consideration of the bi-partition problem. A termination strategy can be based on a threshold value for the modeling error, e.g., the algorithm can be terminated if

$$\left|R_{\phi_m}(P\|\bar{P}(\phi_m)) - R_{\phi_{m-1}}(P\|\bar{P}(\phi_{m-1}))\right| \le \varepsilon_{\text{tol}}$$

where $\phi_m := \phi^{(i_{\min}^{(m)})}$, $\phi_{m-1} := \phi^{(i_{\min}^{(m-1)})}$, and ε_{tol} is the prespecified tolerated error.