Distributed model predictive control for building demand side management

DOI:
10.1109/ISGTEurope.2017.8260293

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
European Control Conference

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.
Distributed model predictive control for building demand side management

Alessandra Parisio, Salvador Pacheco Gutierrez

Abstract—Demand side management is widely acknowledged as an important source of flexibility and then an essential element to balance supply and demand more effectively. A fundamental challenge is to enable buildings to participate in demand side services without violating indoor comfort. In this paper, we present a distributed Model Predictive Control (MPC) approach to demand side management in buildings. The proposed MPC scheme encompasses heating/cooling systems, onsite generation and storage technologies, which are integrated into a unified management framework along with standard objectives (e.g., heating/cooling). The distributed algorithm, based on an active-set method, allows adjustments of multiple setpoints and enables the building to participate in balancing programs while minimising the energy use without violating the indoor comfort. Numerical results with real data from one university building show the promising performance and computational tractability of the proposed approach, which can enable practical implementations on building platforms.

I. INTRODUCTION

The next-generation energy grid and urban environment need to be smarter to deal with the growing energy demand and achieve environmental goals. In order to smarten the energy grid, significant research is to be devoted to buildings due to their large share of energy use, and the fact that they can increasingly integrate not just inelastic loads, but also distributed generation, storage resources and flexible loads. Demand for new sources of flexibility are raising interest in the interaction between energy sectors, like electricity, heating/cooling and gas, which, in combination with demand response and energy storages, can offer numerous benefits, e.g., providing flexibility to counteract the intermittency of renewable energy sources [1], [2]. Despite the recent advances in the area of energy management systems for intelligent buildings, significant effort is still required [2], [3]. Predictive automation systems are becoming available as expensive tailored configurations with limited functionalities (e.g., Ecopilot, GridEdge, Verdigris), which do not achieve the expected theoretical results, whilst holistic and standardised solutions are not available yet [4], [3]. This paper addresses the need of providing an overarching control framework that optimally coordinates heating/cooling systems, onsite generation, and storage, with the grid signal. Thus the building is enabled to support the grid for balancing generation and demand. We include in the control framework multiple thermal zones and the interaction among electricity, heating and gas systems through power-to-heat technologies, thermal energy storages and micro-CHP plants. In fact, studies [5] have shown that using electric power to produce heat (power-to-heat technologies, such as heat pump), often combined with heat storage, are a promising solution for increasing flexibility, which may contribute to both decarbonizing heat supply and integrating variable renewable electricity. We further account for the impact of weather condition and occupancy on the building thermal dynamics. The proposed control framework adopts a distributed approach, which is beneficial for implementation on real platforms.

Literature review: Among the various approaches adopted within the energy management literature, such as game theory and approximate dynamic programming, Model Predictive Control (MPC) [6] has received particular attention, because of its capability to integrate economic, social and environmental aspects, handle the future behavior of the system, compute control actions based on an optimal control problem including technical and operating constraints, as well as make the controlled system more robust against uncertainty [4], [7], [3]. Large-scale extensive simulation and experimental studies have shown that Model Predictive Control (MPC) can yield significant energy savings compared to conventional control strategies (e.g., [8], [9], [10]). It is expected that it will become a common solution for use in building energy management [7]. Decentralised and distributed MPC frameworks have been proposed for controlling Heating, Ventilation and Air Conditioning (HVAC) systems (e.g., [11], [12], [13]). The aforementioned studies either consider decoupled thermal zones ([11]) or apply the method of alternating direction method of multipliers (ADMM) ([13]). Authors in [12] propose a hierarchical economic MPC scheme, including a simple model of an active thermal energy storage without including losses. The high-level MPC solves a centralised problem considering an averaged thermal model of the building, whilst the low-level MPC controllers minimise the local energy use based on the cooling duties sent by the high-level MPC. Some studies have demonstrated the potential benefits of MPC frameworks also for buildings-to-grid applications, in particular for frequency service provision of HVAC systems (e.g., [14], [15], [16]). There are a few studies showing promising results in reducing costs and peak power of centralised MPC schemes for residential buildings including heat pumps [17], [18], which however simply model the thermal building dynamics considering one aggregated thermal zone. There are not studies in the literature targeting the design of advanced building control schemes extended both in terms of
assets under control, i.e., distributed generators and storage systems, and functions (ancillary services) [7]. Advanced and novel frameworks are needed to allow also residential users to play a more active role and must be implementable on real building management platforms.

Statement of Contributions: Designing efficient controllers for complex systems such as buildings is demanding, especially considering they can integrate also distributed generation, storage resources and flexible loads. MPC strategies offer a promising solution but generally come with computational issues, since they require solving large-scale optimisation problems in real time. An efficient solution is to distribute the computational requirements over multiple units, but this can lead to high communication costs. Building on the work of [19], [20], we propose a distributed MPC control framework that integrates and optimally coordinates demand response, onsite generation and energy storage. By doing so the building will be able to offer ancillary services to the grid operator without violating indoor comfort. The proposed MPC algorithm is based on a distributed implementation of the globally convergent active-set method for large-scale convex quadratic optimisation problems described in [19]. As traditional active-set methods, this method generates a sequence of subsets of inequality constraints which are active (i.e., they are satisfied with equality at the optimum). However, differently from traditional active-set methods, it exhibits significant improved convergence rates with local superlinear convergence. This is particular beneficial for large-scale control problems with communication delays, as building control applications [20]. The authors in [20] show that the algorithm described in [19] is implementable in a distributed fashion and can overcome also the aforementioned shortcoming of high communication costs. The application to building temperature control is thoroughly investigated: theoretical and experimental results demonstrate that the distributed implementation of the active-set method in [19] outperforms dual decomposition and ADMM especially in building control applications, requiring much smaller number of iterations to converge and significantly reducing the computation time. In this paper, we address the problem of coordinating and optimising demand response services, onsite generation and energy storage technologies along with along with more standard objectives (e.g., control of the indoor air quality). Building on the work in [20], we show that this problem can be formulated such that the active-set method in [19] can be applied and implemented in a distributed fashion. Furthermore, we propose a suitable initialisation algorithm.

Outline of the manuscript: Section II presents the model of the building and its onsite generation, while Section III describes the distributed MPC algorithm. Section IV provides and discusses numerical results, while Section V summarizes our conclusions and proposes future directions.

II. Modeling

In this section we outline the modeling of the main components considered in the proposed control framework.

Because of the scope of the paper and due to space constraints, we focus on the modeling of heating systems, thermal zones, micro-CHP and thermal energy storage (TES). The interested reader is referred to [21], [22], [23] and references therein for additional modeling details. We aim at using a control-orienting modeling framework, which can capture the main energy-related dynamics while keeping the number of states limited. We point out that the framework can be extended to include more complex thermal dynamics and additional distributed generators and flexible loads.

A. Nomenclature

Table I reports parameters and variables defined in the proposed control framework. In this study we consider discrete-time dynamical systems with a Δk sampling time. Power variables represent the average power over the given sampling period. For simplicity we omit the subscript denoting the time k in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>prediction horizon</td>
</tr>
<tr>
<td>N</td>
<td>number of thermal zones</td>
</tr>
<tr>
<td>P\text{gas}</td>
<td>gas power input to the building</td>
</tr>
<tr>
<td>P\text{grid}</td>
<td>power exchanged with the electrical network</td>
</tr>
<tr>
<td>P\text{gas,B}</td>
<td>boiler gas power input</td>
</tr>
<tr>
<td>P\text{gas,CHP}</td>
<td>micro-CHP gas power input</td>
</tr>
<tr>
<td>P\text{HP}</td>
<td>electrical power input to the heat pump</td>
</tr>
<tr>
<td>Q^\text{heat}</td>
<td>heat power input to thermal zone i</td>
</tr>
<tr>
<td>\eta^\text{CHP}</td>
<td>micro-CHP power efficiency</td>
</tr>
<tr>
<td>\eta^\text{HP}</td>
<td>micro-CHP heat efficiency</td>
</tr>
<tr>
<td>COP</td>
<td>coefficient of performance of the heat pump</td>
</tr>
<tr>
<td>C^\text{a}</td>
<td>air heat capacity</td>
</tr>
<tr>
<td>C^\text{s}</td>
<td>TES heat capacity</td>
</tr>
<tr>
<td>G^\text{a}</td>
<td>solar heat gain coefficient</td>
</tr>
<tr>
<td>A^\text{win}</td>
<td>windows area</td>
</tr>
<tr>
<td>Φ^\text{ext}</td>
<td>incident solar radiation on the windows of thermal zone i</td>
</tr>
<tr>
<td>C^\text{heat}</td>
<td>heat emission (per occupant)</td>
</tr>
<tr>
<td>N^i,people</td>
<td>number of occupants of thermal zone i</td>
</tr>
<tr>
<td>R^\text{a}</td>
<td>thermal resistance of thermal zone i</td>
</tr>
<tr>
<td>T^\text{a}</td>
<td>average temperature of the indoor air of thermal zone i</td>
</tr>
<tr>
<td>h^\text{a}</td>
<td>TES heat transfer coefficient</td>
</tr>
<tr>
<td>A^\text{TES}</td>
<td>TES exposed surface</td>
</tr>
<tr>
<td>T^\text{ext}</td>
<td>outdoor temperature</td>
</tr>
<tr>
<td>T^\text{amb}</td>
<td>temperature of the ambient where the thermal storage is placed</td>
</tr>
<tr>
<td>P\text{HP}, P\text{grid}</td>
<td>bounds on heat pump electrical power</td>
</tr>
<tr>
<td>P\text{gas,CHP}, P\text{grid,CHP}</td>
<td>bounds on micro-CHP gas power</td>
</tr>
<tr>
<td>T^\text{a}, T^\text{a}</td>
<td>bounds on TES temperature</td>
</tr>
<tr>
<td>C^\text{a}T^\text{a}<em>{k+1} = \frac{T^\text{ext}</em>{k} - T^\text{a}<em>{k}}{R^\text{a}} + C N^i,people + G A^\text{win} P^\text{w} + Q^\text{heat}</em>{k}.</td>
<td></td>
</tr>
</tbody>
</table>

B. Thermal zones

We consider multiple thermal zones, each having a dedicated local computing unit. The temperature dynamics of each thermal zone i can be modeled by using a resistance-capacitance circuit analogy

Preprint submitted to 2018 European Control Conference.
Received November 10, 2017.
Remark 1: \( \dot{Q}_{i,\text{heat}} \) can be expanded to specify the contributions coming from different units, such as the ventilation unit and floor heating (e.g., \( \dot{Q}_{i,\text{heat}} = \dot{Q}_{i,\text{floor}} + \dot{Q}_{i,\text{vent}} \)). Both the aspects mentioned above will be analysed in an extended version of the proposed framework under study.

C. Heat pump

The behaviour of a heat pump at each time \( k \) is modeled as
\[
\dot{Q}_{k}^{\text{HP}} = \text{COP} P_{k}^{\text{HP}},
\]
where \( \dot{Q}_{k}^{\text{HP}} \) is the heat generated by the heat pump at time \( k \) and \( \text{COP} = \eta \text{COP}^{\text{Carnot}} \). COP represents the technically feasible whilst the theoretically achievable COP is defined as \( \text{COP}^{\text{Carnot}} \), with \( \eta \) denoting a quality grade. In the numerical evaluation section we consider a ground source heat pump. We point out that the COP of air source heat pumps depends significantly on the ambient temperature, which entails that it would be more sensible to estimate forecasts of the COP at each time \( k \), based, for instance, on weather forecasts. These forecasts can be easily integrated in the proposed control framework.

Remark 2: For the sake of simplicity, in this study we discuss only the heating (winter) scenario. The proposed control framework can also easily include cooling by allowing for a negative \( \dot{Q}_{k}^{\text{HP}} \) and forecasts of the energy efficiency ratio \( \text{EER} = 1 - \text{COP} \).

D. Micro-CHP

The component represents a typical micro combined heat and power (micro-CHP) unit. The micro-CHP at each time \( k \) is described by the following equations: \( P_{k}^{\text{CHP}} = \eta P_{k}^{\text{CHP}} \) and \( \dot{Q}_{k}^{\text{CHP}} = \eta \dot{Q}_{k}^{\text{CHP}} \), where \( P_{k}^{\text{CHP}} \) and \( \dot{Q}_{k}^{\text{CHP}} \) respectively the electrical and heat power produced by the micro-CHP. The modeling can also adopt a data-driven approach, and relevant parameters can be estimated by employing machine learning technique with data from a real-world deployment, as in [21].

E. Boiler

A gas-based boiler is considered and modeled through its efficiency: \( \dot{Q}_{k}^{B} = \eta \dot{Q}_{k}^{B} \), where \( \dot{Q}_{k}^{B} \) is the heat produced by the boiler. Electric boilers can be likewise included in the proposed framework.

F. Thermal energy storage

The dynamics of the TES can be expressed as
\[
C_{A} T_{k+1}^{A} = \dot{Q}_{k}^{s} - \dot{Q}_{k}^{\text{load}} - h^{s} A^{s} (T_{k}^{v} - T_{k}^{\text{amb}}),
\]
where \( \dot{Q}_{k}^{s} = \dot{Q}_{k}^{\text{CHP}} + \dot{Q}_{k}^{B} + \dot{Q}_{k}^{\text{HP}} \) is the heat input to the storage, \( \dot{Q}_{k}^{\text{load}} = \sum_{i=1}^{N} \dot{Q}_{i,\text{load}} \), and the last term represents the heating loss depending on the temperature difference between storage and environment.

III. DISTRIBUTED MODEL PREDICTIVE CONTROL FOR BUILDING DEMAND SIDE MANAGEMENT

In this section we present our distribute MPC framework. We first outline the primal-dual active-set method proposed in [19], then we show that this method can be applied to the problem of tracking a reference power signal from the grid operator when a reserve call occurs.

Further, in III-C, we show how available flexibility to bid for balancing services can be optimised given a baseline building power.

A. Active set algorithm

In this section we outline the algorithm presented in [19] and adopted in [20].

Consider the following strictly convex quadratic problem
\[
\begin{aligned}
\min_{z \in \mathbb{R}^{l}} & \quad \frac{1}{2} z^T H z \\
\text{s.t.} & \quad F z = f \\
& \quad \bar{z} \leq z \leq \underline{z}
\end{aligned}
\]

where \( H \in \mathbb{R}^{l \times l} \) is a symmetric and positive definite matrix, \( F \in \mathbb{R}^{n \times l} \) is a wide matrix, with \( v \leq l \), \( f \in \mathbb{R}^{n} \) and \( \bar{z} \leq z \leq \underline{z} \) (which can take also infinite values). The matrix \( F \) is assumed to have full rank.

Problem 1 is strictly convex. Therefore, if the problem is feasible, the unique solution satisfies the following Karush-Kuhn-Tucker (KKT) conditions
\[
\begin{aligned}
H z + F^T \mu + \bar{X} - \lambda &= 0 \\
F z - f &= 0 \\
\min(\bar{X}, (z - \bar{z})) &= 0 \\
\min(\underline{X}, (z - \underline{z})) &= 0,
\end{aligned}
\]

where \( \mu \in \mathbb{R}^{n} \) are the dual variables corresponding to the equality constraints and \( \bar{X}, \underline{X} \in \mathbb{R}^{l} \) are the dual variables corresponding to the inequality constraints. The set of inequality and equality constraints indices are denoted respectively by \( \mathcal{L} = \{1, \ldots, l\} \) and \( \mathcal{V} = \{1, \ldots, v\} \). The subsets of indices corresponding respectively to the lower-active, upper-active, and inactive optimal primal variables \( z \) are defined as follows
\[
\begin{aligned}
\mathcal{A} &= \{ j \in \mathcal{L} : z_{j} = \bar{z}_{j} \} \\
\mathcal{U} &= \{ j \in \mathcal{L} : z_{j} = \underline{z}_{j} \} \\
\mathcal{I} &= \{ j \in \mathcal{L} : \bar{z}_{j} < z_{j} < \underline{z}_{j} \}.
\end{aligned}
\]

The subsets defined above represent a partition of the set of primal variable indices \( \mathcal{L} \).

One of the main novel features of the approach proposed in [19] with respect to the traditional active-set methods lies in the use of the index sets and the fact that multiple constraints are added to or removed from the active sets during each iteration of the algorithm. The values of the primal-dual variables \( z, \mu, \bar{X}, \underline{X} \) are uniquely determined by the index sets. The algorithm is initialised with a feasible partition of the index sets, \( (\mathcal{A}_{0}, \mathcal{U}_{0}, \mathcal{I}_{0}) \), which is updated at each iteration, along with the values of the primal-dual variables. The algorithm terminates when the active sets do not change from the previous iteration ([19], Theorem 1). At each iteration of the active-set algorithm, the active primal variables, denoted by \( z_{\mathcal{A}} \), are set to their corresponding bound, while the inactive primal variables, denoted by \( z_{\mathcal{I}} \), are free variables. The partition of the index sets must be feasible, i.e., there must exist some \( z_{\mathcal{I}} \) that satisfies the feasibility condition in (2)
\[
\mathcal{Z}_{\mathcal{I}} = \{ z_{\mathcal{I}} : F_{\mathcal{V},\mathcal{I}} z_{\mathcal{I}} = f - F_{\mathcal{V},\mathcal{A}} z_{\mathcal{A}} \}.
\]
The primal variables $z_T$ and the dual variables $\mu \in \mathbb{R}^v$ can be found at each iteration of the algorithm by solving the following subspace minimisation

$$
\begin{bmatrix}
H_{z,I} & F_{y,I} \\
F_{y,I} & 0
\end{bmatrix}
\begin{bmatrix}
z_I \\
\mu
\end{bmatrix}
=
\begin{bmatrix}
-H_{z,A}z_A \\
-F_{y,A}z_A + f
\end{bmatrix}
$$

The dual variables $\lambda_j$, with $j \in A$, and $\lambda_j$, with $j \in A$ are zero. The remaining dual variable are updated as follows:

$$
\begin{align*}
\lambda_j &= H_jz + (F'_j\mu_j), \\
\bar{\lambda}_j &= H_jz + (F\mu)_j,
\end{align*}
$$

where $H_j$ is the $j^{th}$ row of $H$. The active sets at each iteration $m$ of the algorithm are updates as follows:

$$
\begin{align*}
\bar{A}^{m+1} &= \{ j : (z_j > \bar{z}_j \text{ and } j \in I) \text{ or } (\bar{\lambda}_j > 0 \text{ and } j \in \bar{A}^m) \}, \\
A^{m+1} &= \{ j : (z_j < \bar{z}_j \text{ and } j \in I) \text{ or } (\bar{\lambda}_j > 0 \text{ and } j \in \bar{A}^m) \}, \\
I^{m+1} &= \{ j : j \notin \bar{A}^{m+1} \cup A^{m+1} \}.
\end{align*}
$$

The steps of the primal-dual active-set algorithm proposed in [19] are

**Algorithm 1 Active-Set Algorithm**

1: Find $(\bar{A}_0, A_0, I_0)$ and initialise $i = 0$
2: while $(\bar{A}^{m+1} \neq \bar{A}^m)$ and $(A^{m+1} \neq A^m)$ do
3: Solve (3)
4: Update dual variables using (4)
5: Update index sets partition using (5)
6: end while

**B. Distributed implementation for model predictive control**

Building on the work in [20], we illustrate a distributed implementation of the algorithm described above and propose a suitable initialization algorithm for our MPC problem.

We consider a star communication network, where we have a central node with more significant computing power, C-DG in Figure 1, and individual nodes, L-TZ in Figure 1, each corresponding to a thermal zone with local control inputs and limited computational resources. The thermal zones have dynamics coupled only through the control inputs, which are in our case the power setpoints to the local heating and ventilation systems, the onsite generation and TES. However the thermal coupling between thermal zones is commonly negligible, a non negligible coupling can be handled by considering it as a disturbance term, as shown in [20].

1) **Formulation of the MPC problem:** We now show how the problem of tracking a reference power profile at minimum energy use without violating the thermal comfort requirements can be formulated as (1). We first present the overall building model.

The dynamics associated to each individual node L-TZ$_i$ can be written in compact form as

$$
x_{k+1}^i = A^i x_k^i + B^i u_k^i + E^i w_k^i
$$

where the scalars $a^i, b^i$ and the row vector $E^i$ can be easily derived from Section II. In particular $x_k^i = T_s^i$, $u_k^i = Q_{\text{heat}}^i$ and $w_k^i = [T_{\text{ext}}, I^i, N^i, \text{people}^i]$.

The dynamics associated to the central node C-DG can be written in compact form as

$$
x_{k+1}^1 = a^1 x_k^1 + B^1 u_k^1 - 1_N u_k^{\text{zone}} + e^1 w_k^1,
$$

where the scalars $a^1, e^1$ and the row vector $B^1$ can be easily derived from Section II, $1_N$ is a row vector with $N$ 1's elements and $u_k^{\text{zone}}$ is the column vector of the stacked inputs to the thermal zones. In particular $x_k^1 = T^s$, $u_k^1 = [P_{\text{HP}}, P_{\text{gas},B}, P_{\text{gas},CHP}]$, $e^1 = T_{\text{amb}}$ and $B^1 = [\text{COP}, \eta^B, \eta^G]$.

At each time step $k$ we define $x_k := [x_k^1, \ldots, x_k^N, x_k^1]'$, $u_k := [u_k^{\text{zone}} u_k^1]'$ and $w_k := [w_k^1, \ldots, w_k^N u_k^1]'$.

Then the overall building model can be written in a compact form as

$$
x_{k+1} = Ax_k + Bu_k + Ew_k,
$$

where $A = \text{diag}(a^1, \ldots, a^N, a^1)$, $E = \text{diag}(E^1, \ldots, E_N, e^1)$ and

$$
B := \begin{bmatrix}
b^1 & 0 & \ldots & 0_{1 \times 3} \\
0 & \ddots & \vdots & \\
\vdots & \ddots & b^N & \\
-1 & \ldots & -1 & B^1
\end{bmatrix}
$$

**Equality and inequality constraints:** We now denote the state and input trajectories over the prediction horizon $T$ respectively by $x := [x_1', \ldots, x_T']$ and $u := [u_0', \ldots, u_{T-1}']$. We then define $z := [x \ u]'$.

The propagation of the system dynamics (8) over the prediction horizon $T$ generates the following equality constraints

$$
Fz = f,
$$

where

$$
F := \begin{bmatrix}
I & 0 & \ldots & 0 & -B & 0 & \ldots & 0 \\
-A & I & 0 & \ldots & -B & 0 & \ldots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & -A & I & 0 & \ldots & -B
\end{bmatrix}
$$

$$
f := \begin{bmatrix}
Ax_0 + Ew_0 \\
Ew_1 \\
\vdots \\
Ew_{T-1}
\end{bmatrix}
$$

**Fig. 1. Communication implementation for the proposed distributed implementation.**
We point out that \( \mathbf{F} \in \mathbb{R}^{T(N+1) \times T(2N+4)} \) is a wide matrix and has full rank, as required in (1), since \( \mathbf{B} \) has full row rank.

We also include polytopic constraints on states and inputs, 
\[ \mathbf{z} \leq \mathbf{z} \leq \mathbf{z}, \]
where \( \mathbf{z} \) and \( \mathbf{z} \) contain comfort ranges for the indoor temperatures and bounds on TES temperature and on the control inputs, as defined in Table I.

**Objective function:** The quadratic cost function we consider in order to minimise the energy use is
\[
J = \frac{1}{2} \mathbf{x}' \mathbf{S} \mathbf{x} + \mathbf{u}' \mathbf{R} \mathbf{u},
\]
where the diagonal and definite positive matrices \( \mathbf{S} \) and \( \mathbf{R} \) represents penalties on states and on inputs. We point out that the penalty on states will be lower than the penalty on the inputs. The Hessian matrix \( \mathbf{H} \) is given by \( \mathbf{H} = \text{diag}(\mathbf{S}, \mathbf{R}) \). Then the objective function can be re-written as \( J = \mathbf{z}' \mathbf{H} \mathbf{z} \). Hence, the problem of optimally managing the operation of the available energy resources in order to satisfy the comfort requirement at minimum energy use can be formulated as (1).

**MPC problem for building demand side management:**
We now assume that a reserve call occurs at the current point in time, \( k = 0 \). Hence a power signal, \( \mathbf{P}_{\text{grid}} \), is sent by the grid operator; this signal is to be followed over a previously agreed temporal window \( T \) at each \( k = 0, \ldots, T-1 \). We point out that the reference power signal is resulting from a previous contractual agreement between the flexibility service provider and the grid operator. The service provider (the building in this case) committed to be available during a selected availability temporal window and offer a certain amount of flexibility. In III-C we propose a possible formulation of the problem to compute the optimal amount of afforded flexibility with respect to a given baseline building power.

The objective function (10) and the constraints (9) have to be modified in order to account for the reference power signal to track. Therefore, we include an additional inequality constraint for each time \( k \)
\[
P_{\text{grid}}^r - \epsilon_k \leq P_{\text{grid}} \leq P_{\text{grid}}^r + \epsilon_k,
\]
where \( P_{\text{grid}}^r - \epsilon_k \) is the electrical power exchanged with the grid and \( \epsilon_k \) is a vector of stacked \( \epsilon_k \) over \( T \) and define \( P_{\text{grid}} := \begin{bmatrix} z' \epsilon_{\text{grid}} \end{bmatrix} \). Hence, constraints (12) can be further re-written as \( K_{\text{grid}} z_{\text{grid}} \leq k_{\text{grid}} \), where \( K_{\text{grid}} \) is a matrix of appropriate size and \( k_{\text{grid}} \) is a column vector containing stacked positive and negative values of the reference power signal \( P_{\text{grid}}^r \).

We can accordingly modify matrix \( \mathbf{H} \) such to include the penalty on the deviations from the reference signal, \( \mathbf{R}_{\text{grid}} \), as follows
\[
H_{\text{grid}} := \text{diag}(\mathbf{S}, \mathbf{R}, \mathbf{R}_{\text{grid}}).
\]
As shown in [20] (Theorem 2), in order to apply Algorithm 1 we have to convert the additional inequality constraints (12) into equality constraints by introducing slack variables \( \sigma \)
\[
K_{\text{grid}} z_{\text{grid}} + \sigma = k_{\text{grid}}
\]
\[
\sigma \geq 0.
\]
The vector of decision variables has to include also the slack variables but this implies that \( H_{\text{grid}} \) is positive semi-definite and not positive definite. In order to address this issues, as proposed in [20], we add a Lagrangian penalty to the cost function \( J \) for the introduced equality constraints in (13). The Hessian of \( J \) is accordingly modified as follows
\[
H_{\text{aug}} := \begin{bmatrix} H_{\text{grid}} + \rho K_{\text{grid}}' K_{\text{grid}} & K_{\text{grid}}' & K_{\text{grid}}' \end{bmatrix}
\]
\[
\begin{bmatrix} K_{\text{grid}} & K_{\text{grid}} & K_{\text{grid}} & I \end{bmatrix}
\]
which is positive definite. The penalty factor \( \rho \) is greater or equal to 1.

Our MPC problem to be solved at each point in time, based on real measurements from the building and updated forecasts of disturbances, can be formulated as the following strictly convex quadratic problem
\[
\begin{array}{ll}
\text{minimize} & \frac{1}{2} z_{\text{aug}}' H_{\text{aug}} z_{\text{aug}} \\
\text{subject to} & F_{\text{aug}} z_{\text{aug}} = f_{\text{aug}} \\
& z_{\text{aug}} \leq z_{\text{aug}} \leq z_{\text{aug}}
\end{array}
\]
where \( z_{\text{aug}} \) is the vector \( z_{\text{grid}} \) stacked with the slack variables \( \sigma \), \( F_{\text{aug}} \) and \( f_{\text{aug}} \) are respectively the matrix \( F \) and the
column vector \( f \) from (9) appropriately modified to include the introduced decision variables \( \epsilon_{\text{grid}} \) and \( \sigma \). The polytopic constraints representing bounds on all the decision variables are also accordingly augmented.

We point out that it is possible to tune the tradeoff between the energy use and the reference tracking by appropriately tuning the penalty on the inputs and on the variables representing the deviation from the reference power signal. Further, the MPC problem can be easily extended to soften the thermal comfort constraints and increase the available flexibility.

2) Distributed implementation: The distributed implementation of the active-set algorithm is described in Algorithm 2. Notice that the dual variables and the active sets corresponding to TES and onsite generation are updated in the central node C-DG (see Figure 1). We propose a simple

\[ u_k^{\text{grid}} = \frac{\left( \text{COP} P_k^{\text{grid}} - P_k^{\text{grid}} \right)}{\eta_k}, \quad u_k^{\text{grid}} = \frac{\left( \text{COP} P_k^{\text{grid}} - P_k^{\text{grid}} \right)}{\eta_k} \]

Algorithm 2 Distributed Active Set Algorithm

1: Each node L-TZ, and the central node C-DG initialise their active sets according to Algorithm 3
2: \text{while} \( (A^{m+1} \neq A^m) \) and \( (A^{m+1} \neq A^m) \) \text{do}
3: Each node L-TZ sends current active sets to C-DG
4: C-DG solves (3)
5: C-DG sends \( x_I = \epsilon_{\text{grid}} \) and \( \mu \) to nodes L-TZ
6: Each node L-TZ, and C-DG update their corresponding dual variables and active sets by using (4) and (5).
7: \text{end while}

initialisation algorithm for our MPC problem, summarised in Algorithm 3. In the first step of Algorithm 3, \( u_k, \forall k \), are computed as follows

\[
\begin{align*}
\text{if } P_k^{\text{grid}} < 0, & \quad P_k^{\text{gas,CHP}} = P_k^{\text{grid}}, & u_k = u_k^{\text{grid}} \\
\text{if } P_k^{\text{grid}} \geq 0, & \quad P_k^{\text{grid}} = \text{COP} P_k^{\text{grid}}, & u_k = u_k^{\text{grid}}(15)
\end{align*}
\]

Algorithm 3 Initialisation Algorithm

1: Set \( u_k, \forall k \), using (15) and \( \epsilon_{\text{grid}} = 0 \)
2: Each node L-TZ obtains the states, \( x_i, \forall k \), using (6)
3: C-DG obtains the states, \( x_i, \forall k \), using (7)
4: \( I_0 = \{ j : x_j > x_j^{\text{min}} \} \) \cup \{ j : x_j > x_j^{\text{max}} \}
5: \( I_0 = \{ j : x_j < x_j^{\text{min}} \} \) \cup \{ j : x_j < x_j^{\text{max}} \}
6: \( I_0 = \{ j : j \notin A_0 \cup A_0 \} \)

**Proposition 1:** Consider Problem (14) and assume it is feasible. Then Algorithm 3 satisfies the feasibility condition \( Z_I \neq \emptyset \)

**Proof:** Matrix \( F_{\text{aug}} \) is still a wide matrix, since we augmented \( F \) with 2\( T \) variables and included 2\( T \) equality constraints. This matrix will be given by \[ \begin{bmatrix} F & 0 \\ \text{K}_{\text{grid}} & I \end{bmatrix} \]. Algorithm 3 will remove some columns from the first \( T(N+1) \) columns of \( F_{\text{aug}} \), which corresponds to the active elements of \( x \). The matrix \( F_{\text{aug},I} \) will still include the block diagonal matrix \( B \) and the identity matrix \( I \), hence it will be full row rank. Therefore, \( Z_I \neq \emptyset \)

C. Optimisation model for balancing services

We assume that building energy users can participate in different balancing services as flexibility providers. They can be called to: i) either reduce their generation or increase their demand when there is an excess of energy in the system (denoted by \( BS_{\text{UP}} \)); ii) either increase their generation or decrease their demand when there is a deficit of energy (denoted as \( BS_{\text{DOWN}} \)). The providers can bid on each of the two balancing services during specific availability windows established by the grid operator.

The following optimisation problem can be used to compute the afforded flexibility, based on the baseline power and heat consumption and the economic incentives offered by the grid operator for each of the two balancing services (denoted by \( c_{\text{UP}} \) and \( c_{\text{DOWN}} \) respectively)

\[
\min \sum_{k=0}^{T-1} \left( -c_{\text{UP}} \Delta P_k^{\text{UP}} + c_{\text{DOWN}} \Delta P_k^{\text{DOWN}} + c_{\text{gas}} \Delta P_k^{\text{gas}} \right)
\]

s.t.

\[
\begin{align*}
\Delta T_{k+1}^{\text{grid}} &= \Delta T_k^{\text{grid}} - \sum_{i=1}^{N} \Delta Q_i^{\text{heat}} + \Delta T_k^{\text{Heat}} - T_{\text{BL}} \\
\Delta Q_k^{\text{heat}} &= \text{COP} P_k^{\text{grid}} + \eta_k B P_k^{\text{gas,B}} + \eta_k^{\text{CHP}} P_k^{\text{gas,CHP}} \\
\Delta P_k^{\text{grid}} &= \Delta P_k^{\text{grid}} + \Delta P_k^{\text{gas,B}} + \Delta P_k^{\text{gas,CHP}} \\
\Delta P_k^{\text{grid}} &= \Delta P_k^{\text{grid}} - \eta_k B P_k^{\text{gas,B}} - \eta_k^{\text{CHP}} P_k^{\text{gas,CHP}} + \sum_{r=1}^{N} \gamma_r \Delta Q_r^{\text{heat}} \\
T_k^{\text{DOWN}} - \Delta T_k^{\text{grid}} &\leq T_k^{\text{UP}} - \Delta T_k^{\text{UP}} \\
-\Delta P_k^{\text{DOWN}} &\leq \Delta P_k^{\text{UP}} - \Delta P_k^{\text{UP}} \\
\Delta P_k^{\text{grid}} &\leq \Delta P_k^{\text{grid}} - \Delta P_k^{\text{grid}}
\end{align*}
\]

where \( T_k^{\text{BL}} \) is the TES temperature at each time \( k \) resulting from baseline power and heat consumption, \( \Delta Q_k^{\text{heat}} \) and \( \Delta Q_k^{\text{heat}} \) are the acceptable flexibility in the heat requirements of each thermal zone \( i \) at each time \( k \) such that the indoor comfort is not violated. The binary numbers \( T_k^{\text{UP}} \) and \( T_k^{\text{DOWN}} \) model whether or not the availability window for \( BS_{\text{DOWN}} \) and \( BS_{\text{UP}} \) is active at time \( k \), with \( T_k^{\text{UP}} + T_k^{\text{DOWN}} \leq 1 \) for each \( k \) (i.e., the grid operator does not ask the flexibility provider for both services at the same time). The optimal values of \( \Delta P_k^{\text{grid}} \), \( \forall k \), represent the afforded flexibility, i.e. optimal deviations from the baseline power such that the profits are maximised and the thermal comfort is satisfied. The baseline power and heat consumptions and the bounds in Problem (16) can be estimated out of historical data [24].

IV. NUMERICAL EVALUATION

In this section we present the numerical evaluation of the Algorithm 2 presented in Section III. The active-set method was applied to a university building consisting of several rooms and laboratories. The 4 story building is equipped with three air handling units (AHU) located on the top floor and supplying ventilation to different areas of the building according to their use, i.e. laboratories, common areas and pilot plants. For each one of the AHUs, the supplied air is
heated by means of hot water coils using a ground source heat pump (GSHP) and backed up by a set of two boilers. The variable-air-volume (VAV) terminals supply the air to the thermal zone under control, which in this case, is composed of several rooms and spaces.

A. Simulation Setup

The building and its control logic was implemented using EnergyPlus (E+) and Sketchup plugin. The real and simulated building can be observed in Fig 2. The communication between EnergyPlus and MATLAB was conducted using the MLE+ library through the Building Control Virtual Test Bed (BCVTB). The model developed in E+ was validated against real building data.

First, the internal temperature computed with the E+ model of each thermal zone was compared against real measured data. Weather data of actual building location was utilised. In Figure 3 the real and simulated internal temperature profiles of one of the thermal zones are depicted for 7 days in March 2016. The model generated utilising E+ is capable to follow quite accurately the dynamics of the internal temperature: the error does not exceed 0.4°C. Similar results were obtained for the remaining thermal zones.

![Fig. 2. James Chadwick building: a) real building b) SketchUp/Openstudio model](image)

As mentioned above, the heat is provided by the GSHP and two boilers. This system consists of a water tank acting as a thermal storage, connected to the GSHP and the boilers loops. The logic of the system is simple and consists in maintaining the temperature of the hot water in the thermal storage within a given range.

The GSHP and the thermal storage are then assigned to the central node C-DG (Figure 1). Hence, the temperature of the water in the tank is modelled as $T^s$ in (7), while the input vector is $u^s = [P_{HP} \ P_{gas,B}]^T$. The elements of the input vector can be seen as the aggregated power consumption of the heat pump (HP) and the boilers (B). Since there is not micro-CHP utilised in the building, $u^s$ contains only two elements.

The dynamics of each individual node $L_{- TZ_i}$, corresponding to a thermal zone, can be represented by (6). Thus, the heat flow required by the AHU of each thermal zone, $u^i = Q^i_{heat}$, is provided by the heat pump, $Q^{HP}_k = \text{COP} \ P_{HP}^k$, and the boilers, $Q^B_k = \eta^B \ P_{gas,B}^k$, through the thermal storage.

B. Result Evaluation

The algorithm was implemented in MATLAB utilising the MA57 LDL solver to perform the subspace minimisation, as suggested in [20]. For the numerical experiment conducted for 22nd March 2016, the maximum computation time of the MPC iterations is 2.035 seconds, which is basically the time required to solve the subspace minimisation on a Windows system Core i7 @ 2.80 GHz.

Since the micro-CHP is not included, we focus on the case with $P_{s-grid} > 0$. In order to obtain the reference power profile, Problem (16) is solved, where historical data from the day before the grid operator call are considered as baseline heat and power consumption. Forecasts of the disturbances are included based on E+ and weather forecasts. The heat inputs to each thermal zone and the power setpoints to the GSHP and the TES are computed by the proposed Algorithm 2 at each time step. Algorithm 3 is used to initialise the partition of the index sets. The obtained inputs and power setpoints are then applied to the E+ model in order to obtain the actual power demand required by the university building, based on the actual disturbances on 22nd March 2016.

The zone temperature comfort range is set between 19°C and 23°C. The bounds on the water temperature of the thermal storage are set to 60°C and 70°C. The sampling time is 15 minutes and the prediction horizon is 5 hours.

![Fig. 3. Comparison of real and modelled internal temperature using EnergyPlus](image)

![Fig. 4. Comparison between the actual power demand and reference power profiles](image)

Figure 4 depicts the reference power signal and the power profile resulting from the application of the Algorithm 2. The power profile is mainly due to the fact that the pumps that form part of the GSHP can drastically change the flow rate in the hot water loop by being turned on and off.

We can notice that the reference power profile is reasonably followed by the actual power profile demanded by the building in order not to violate the internal thermal comfort. The actual power demand could be lower (see Figure 4 between 5:00 and 5:00) or higher (see Figure 4 between 12:00 and 13:00) depending on the actual disturbance profiles and heating requirements from the thermal zones. We recall...
that the comfort bounds are not softened and the tracking results depend on the penalty on the decision variables representing the deviation from the reference signal, \( \epsilon_{\text{grid}} \). In this case, the comfort has priority and the building is not able to exactly follow the previously computed reference power profile. The tracking results can be improved by varying the penalty on the control inputs.

Furthermore, the updating strategy for finding a feasible partition of the index sets at each iteration of the algorithm used in this paper, however extremely efficient, can lead to cycling, especially when applied to solve ill-conditioned problems. We are currently investigating alternative updating strategies, which require slightly additional computational effort, but can guarantee the global convergence in any case \([19]\).

V. CONCLUSIONS AND FUTURE STUDIES

In this paper, we propose a distributed Model Predictive Control (MPC) framework that tracks a reference power signal sent by the grid operator while minimising the energy use without violating the indoor thermal comfort.

Building on the work in \([19], [20]\), we propose a distributed implementation of an effective primal-dual active-set method, which requires smaller number of iteration to converge with respect to other methods. Furthermore, we present a suitable initialisation algorithm. Practical implementation can benefit from the distributed approach of the proposed control framework. Numerical results based on real measurements from an actual university building show promising results. A more extensive numerical evaluation of the resented control framework is under study.

Future works will include the extension of the modeling to include more complex building dynamics and additional onsite generation, as well as the extension of the control framework to incorporate statistics of the disturbances, based on historical data. Another relevant direction is to investigate suitable tuning strategies of penalty on the inputs, in particular the tradeoff between the energy use and the tracking of the reference power signal.

REFERENCES


