Offset-Free Model Predictive Temperature Control for Ultrasound-Based Hyperthermia Cancer Treatments


Abstract—Heating cancer cells over an extended period of time, referred to as hyperthermia, has been proven to enhance the effects of chemotherapy and radiotherapy without inducing additional toxicity or undesirable side effects, and is therefore considered a highly valuable adjuvant therapy in cancer treatment. In this work, a model predictive control (MPC) setup is developed for improving performance and robustness in regulating the temperature for magnetic-resonance-guided high-intensity focused ultrasound (MR-HIFU) hyperthermia treatments. The proposed control design incorporates a disturbance estimator as encountered in offset-free MPC that is able to remove the steady-state temperature error caused by plant-model mismatch. For the considered healthcare application, such modeling errors are inevitable in practice due to the high variability of tissue properties in patients, some of which even exhibit time- and temperature-dependent behavior due to the body’s thermoregulatory response, combined with the fact that extensive model identification is undesirable in the clinic. The controller’s performance is demonstrated by means of in vivo experiments on a porcine thigh muscle using a clinical MR-HIFU treatment setup.

Index Terms—High-intensity focused ultrasound, hyperthermia, model predictive control (MPC), offset-free control, oncology.

I. INTRODUCTION

MILD local hyperthermia involves the heating of a specific target volume inside the body, typically containing the tumor, to temperatures of 39 °C–45 °C for up to about 90 min while preserving nonelevated temperatures in healthy tissue. Clinical trials have provided ample evidence that hyperthermia is a potent adjuvant therapy for cancer treatment (see [1]–[3] and the references therein). It locally sensitizes the treated tissue to the effects of chemotherapy and radiotherapy while leaving the untreated (healthy) tissue unaffected. Consequently, similar or improved treatment effectiveness can be achieved using lower doses of radiation and drugs, thereby reducing the severity of the negative side effects typically associated with cancer treatment [4], [5]. Moreover, using temperature-sensitive liposomes, it enables localized heat-mediated delivery of anticancer drugs, allowing for a further reduction of the required systemic drug concentrations [6], [7]. Most importantly, hyperthermia itself is nontoxic and therefore introduces no additional toxicity-related side effects, making it highly appealing for clinical application to increase treatment success rates and improve quality of life for the patients.

A particularly well-suited technology for hyperthermia is magnetic-resonance-guided high-intensity focused ultrasound (MR-HIFU). This entails the combined use of powerful and millimeter-accurate heating using ultrasound waves and real-time volumetric thermometry using an MRI scanner [8], [9]. Using an external HIFU applicator, this technology allows for a completely noninvasive treatment, which is highly desirable for patient comfort, and eliminates the need for postinterventional wound care. However, realizing the desired temperature distribution and accurately maintaining it over the course of an entire treatment in a clinical setting is no simple task, resulting in a need for the development of adequate feedback controllers. This is particularly true for mild hyperthermia since its beneficial effects have been found to be strongly correlated with the tissue temperatures truly achieved during treatment [10]–[12].

Currently, most implementations of MR-HIFU for temperature control in thermal therapies use predetermined sonication plans, possibly extended with simple feedback controllers that make (minor) online adjustments. Examples include binary strategies to scale the sonication power and/or length of the heating intervals in the sonication protocol [13], [14], ad hoc proportional–integral–derivative (PID)-based methods [15], [16], or some hybrid form of PID and bang-bang control [17]. In addition to not exploiting the full potential control freedom offered by the setup, the major drawback of such designs is their inability to take the body’s future thermal...
behavior and the restrictive actuator constraints into account when computing the control inputs, which negatively affects treatment quality and duration.

In this respect, we believe that model predictive control (MPC) offers superior potential for temperature control in hyperthermia treatments. This has already been recognized by other researchers, see, for example, [18] for the control of a scalar thermal dose parameter based on a single-point or 1-D model, [19] (using heavily simplified models and a fixed sonication trajectory) and [20] for thermal dose control in 2-D systems, and more recently [21] for temperature control in 2-D systems with online adaptation of the heating location and power.

These works illustrate clearly the potential of MPC for hyperthermia. However, for accurate temperature control in mild hyperthermia treatments in a clinical setting, a more sophisticated MPC design is required. One important reason for this is the large variation observed in the tissue properties in reality, which are typically patient- and tumor-specific, and can additionally vary dramatically as functions of time, space, and temperature [22]–[24]. Extensive pre-treatment model identification is undesirable from a clinical perspective since, besides improving treatment quality, the goal is also to reduce treatment time and avoid unnecessary heating. Consequently, model discrepancies are typically inevitable in practice and may result in significantly deteriorated control performance. It is the objective of this article to provide a significant step forward in the development of an MPC setup that is able to robustly regulate the temperature in the tumor over a prolonged period of time, despite the presence of considerable and possibly varying model uncertainty.

To this end, as the main contribution of this work, we develop a temperature controller for MR-HIFU hyperthermia based on an offset-free MPC approach. Using a disturbance model, the effects of constant and slowly varying plant-model mismatch can be captured [25], [26]. An observer is employed to identify the corresponding disturbance while simultaneously providing temperature estimates that improve upon the noise-corrupted MR thermometry measurements. This enables the MPC to compensate for the model-error-induced temperature offset in steady state, recovering the optimal performance as achievable in case of zero plant-model mismatch, thereby allowing for significantly enhanced temperature control and treatment quality.

In [27], we have presented a preliminary version of this control setup. Using simulations, it was shown that the proposed solution is able to identify and compensate for model discrepancies. In this work, we present results obtained in porcine in vivo experiments, demonstrating the developed feedback setup’s ability of achieving desired steady-state heating despite the presence of realistic plant-model mismatch in a clinical setting. Furthermore, to improve modeling accuracy without increasing computation time, in this article, we extend the modeling procedure for the controller and observer separately, additionally accounting for computation/communication delays to reduce modeling error in the temporal discretization of the state-space models. Finally, more details are provided regarding the design considerations of the cost function, which has been changed with respect to [27].

The remainder of this article is organized as follows. First, Section II discusses the MR-HIFU hyperthermia setup and treatment, motivating the development of offset-free MPC for this application. In Section III, we describe the thermal model and its spatial discretization, which is used in the observer and controller design presented in Section IV. In Section V, the results of the in vivo experiments are discussed to illustrate the algorithm’s potential for clinical application. Finally, Section VI summarizes the key achievements and the corresponding observations.

II. SYSTEM AND TREATMENT DESCRIPTION

In this section, we introduce the hyperthermia treatment and setup and motivate the use of MPC in this application.

A. MR-HIFU Hyperthermia Setup

Although the MPC setup we propose is generic in nature, in this article it is designed for a clinical MR-HIFU system consisting of a Profound Sonalleve, shown in Fig. 1, and a Philips 3T Achieva. The former is a dedicated trolley-tabletop in which an MR-compatible HIFU transducer is integrated, and the latter is an MRI scanner which we use for noninvasive near-real-time thermometry. This system is already being used in clinics to noninvasively treat uterine fibroids and for incision- and radiation-free palliative treatment of pain associated with bone metastases. A custom software layer based on [21] has been developed to connect the MPC algorithms designed in this work to the MR-HIFU setup.

B. HIFU Applicator

This system uses a phased-array HIFU transducer to generate the ultrasound waves. It consists of 256 elements of which the phases and amplitudes can individually be chosen such that by interference a focal spot is created, thereby enabling powerful and millimeter-accurate heat delivery to internal tissues, as depicted in Fig. 2. By modulating these settings, which is referred to as electronic beam steering, the focal spot can be repositioned to up to 20 distinct locations per second within a 16-mm-diameter circle around the transducer axis. By moving the focal spot through the tumor area, all
of the cancerous tissue can be heated. Treating larger regions additionally requires mechanically displacing the transducer (see [14] for example). In this article, however, we limit ourselves to the development of a controller for the treatment of small tumors and will therefore consider electronic beam steering only.

The focal spot has an axial length of approximately 7 mm, which is significantly larger than its narrow radial width of about 2 mm, as depicted in Fig. 3. Here, it is also shown how for different focus locations in the focal plane, the acoustic beams (each individually indicated in light gray) significantly overlap each other before and behind the focal plane in the direction of the beam axis (resulting in darker gray in the figure). Consequently, the resulting temperature distribution is fairly homogeneous in the axial direction near the focal plane, whereas the in-plane temperature gradients may be large. Since in this work we consider only small treatment volumes, the temperature map in the focal plane characterizes the tissue temperature in the entire treatment volume, allowing us to design our MPC setup for temperature control only in the focal plane using 2-D models. Note that, additionally, all (constant and slowly varying) neglected out-of-plane effects, e.g., heat conduction in case of a nonzero axial temperature gradient, will be identified and compensated for by the offset-free algorithm.

### C. Motivation for Offset-Free MPC in Hyperthermia

The key to successful application of mild local hyperthermia is to maintain a steady temperature elevation above 41 °C in the region of interest (ROI), typically the tumor and some adjacent tissue, during the entire treatment. The temperature sweet spot for treatment quality is at 42 °C. This is required to fully benefit from the desired heat-induced effects, such as increased blood perfusion, aiding drug delivery to otherwise poorly perfused parts of the tumor when combined with chemotherapy [1], [28], or the inhibition of DNA repair mechanisms during the crucial time window after radiation therapy [4], [12]. On the other hand, overheating can also be detrimental to successful treatment since some of these mechanisms exhibit reversal effects at higher temperatures (above 43 °C) [28], [29]. Furthermore, temperature elevations above 41 °C outside the ROI must be avoided to prevent sensitization of (or damage to) healthy tissue.

Compared to existing binary or PID-based strategies applied in MR-HIFU hyperthermia, we believe that MPC-based approaches can deliver superior closed-loop properties, as they are able to achieve faster and more uniform heating by exploiting beneficial (future) behavior, e.g., heat transfer by conduction, and can explicitly consider actuator constraints, such as the inability to actively remove heat from inside the body using HIFU. Unfortunately, model-based strategies are also inherently accompanied by the possibility for modeling errors.

As discussed in [24], accurate thermal and thermoregulatory modeling for hyperthermia treatments is especially difficult due to the high variability of the spatially distributed and time/temperature-varying tissue properties [22], [23]. In addition, attempting to capture all these effects using extensive personalized pre-treatment model identification is highly undesirable in the clinic, where besides improving treatment quality, the aim is also to reduce treatment times and to improve safety by avoiding unnecessary (over)heating. As a result, plant-model mismatch is typically inevitable in practice. If inadequately accounted for, however, such modeling errors may result in insufficient heating of the tumor or in the overheating of healthy tissue, which would significantly deteriorate treatment quality and safety.

Given this situation, we propose an MPC setup novel for hyperthermia inspired by offset-free MPC [25], [26]. Fig. 4 shows the resulting feedback scheme, where the observer provides the temperature and disturbance estimates from the noninvasively sampled MR thermometry data, such that the MPC scheme can compute the optimal power distribution to be generated by the extracorporeal HIFU actuator.

### III. THERMAL MODEL

In this section, a model of the body’s thermal response will be discussed and spatially discretized, yielding the state-space model from which the observer and controller models can be derived.

#### A. Bioheat Model

The tissue’s thermal behavior is modeled using Pennes’ bioheat equation [30] given by

\[
\rho c \frac{\partial T(r, t)}{\partial t} = \nabla \left( \kappa \nabla T(r, t) \right) + Q_a(r, t) - w_b(r, t, T) c_b(T(r, t) - T_b)
\]

(1)
where $T : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is the temperature profile. In particular, $T(r, t)$ denotes the temperature at time $t \in \mathbb{R}_{\geq 0}$ and location $r \in \mathbb{R}^3$ with $\Omega$ being the patient domain in the focal plane. Furthermore, $\rho : \Omega \to \mathbb{R}_{\geq 0}$ denotes the tissue’s volumetric mass density, $c : \Omega \to \mathbb{R}_{\geq 0}$ denotes the specific heat capacity, $\kappa : \Omega \to \mathbb{R}_{\geq 0}$ denotes the thermal conductivity, and $w_b : \Omega \times \mathbb{R}_{\geq 0} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ denotes the blood perfusion coefficient. The blood’s specific heat capacity and temperature are given by $c_{ph} \in \mathbb{R}_{\geq 0}$ and $T_b \in \mathbb{R}$, respectively. Note that these tissue properties are typically spatially varying in reality due to tissue heterogeneity or blood vessels, for example. In addition, the blood flow coefficient $w_b$ can also depend nonlinearly on time and temperature due to the body’s thermoregulatory response, which is patient- and tumor-specific [22]–[24], [31]. Although low-power pre-treatment test sonications may serve as a practical solution to obtain initial estimates of the tissue properties (see [21]), fully accurate models are nearly impossible to obtain due to the unavailability of extensive model identification for the reasons previously mentioned. We therefore propose to assume spatially homogeneous tissue properties that remain constant in time, reducing (1) to

$$\frac{\rho c}{\partial t} \nabla^2 T(r, t) + Q_a(r, t) - w_b c_h (T(r, t) - T_b) = 0$$

(2)

and design a feedback controller that is able to identify and compensate for the resulting model mismatch. Note, however, that our MPC setup can be directly applied to a fully inhomogeneous tissue model as well, in case such a model would be available.

The power deposition density $Q_a : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ depends on the acoustic deposition intensity $F : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and scales linearly with the sonication power $P : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ as described by

$$Q_a(r, t) = F(r, t) P(t), \quad r \in \Omega, \quad t \in \mathbb{R}_{\geq 0}.$$  

(3)

In reality, for a given focus location $r_f(t)$ at some time $t$, the intensity $F(r, t)$ also depends on spatially varying factors, such as the acoustic properties of the (intermediate) tissue. In our approach, however, $F$ is modeled by a radially symmetric 2-D Gaussian distribution centered around the focus location $r_f$ with standard deviation $\sigma_f = 2.4$ mm

$$F(r, t) = \frac{a}{2 \pi \sigma_f^2} \exp \left(-\frac{\|r - r_f(t)\|^2}{2 \sigma_f^2}\right)$$

(4)

with $r \in \Omega, t \in \mathbb{R}_{\geq 0}$ and acoustic energy absorption coefficient $a \in \mathbb{R}_{\geq 0}$.

B. Spatially Discretized State-Space Model

Recall that we only consider the 2-D focal plane and assume zero out-of-plane interaction, justified by the temperature distribution being fairly homogeneous in the axial direction, as discussed in Section II-B. Spatial discretization of (2)–(4) is done using the central difference scheme, which in [21] has been concluded to best combine model simplicity with predictive power for the considered system, on a $44 \times 44$ grid with a voxel size of $1.85 \times 1.85$ mm$^2$. As boundary condition, we prescribe the outward heat flux corresponding to a fixed temperature $T_b$ just outside our grid’s edges and shift the origin of (2) to the blood temperature $T_b$. The resulting continuous-time space-state dynamics are given by

$$\dot{x}(t) = A_c x(t) + B_c u(t)$$

(5a)

where the states $x(t) \in \mathbb{R}^n_x$ with $n_x = 44^2 = 1936$ represent the voxel’s temperature elevations with respect to $T_b$ at time $t \in \mathbb{R}_{\geq 0}$. Due to the central difference scheme, $A_c$ is a sparse matrix containing at most five nonzero elements per row/column. For the input, we choose to allow sonication at the voxel centers within the ROI, which will be referred to as the sonication points. The input $u(t) \in \mathbb{R}^n_u$, $n_u = 60$, represents the applied acoustic power at the sonication points, which may contain at most one nonzero element at any time $t$, corresponding to the sonication point that coincides with the focus location $r_f(t)$ using electronic beam steering. Each individual column of $B_c$, therefore, captures the rate of increase of the voxel temperatures when applying unit sonication power at the corresponding individual sonication point.

MR thermometry does not yield continuous measurement, but instead induces a sample time of $T_s = 3.7$ s when employing the acquisition protocol used in this article. The spatial discretization in (5a) is specifically chosen such that the voxel centers coincide with the points measured by MR thermometry. Hence, the measurements can be modeled by

$$y_k = x(t_k) + v_k$$

(5b)

with $k \in \mathbb{N}$ connecting to real time $t_k = k T_s$ and $y_k = y(t_k) \in \mathbb{R}^n_y$ representing the measured voxel temperatures corrupted by $v_k \in \mathbb{R}^n_y$, where in the monitored area of interest, the latter consists mainly of sensor noise, which has been experimentally determined to typically be zero-mean and Gaussian distributed.

IV. CONTROLLER DESIGN

In this section, we first introduce the disturbance model and discuss the sampled-data setup, based on which we then derive the discrete-time state-space models used by the controller and observer. The cost function is designed using the control
To facilitate the derivation of the required discrete-time model induced by model mismatch or slowly varying disturbances, we formulate the constrained optimization problem for MPC, enabling it to compensate for the steady-state offset resulting from model error.

A. Disturbance Model

To incorporate integral action in the MPC setup, a disturbance model together with a disturbance estimator can be used [25], [26]. The disturbance estimate is provided to the controller, enabling it to compensate for the steady-state offset induced by model mismatch or slowly varying disturbances. To facilitate the derivation of the required discrete-time models, let us augment the continuous-time dynamics (5a) to

\[
\dot{x}(t) = A_c x(t) + B_c u(t) + d(t) \quad (6a)
\]

\[
\dot{d}(t) = 0 \quad (6b)
\]

such that it includes the additive disturbance \(d(t) \in \mathbb{R}^n_d\), which is assumed to be constant in time. This particular choice of disturbance model allows for the construction of a stable discrete-time estimator, as will be shown in Section IV-D, which in turn enables the compensation of the steady-state offset resulting from model error.

B. Sampled-Data Setup

We design a discrete-time controller for a continuous-time system. The sampled-data workflow is schematically visualized in Fig. 5. Here, the sampled output \(y_k\) is given by (5b).

The discrete-time observer variables \(\hat{x}_k, \hat{d}_k \in \mathbb{R}^n\) represent the estimates of the state \(x(t_k)\) and disturbance \(d(t_k)\), which become available at (approximately) the same time as the measurement \(y_k\) due to negligible observer computation time. As indicated in the figure, for the considered system, there is a delay of \(T_0 \approx 1.1\) s between obtaining a temperature map and applying the corresponding new heating plan, which is caused mainly by the numerical optimization in MPC and a significant communication and data processing overhead between the MPC and the dedicated MR-HIFU software.

As discrete-time input, we define \(u_k \in \mathbb{R}^n_u\), which represents the applied sonic power averaged over one sampling interval at each of the sonication points. This is achieved by using electronic beam steering to rapidly switch the focal spot \(r(t)\) between all sonication points at which heating is desired in such a manner that, on average over one sampling interval, the power as requested by \(u_k\) is injected. Recall that the input \(u(t)\) in (6) can contain at most only one nonzero element. However, due to the high steering frequency with respect to the slow sampling and system dynamics, the input in (6) can be well approximated using a zero-order hold (ZOH) given by \(u(t) = u_k\) for \(t_k + T_0 \leq t < t_k + 1 + T_0\) (note the shift of the interval by \(T_0\) due to the computation/communication delay), where \(u_k\) may contain more than one nonzero element.

Finally, in temporally discretizing the continuous-time dynamics (6) to obtain the controller state-space models, we account for the delay \(T_0\) without increasing the computational complexity of the numerical optimization in MPC. That is, we first compute the prediction sequence’s initial conditions \(x_{0|k}, d_{0|k} \in \mathbb{R}^n\), representing the predicted state and disturbance at time \(t_k + T_0\), and then synchronize the prediction instants \(t_{i|k}, i, k \in \mathbb{N}\), in the MPC horizon with the (future) time instants \(t_{k+i} = t_k + T_0 + iT_s\) at which the control inputs will be updated, as shown in Fig. 5. As a result, a state update using the prediction model derived in Section IV-C requires only a single input value, as opposed to two in case the predictions and controls were not synchronized.

C. Prediction Model

As discussed in Section IV-B, the discrete-time state-space prediction model is then given by

\[
x_{i+1|k} = A x_{i|k} + B u_{i|k} + B_d d_{i|k} \quad (7a)
\]

\[
d_{i+1|k} = d_{i|k} \quad (7b)
\]

\[
y_{i|k} = x_{i|k} \quad (7c)
\]

where \(x_{i|k}, d_{i|k}, y_{i|k} \in \mathbb{R}^n\) and \(u_{i|k} \in \mathbb{R}^n_u\) denote the predicted states, disturbance, outputs, and control inputs, respectively, at \(i \in \mathbb{N}\) time steps ahead of the prediction sequence’s starting time \(k \in \mathbb{N}\). Note that we assume the disturbance to be constant over the prediction horizon and that we do not incorporate measurement noise in the prediction model since (approximately) \(E(t_{i|k}) = 0\). Here, we use the forward Euler method to temporally discretize (6), i.e., \(A = I_n + T_s A_c, B = T_s B_c\), and \(B_d = T_s B_d\), as this fully preserves the sparseness present in the continuous-time state-space matrices, which is crucial to reduce the computation time of the numerical optimization step in the MPC, while also yielding sufficient predictive accuracy. In particular, for the considered system, this yields a (numerically) stable approximation for \(0 < T_s < -2/\lambda(A_c) = 8.35\) s (where we use the fact that all eigenvalues of \(A_c\) are real and negative and where \(\lambda(A_c) < 0\) denotes the most negative eigenvalue of \(A_c\)), and when simulating entire treatment given typical input trajectories for \(T_s = 3.7\) s, it induces an approximation error in the order of 0.01 °C in steady state and at most around 0.1 °C during the transient. Moreover, note that since we have synchronized the time instants of input application and state prediction, (7a) indeed contains only a single control input term, which is beneficial for computation time.

In addition to (7), we derive the model that computes the prediction’s initial conditions as shown in Fig. 5, which is given by

\[
x_{0|k} = A_0 x_k + B_0 u_{k-1} + B_{d,0} d_k \quad (8a)
\]

\[
d_{0|k} = \hat{d}_k. \quad (8b)
\]

Since these need only be computed once before numerical optimization, matrix sparseness is of negligible importance.
To avoid numerical approximation errors, we obtain the matrices by exact discretization using \( A_0 = e^{T_x A} \), \( B_0 = \int_{t_0}^{T_x} e^{A \tau} d \tau B_c \), and \( B_{d,0} = \int_{t_0}^{T_x} e^{A \tau} d \tau \), of which the latter two expressions exploit ZOH on the corresponding (disturbance) inputs.

### D. State and Disturbance Estimation

We use an observer to obtain temperature estimates in which the effects of measurement noise are reduced with respect to the MR thermometry readings and to identify the disturbance that captures the offset induced by plant-model mismatch. The state and disturbance estimator is given by

\[
\dot{\hat{x}}_k = \hat{x}_{k-1} + L_x (y_k - \hat{y}_k) \quad (9a)
\]

\[
\dot{d}_k = \hat{d}_{k-1} + L_d (y_k - \hat{y}_k) \quad (9b)
\]

where

\[
\hat{y}_k = \hat{x}_{k-1} + \hat{B}_1 u_{k-1} + \hat{B}_2 u_{k-2} + \hat{B}_d \hat{d}_{k-1} \quad (9c)
\]

denote the model-based output and state estimates at time \( k \) before applying the measurement-based correction to obtain \( \hat{x}_k \). For the same reasons as for (8), the observer model’s discrete-time matrices are obtained by exact discretization of (6) as given by \( \hat{A} = e^{T_x A} \), \( \hat{B}_1 = \int_{t_0}^{T_x} e^{A \tau} d \tau B_c \), \( \hat{B}_2 = \int_{t_0}^{T_x} e^{A \tau} d \tau B_c \), and \( \hat{B}_d = \int_{t_0}^{T_x} e^{A \tau} d \tau \), where the expressions for \( \hat{B}_1 \) and \( \hat{B}_2 \) again follow from the fact that we use ZOH on the inputs (see also [32]).

We use [25, Proposition 1] to verify the observability of the augmented discrete-time state-space model on which the observer is based (i.e., (9) with \( L_x = L_d = 0 \)), which is derived using the disturbance model (6). That is, we first verify that the nominal system (excluding the disturbance) is observable due to the full (noisy) state measurement (5b). Second, we evaluate the rank condition in [25, Proposition 1], for which in our case, it holds that

\[
\text{rank} \begin{bmatrix} \hat{A} & \hat{B}_d \\ I & 0 \end{bmatrix} = 2n
\]

(10)

from which we conclude the augmented model to be observable. Consequently, an asymptotically stable estimator is known to exist for the proposed disturbance model. This property also guarantees that for a stable closed-loop system, a feasible steady-state setpoint, and constant disturbances, there will be zero offset between the realized and desired temperature distributions in case no constraints are active at steady state (see [33, Th. 1]).

By observability of the augmented model, \( L_x \) and \( L_d \) can in principle be used for arbitrary pole placement of the estimation error system resulting from (9). However, determining the observer gains using exact pole placement methods or computing them as the Kalman gains typically results in a prohibitively large computational burden due to the large augmented state dimension \( 2n_x = 3872 \). Therefore, we set \( L_x \) to be a diagonal matrix instead, which, due to \( y_k \) in (5b) being the noisy full state measurement, is a practical solution with good performance for reducing the propagation of noise into the state estimate \( \hat{x}_k \). Similarly, we set \( L_d \) diagonal to achieve low-pass filtering of the measured plant-model mismatch, which is possible since \( \hat{d}_k \) is chosen to contain an estimate of the constant and low-frequency mismatch for each state element individually. Using a polyacrylamide tissue-mimicking phantom as in [34], but with the ink replaced by water, the matrices have been tuned experimentally, resulting in

\[
L_x = 0.25 I_{n_x}, \quad L_d = 0.01 I_{n_x}
\]

(11)

which yield desirable estimator behavior with stable closed-loop poles.

### E. Cost Function

The temperature objectives have been discussed in Section II-C and are schematically depicted in Fig. 6 in cross-section perspective. The concentric circular regions \( \mathcal{R} \) and \( \mathcal{S} \) inside the patient domain \( \Omega \), for which it holds that \( \mathcal{R} \subset \mathcal{S} \subset \Omega \), denote the ROI and the region outside of which healthy tissue must be safeguarded from overheating, respectively. \( \mathcal{T} : \mathcal{R} \to \mathbb{R} \) and \( \mathcal{T} : \Omega \to \mathbb{R} \) represent the location-dependent (desired) lower and upper temperature bounds, respectively, defining the desired temperature range (green). Underheating and overheating occur for temperatures within the (light) blue and (light) red areas in the figure, respectively. \( \mathcal{L} \) is nonzero and defined only within \( \mathcal{R} \), where sufficient heating is desired. \( \mathcal{T} \) features an elevated plateau on \( \mathcal{S} \), preventing reversal of the temperature-dependent mechanisms due to overheating, and has a lower value on \( \Omega \setminus \mathcal{S} \) to protect healthy tissue. Within the aforementioned temperature range, optimal treatment efficacy is achieved when a flat temperature distribution is realized in the ROI, which is described by the reference temperature \( T_r : \mathcal{R} \to \mathbb{R} \).

To formulate these objectives mathematically, let us denote the \( n_z < n_s \) elements of \( x_k \) corresponding to a point within \( \mathcal{S} \) by \( z_k = H x_k \in \mathbb{R}^{n_z} \), where \( H \in \{0,1\}^{n_s \times n_z} \) is a matrix with one in each row (and at most one per column). Furthermore, let \( z_r \in \mathbb{R}^{n_z} \), \( \epsilon \in \mathbb{R}^{n_z} \), and \( \mathbf{x} \in \mathbb{R}^{n_s} \) denote the voxel-wise temperature elevation reference, lower bound, and upper bound corresponding to the values of \( T_r \), \( \mathcal{T} \), and \( \mathcal{L} \) on the
discrete voxel locations, respectively. This allows for the ROI voxel temperature deviations with respect to the reference to be given by \( z_k - z_r \). The maximum violations of the temperature bounds are defined by \( \tau_k = \tau(x_k) = \| \max(x_k - T, 0) \|_\infty \in \mathbb{R}_{\geq 0} \) and \( \xi_k = \xi(z_k) = \| \max(z - z_k, 0) \|_\infty \in \mathbb{R}_{\geq 0} \), where \( 0_n \) denotes a zero vector of length \( n \) and the maximum operator is used elementwise, and are collected in \( \epsilon_k = [\xi_k \, \tau_k]^T \in \mathbb{R}_{\geq 0}^n \), \( n_x = 2 \).

The cost function is given by

\[
V_N(z_k, \epsilon_k) = \sum_{i=0}^{N} \ell(z_{i|k}, \epsilon_{i|k})
\]

where \( z_k = (z_{0|k}, \ldots, z_{N|k}) \) is the sequence of predicted performance variables \( z_{i|k} = Hx_{i|k} \) at time instant \( k \in \mathbb{N} \), and similarly, \( \epsilon_k = (\epsilon_{0|k}, \ldots, \epsilon_{N|k}) \) is the prediction sequence of the slack variable vector \( \epsilon_{i|k} = [\xi_{i|k} \, \tau_{i|k}]^T \). The stage cost is given by

\[
\ell(z_{i|k}, \epsilon_{i|k}) = (z_{i|k} - z_r)^T Q(z_{i|k} - z_r) + f_r^T \epsilon_{i|k}
\]

where \( Q \) is a positive-definite weighting matrix and \( f_r \in \mathbb{R}_{\geq 0}^2 \). The critical importance of achieving temperatures within the desired range \( T \leq T \leq \bar{T} \) is expressed by using a linear penalty on \( \epsilon \) such that its contribution to the cost does not vanish quadratically when approaching zero. In addition, the weighting in \( f_r \) is typically chosen relatively high compared to the weighting in \( Q \) such that any (significant) violation of the temperature range will be the dominant component of the cost. When the soft constraints are (nearly) satisfied, i.e., when \( \epsilon_k \approx 0 \), the contribution of the quadratic term will be dominant, enforcing temperature tracking of the voxels within the ROI to the optimal treatment temperature \( T_r \).

Although it is a natural choice to use slack variables and soft constraints to incorporate the desire for temperatures within the range \( L \leq T \leq U \) in the cost function, the manner in which these are implemented deserves considerable attention. We choose to penalize the ROI voxels’ magnitudewise largest violations using the \( \infty \)-norm, as opposed to introducing a slack variable for each voxel individually and using a penalty on the 1-norm of all slack variables. Compared with the latter, the former has two main advantages. First, it requires only two slack variables per predicted time step (one for the upper bound and one for the lower bound) compared to \( 2n_x \) slack variables per time step (\( n_x \) per bound), thereby introducing less complexity in the MPC optimization problem. Second, it yields superior temperature homogeneity, which is desirable from a clinical point of view. Although this effect is important during the entire treatment, it is especially visible during the initial heat-up phase, and hence, we will use this interval to illustrate the behavioral advantage of the \( \infty \)-norm approach.

To this end, compare the temperature cross sections in Fig. 7(a) obtained through simulation with \( n_k = 0 \) using the \( \infty \)-norm to those in Fig. 7(b) corresponding to using the 1-norm. In the former, relatively more heating is applied toward the edge of the ROI, where the temperature is naturally lowest due to outward heat diffusion, leading to a more uniform temperature distribution within the ROI (gray area), whereas in Fig. 7(b), a much larger difference exists between the maximum ROI voxel temperatures (■) and the minima at the ROI edge (●). In addition, in Fig. 7(c), we show the temperature profiles obtained when omitting the slack variable penalty altogether, i.e., when using (13) with \( f_r = 0 \). Besides lacking the

![Fig. 7. Evolution of the voxel temperatures (×, ●, and ■) of the cross section at \( r_x = 0 \) m during heat-up when using (a) an \( \infty \)-norm penalty, (b) a 1-norm penalty, and (c) no penalty on the slack variables, at the time instants indicated in (d) (dotted), alongside the ROI (gray area), and the temperature range's upper and lower bounds (dashed line) and setpoint temperature (dashed–dotted line). In (d), the difference between the maximum (■) and minimum (●) temperature in the ROI is shown for all cases.](image-url)
ability to explicitly counteract temperatures outside the desired range, using \( f_\epsilon = 0 \) results in similar heat-up behavior as with the 1-norm approach. Finally, in Fig. 7(d), it is clear to see that the resulting temperature range inside the ROI is significantly narrower during heat-up when including the cost on the uniform lower bound following from the \( \infty \)-norm penalty (and even also slightly better directly thereafter), from which we conclude that using the \( \infty \)-norm penalty is indeed the most favorable of the considered options.

\[ F. \text{Optimization Problem} \]

The control objective can now be formulated as the constrained optimization problem

\[
\min_{u_k} V_N(z_k, \epsilon_k) \quad (14a)
\]

where \( u_k = (u_{0,i}, \ldots, u_{N-1,i}) \), subject to

\[
x_{i+1,j} = Ax_{i,j} + Bu_{i,j} + Bd_{i,j}, \quad \forall i \in \mathbb{N}_{[0,N-1]} \quad (14b)
\]

\[
d_{i+1,j} = d_{i,j}, \quad \forall i \in \mathbb{N}_{[0,N-1]} \quad (14c)
\]

\[
x_{0,i} = A_0 x_{i,j} + B_0 u_{i,j-1} + B_{d,0} d_k \quad (14d)
\]

\[
d_{0,i} = d_k \quad (14e)
\]

\[
z_{i,j} \geq z - 1_n z_{i,j}, \quad \forall i \in \mathbb{N}_{[0,N]} \quad (14f)
\]

\[
x_{i,j} \leq x + 1_n x_{i,j}, \quad \forall i \in \mathbb{N}_{[0,N]} \quad (14g)
\]

\[
0_n \leq \epsilon_{i,j}, \quad \forall i \in \mathbb{N}_{[0,N]} \quad (14h)
\]

\[
0_n \leq u_{i,j} \leq 1_n \bar{u}, \quad \forall i \in \mathbb{N}_{[0,N-1]} \quad (14i)
\]

\[
1_n^T u_{i,j} = \bar{w}_\Sigma, \quad \forall i \in \mathbb{N}_{[0,N-1]} \quad (14j)
\]

where \( \mathbb{N}_{[a,b]} = \{a, a+1, \ldots, b\} \), \( a, b \in \mathbb{N} \) and \( a \leq b \), denotes the set of natural numbers from \( a \) through \( b \), and \( 1_n \) represents an all-ones vector of length \( n \). Equality constraints (14b) and (14c) ensure satisfaction of the dynamics according to (7), with the prediction sequence’s initial conditions given by (14d) and (14e) corresponding to (8). Inequalities (14f) and (14g) capture the temperature bounds as soft constraints using the slack variables bounded from below in (14h). Finally, (14i) and (14j) describe the actuator constraints, where \( \bar{u} = 15 \text{ W} \) and \( \bar{w}_\Sigma = 60 \text{ W} \) are the maximum allowable power applied to a single sonication point and to the entire treatment region, respectively.

Furthermore, a horizon of \( N = 5 \) is used, and for the weights, we choose

\[
Q = \frac{1}{\mu_c} I, \quad f_\epsilon = \begin{bmatrix} 10 \\ 100 \end{bmatrix} \quad (15)
\]

which are normalized with respect to the number of corresponding variables (note that \( \xi \) and \( \tau \) are scalar) for more intuitive balancing of the objectives’ relative contribution to the cost function.

We choose to incorporate the desired temperature range objective as soft constraints to prevent the optimization problem from becoming infeasible, as proper controller operation is crucial for patient safety and treatment quality. Note that this is of particular importance for the considered application, as the minimum temperature bound will be directly violated during heat-up. In addition, violations may occur as a result of unexpected heat sinks or sources, or due to measurement artifacts that may occur in MR-based thermometry due to patient motion or magnetic field drift.

\[ \text{Remark 1:} \] The stage cost (13) is improved with respect to [27] by excluding the cost on the predicted input power \( u_{i,k} \). That is, removing this penalty slightly increases control effort, but consequently also removes the associated downward shift of the tumor temperature away from the reference \( z_* \). This is desirable for hyperthermia treatments, where achieving optimal tumor temperatures, and thereby optimal treatment quality, is of the utmost importance, in comparison to which the desire to reduce the control energy is negligible. In addition, note that the control effort is, in fact, upper bounded by the hard input constraints (14i) and (14j). Similarly, we chose not to include an input rate penalty in (13), since reducing changes in the input signal is also of little to no importance compared to the temperature objectives. Moreover, the input oscillations resulting from the propagation of noise into the state and disturbance estimates are reduced to a negligible level by adequate filtering in the observer, as will be shown in Fig. 11 in the upcoming discussion of the in vivo experiment.

\[ \text{Remark 2:} \] The cost function (12) describes an economic cost, as the setpoint \( z_* \) does not correspond to a reachable steady-state solution of the considered system (7). This is due to the combination of the ROI radius being equal to the lateral range using electronic beam steering, meaning that all sonication points are inside the ROI, the finitely narrow power deposition in (4), and the fact that the nonnegativity constraints on the inputs (14i) are active at steady state. As a result, although having satisfied (10), we cannot guarantee that \( z_k \to z_* \) for \( k \to \infty \). Instead, using the fact that the estimator is asymptotically stable and the disturbances are (nearly) constant, we only claim to be able to remove the part of the steady-state offset resulting from the plant-model mismatch, recovering the performance as optimally achievable without modeling error. An alternative approach would be to formulate the objective as a tracking problem with respect to a feasible steady-state setpoint. To this end, a target selector as discussed in [35] and [36] could be used to determine the economically optimal steady state and corresponding input \( (x^*, u^*) \) as a function of the current disturbance estimate by solving at each time \( k \)

\[
(x^*, u^*) = \arg \min_{x, u} f(z, \epsilon) \quad (16)
\]

with \( z_s = H x_s \) and \( \epsilon_s = [\epsilon(z_s), \tau(x_s)]^T \), subject to

\[
[I - A - B] x^* \quad (17)
\]

and inequality constraints similar to (14f)–(14j). However, such a setup has been observed to deteriorate performance compared with the current design during the transient of \( d_k \) and in the presence of measurement noise, while yielding at best similar performance after convergence of the disturbance estimate (see also [37]). In addition, this increases complexity (and thereby computation time), which is particularly unwanted in the clinic, where any additional complexity is regarded as a potential source of errors. We have, therefore,
chosen to omit such a target selector and instead optimize the economic cost in (12) directly.

V. IN VIVO EXPERIMENTS

In this section, the experimental results obtained on an in vivo porcine thigh muscle are presented to demonstrate the effectiveness of the designed feedback control setup for removing the part of the treatment temperature steady-state offset caused by plant-model mismatch.

A. Implementation

The MR-HIFU system, consisting of the combination of an MRI scanner and a dedicated HIFU therapy platform, has been discussed in Sections II-A and II-B. The MPC and estimator algorithms were implemented in MATLAB 2017b and executed in real time using Python 2.7 via the MATLAB engine API, running on an HP Z800 Workstation with Intel Xeon X5650 CPU at 2.67 GHz and 12 GB RAM. The communication with the MR-HIFU system was achieved using the matMRI and matHIFU toolboxes [38]. For fast numerical optimization, we used Gurobi 7.0.2 to solve a sparse formulation of the MPC problem (14), which after comparative testing with a dense formulation was concluded to result in significantly smaller computation times.

In particular, the solver times recorded during the experiments were fairly consistent, averaging at approximately 0.26 s when the disturbance estimator was disabled, i.e., when running MPC without offset-free capabilities, and around 0.33 s for the offset-free MPC (14). Since the computation time of the observer (9) is negligible (≤5 ms in both cases), this difference illustrates the additional computational time associated with using an offset-free approach. Recalling that the sample time is $T_s = 3.7$ s, we find that the resulting computation times for both MPC strategies are still well within the bounds for real-time implementation. Given the significant performance improvement demonstrated by the experimental results shown next, this strongly motivates the offset-free MPC approach.

B. Tissue Parameters

Using our proposed approach, we assume the treatment area to have homogeneous tissue properties and rely on the disturbance estimator to identify the effects of (inhomogeneous) parameter error. Baseline values for the relevant muscle tissue parameters have been obtained from the IT’IS Foundation database [39], which have then been slightly adapted according to the identification procedure in [21] based on low-power test sonications. For the thermal model equations (2)–(4), from which the continuous-time state-space model (6) and in turn the discrete-time controller and observer models (7)–(9) are derived, however, we have intentionally introduced additional parameter mismatch that results in a negative temperature offset in closed loop. In particular, corresponding to the observations in [21], this is achieved by underestimating the thermal conductivity $\kappa$ in (2) and overestimating the acoustic energy absorption coefficient $\alpha$ in (4), resulting in the parameters listed in Table I, and a plant-model mismatch as visualized in the Appendix by an open-loop model validation.

This is done for two reasons. First, for the purpose of this article, it allows for a better illustration of the controller’s ability to achieve the desired treatment temperature despite significant model error. Second, although any plant-model mismatch is obviously undesirable in general, in case it is inevitable due to large model uncertainty, for patient safety it is preferred to have a model bias toward underheating instead of toward overheating. That is, in case of the former, no harm is unintentionally inflicted to the tissue, and, after proper identification of the mismatch, increasing the sonication power will directly contribute to achieving the desired temperature. Contrarily, overheating the tissue may induce hyperthermic reversal effects detrimental to treatment success or directly cause unwanted tissue damage and, after identification, cannot be actively counteracted using HIFU (as expressed by the nonnegativity constraints on the inputs $u_k \geq 0$).

C. Initial Disturbance Estimator Inactivity

In line with the desire to demonstrate the benefits of using the proposed offset-free implementation, the disturbance estimator is disabled during the initial heat-up phase by setting $\delta_k = 0$ and $L_d = 0$ for all $k \in \mathbb{N}$ corresponding to $t_k < 180$ s. This is, however, also desirable from a clinical point of view, as hereby we avoid a potential and unnecessarily large build-up of $\delta_k$ that would otherwise lead to increased temperature overshoot or undershoot, see Remark 3.

Remark 3: Initially disabling the disturbance estimator is especially useful in case of a mismatch in the input matrix $B_c$ in (6) (from which the observer and controller input matrices are derived), e.g., resulting from an incorrect value for the acoustic absorption coefficient $\alpha$, combined with rapid and substantial changes in the input $u_k$ such as typically occurring in the transient behavior during heat-up (especially when transitioning from heat-up toward steady state). This effect is due to two reasons. First, a model error in $B_c$ is multiplicative in nature, whereas we use an additive disturbance in the augmented model (6) and corresponding observer (9). Second, our disturbance estimator (9) is designed to estimate steady-state (or slowly varying) errors, while the input $u_k$ (and thus the mismatch in the terms $B_c u_k$, $i = 1, 2$) can vary rapidly. Although this could possibly be tackled by using different disturbance models or a more sophisticated adaptive controller including online parameter estimation, in the context of MR-HIFU hyperthermia the resulting performance improvement is believed to be minor for the following reasons. First of all, due to its general form, the design proposed in this work is able

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>1090</td>
<td>kg/m³</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>3421</td>
<td>J/(kg·K)</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>0.3773</td>
<td>W/(m·K)</td>
</tr>
<tr>
<td>$\omega_k$</td>
<td>3.5289</td>
<td>kg/(s·m³)</td>
</tr>
<tr>
<td>$c_b$</td>
<td>3617</td>
<td>J/(kg·K)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.06141</td>
<td>-</td>
</tr>
</tbody>
</table>

TABLE I

Tissue Parameters

Authorized licensed use limited to: Eindhoven University of Technology. Downloaded on October 17,2021 at 17:23:58 UTC from IEEE Xplore. Restrictions apply.
to fully capture the effects of any constant or slowly varying model mismatch and disturbance in steady state, rather than only specific types of model errors. Furthermore, our control setup already achieves heat-up times of less than 1 min, which is short with respect to the envisioned 90-min treatment time, after which the disturbance estimator can be enabled. In fact, the maximum power limit is already typically reached during heat-up, indicating that there is relatively not much room for shortening the heat-up phase.

D. Results

To test the performance of the developed offset-free controller and observer setup under realistic clinical conditions, experiments have been conducted on an in vivo porcine thigh muscle. The focal plane MR image is shown in Fig. 8, indicating the treatment area (green) and ROI (red). We set $r_x = r_y = 0$ m at the center of the ROI, with positive $r_x$- and $r_y$-axis directions toward the ventral (belly) and cranial (head) side, respectively. In Fig. 9, the results of a 30-min MR-HIFU hyperthermia treatment are presented. Note that the baseline body/blood temperature of the treated pig was $T_b \approx 38^{\circ}C$. The mean (solid line) and minimum/maximum (dashed line) voxel temperatures are shown corresponding to, with a temperature shift of $T_b$, both the state estimate in the ROI $\hat{z}_k = H\hat{x}_k$ (black thick line) and the measurement in the ROI $z_{y,k} = Hy_k$ (red thin line), together with the ROI’s reference temperature $T_r$ and lower and upper bounds of the desired temperature range $T_l$ and $T_u$, respectively. The disturbance estimator is disabled for $t_k < 180$ s indicated by the gray area. In Fig. 9(a), the temperature indicators are given for the full length of the treatment, showing that after heat-up and disturbance estimator convergence, the entire ROI is heated to within the desired temperature range, with the mean temperature settling at $T_r$. Moreover, it illustrates the benefit of using a model-based observer to obtain temperature estimates that contain significantly reduced noise effects compared to the corrupted measurements.

Closer inspection of the interval where $0 \leq t_k \leq 400$ s in Fig. 9(b) reveals the effect of the existing plant-model mismatch on the treatment temperature in closed loop. Due to the overestimation of the achievable temperature increase, it appears from the observer as if the minimum ROI temperature has exceeded the desired $41^{\circ}C$ at $t_k \approx 26$ s, whereas the measured (noisy) minimum ROI temperature is around $39.7^{\circ}C$ at this time. It can also be seen that with the disturbance estimator disabled, the modeling error results in a discrepancy between the measured and estimated temperatures that settles around $0.3^{\circ}C$ on average at $t = 180$ s. Consequently, a negative mean temperature offset of almost $0.5^{\circ}C$ can be observed, assuming that the effect of $v_k$ on the mean of $z_{y,k}$ is sufficiently suppressed by averaging and, therefore, $z_{y,k}$ is representative of the actual mean ROI temperature. After activating the disturbance estimator, i.e., setting $L_d = 0.01I_x$, as defined in (11) for $t_k \geq 180$ s, the state and disturbance

![Fig. 8. Sagittal focal plane MR image of the porcine right thigh area, ranging from halfway the abdomen (cranial) to the ankle (caudal), indicating the location of the ROI (red circle).](image)

![Fig. 9. Mean (solid line) and minimum/maximum (black dashed line) voxel temperatures of the ROI corresponding to the measurements $z_{y,k} = Hy_k$ (red thin line) and state estimates $\hat{z}_k = H\hat{x}_k$ (black thick line) with the disturbance estimator disabled for $t < 180$ s (gray area), shown for (a) the entire experiment and (b) zoomed in on the first 400 s of the experiment [dotted rectangle in (a)].](image)
estimates converge within approximately 1 min, allowing for the feedback setup to achieve the desired temperature distribution, as would typically be the result in case of no model error.

Temperature cross sections at $r_x = 0$ m, i.e., through the center of the ROI in the $r_y$-direction, are shown in Fig. 10 corresponding to the averages of the state estimate $\hat{x}$ (black ◦) and measurement $\bar{y}$ (red ×) with corresponding $2\sigma$ confidence intervals (red/gray areas) when the disturbance estimator is (a) inactive over the interval $154 \leq t_k \leq 180$ s, and (b) active during $374 \leq t_k \leq 400$ s, alongside the $T_{e}$ and $T_r$ (dashed black) and $T_{c}$ (dash-dotted black).

Fig. 10. The 7-sample averaged temperature cross sections at $r_x = 0$ m corresponding to $\bar{y}$ (black ◦) and the measurement $\bar{y}$ (red ×) with corresponding $2\sigma$ confidence intervals (red/gray areas) when the disturbance estimator is (a) inactive over the interval $154 \leq t_k \leq 180$ s, and (b) active during $374 \leq t_k \leq 400$ s, alongside the $T_{e}$ and $T_r$ (dashed black) and $T_{c}$ (dash-dotted black).

The corresponding total input power per time step is shown in Fig. 11. The benefit of using an observer, reducing the propagation of noise effects into the control input, is expressed in the small magnitude of the power fluctuations occurring during the treatment. The figure also clearly visualizes the large acoustic power required for initial heat-up limited only by the input constraint (14), and the increased power demand after activating the disturbance estimation that eventually brings the tissue in the ROI to its desired temperature. Interestingly, the requested sonication power in this experiment also exhibits a steady increase for $t_k \geq 800$ s during which the tissue temperature does not significantly change. This is most likely due to the tissue’s thermoregulatory response, increasing perfusion when having been exposed to elevated temperatures for an extended period of time, as will be discussed in more detail in Section V-E. This clearly demonstrates the controller’s ability to maintain optimal performance also in the presence of slowly varying model mismatch.

Possibly even more remarkable, however, is that the power increase can be observed to be seemingly linear. This is of special importance for medical science since measuring and modeling the transient behavior of the body’s thermoregulatory response to thermal stress has proved to be difficult [24]. By providing the means for accurately controlling the hyperthermic conditions and identifying tissue changes, our controller may serve as an enabler for research into better physiological tissue models.

E. Spatially and Temporally Varying Disturbance Estimation

For the considered experiment, the ROI contained significantly vascularized tissue, as can be seen from the focal plane MR image shown in Fig. 12, which corresponds to the area marked in Fig. 8 (green box). The outer edges of the ROI $R$ and the area including the surrounding tissue $S$ are marked by the red solid circle and the black dashed circle, respectively. Tissue areas with stronger perfusion, typically
Fig. 12. MR image of the treatment area (green box in Fig. 8), highlighting the intensity of the vascularization (grayscale) in and around the ROI $R$ (red circle) and $S$ (dashed black circle).

Fig. 13. Estimated voxel-wise temperature mismatch $\hat{B}_d \hat{d}_k$ at time (a) $t_k \approx 800$ s and (b) $t_k \approx 1800$ s for the tissue area containing the ROI $R$ (red solid circle) and $S$ (black dashed circle). In (a) and (b), the subfigures herein also include the MR image in the background (gray, with increased contrast for visibility). Both Fig. 13(a) and (b) illustrate that the proposed disturbance estimator is able to capture spatially varying model discrepancies. Moreover, when comparing the temperature mismatch estimate before the input power increases at $t_k \approx 800$ s in Fig. 13(a) to the estimated mismatch at $t_k \approx 1800$ s in Fig. 13(b), it is clear to see that the estimated local heat loss has increased in some areas, demonstrating the estimator’s ability to identify model mismatches that (slowly) change over time.

Especially inside the ROI, where heating is applied and, consequently, the temperature is elevated, negative values for the estimated temperature mismatch (cyan/blue) can be seen to largely coincide with the more strongly perfused tissue (gray). Also, at the edge of and outside the ROI, a striking spatial correlation can be observed, most notably so in the areas around the coordinates $(-0.01, 0.003)$, $(0.004, -0.008)$, and $(0.007, 0.006)$. In the regions outside the ROI where neither sonication occurs, and thereby modeling errors in the overestimated $B$ are not expressed, nor significant perfusion is present, the disturbance estimate mostly takes on small positive values (yellow). This is the result of the estimator identifying the effects of underestimating the thermal conductivity $\kappa$, by which more heat is transported outward from the ROI than expected by the nominal model.

In Fig. 14, a spatial resemblance can also be observed between the disturbance estimate and the heating that is applied using this value for $\hat{d}_k$ at time $t_k \approx 1800$ s. Here, the active and inactive sonication points (green and black dots, respectively) are shown inside the ROI (solid red circle). The radii of the green circles drawn around the active sonication points represent the relative magnitude of the applied acoustic power. The resulting spatial distribution of $B_{u_k}$ and the active sonication points can be seen to correspond well to the highlighted tissue in Fig. 12 and the temperature mismatch in Fig. 13(b).

From this, we draw several conclusions. First, neglecting perfusion in (2) may indeed be the cause of significant unmodeled spatially varying heat loss when the ROI containing networks of small blood vessels and the tissue in their vicinity, are visualized in darker gray. In Fig. 13, the estimated voxel-wise temperature mismatch $\hat{B}_d \hat{d}_k$ is shown for two different time instants. In addition, the subfigures herein also include the MR image in the background (gray, with increased contrast for visibility). Both Fig. 13(a) and (b)
contains (considerably) vascularized tissue. Second, the effects of perfusion increasing over time due to the body’s thermoregulatory response can be of sufficient magnitude to require an increasing steady-state power demand, as shown in Fig. 11. Next, note that although the identified heat loss shows an obvious similarity to the MR perfusion image, it is not an exact match. In addition, relating the MR image to the magnitude of the perfusion-related heat loss is not straightforward. In fact, even the relative magnitude of these effects does not exactly correspond to the MRI data, as shown in Fig. 13(b) by the disturbance estimates at (−0.001, −0.005) and (−0.002, 0.002) being larger in magnitude than the estimate at (−0.001, 0.008), whereas from Fig. 12 the latter area is seen to be most strongly perfused. Therefore, we conclude that accurately modeling the thermal effects of perfusion a priori based on an MR image as in Fig. 12 may be considered as extremely difficult, even more so due to the unavailability of extensive model identification (as previously mentioned). These conclusions motivate our offset-free MPC approach, involving the use of a simple homogeneous thermal model (2) and incorporating a disturbance estimator to allow for the identification and compensation of the potentially time-, space-, and temperature-dependent unmodeled thermal effects.

VI. Conclusion

In this article, an offset-free MPC scheme is presented to enhance the performance and robustness in temperature control using an MR-HIFU hyperthermia system for cancer treatment. Since plant-model mismatch inevitably occurs when treating different patients and tumors, and accurate thermal modeling is impracticable for MR-HIFU hyperthermia, we propose to assume a simple model with homogeneous tissue properties and (near) database parameter values, and then include a disturbance estimator to cope with the resulting model mismatch. We showed that this estimator is able to identify the effects of constant and slowly varying modeling errors (and disturbances), enabling the controller to eliminate the steady-state offset otherwise resulting from such errors. By means of experiments on the thigh muscle of a living anesthetized pig using a clinical MR-HIFU system, we have verified the feedback setup’s effectiveness in terms of recovering optimal treatment temperatures despite the presence of realistic plant-model mismatch.

APPENDIX

OPEN-LOOP MODEL VALIDATION

The plant-model mismatch in the in vivo experiment is visualized by feeding the inputs applied during the experiment, of which the total power is shown in Fig. 11, to the nominal model (5) in open loop. The nominal model’s resulting minimum/maximum (dashed line) and mean (solid line) tumor temperatures are shown in Fig. 15 (red line), and compared to the thermal response of the porcine thigh muscle measured during the experiment (black line). Clearly, there exists a severe mismatch, which, in addition, can be seen to grow over time due to the increased perfusion caused by the body’s thermoregulatory response (see also Figs. 11 and 14).

Fig. 15. Mean (solid line) and minimum/maximum (dashed line) voxel temperatures of the ROI corresponding to the experiment measurements (black line) and generated by the nominal model (red line) when supplied with the inputs applied during the in vivo experiments. The gray background indicates when the disturbance estimator was disabled during the in vivo experiments.

This illustrates the degree of model error that our offset-free MPC was able to compensate for during the in vivo experiment (see Fig. 9).

REFERENCES


B. Bram de Jager received the M.Sc. degree (cum laude) from the Dynamics and Control Group, Department of Mechanical Engineering, Eindhoven University of Technology (TU/e), Eindhoven, the Netherlands, in 2016, where he is currently pursuing the Ph.D. degree. His research focuses on model predictive temperature control strategies for magnetic-resonance-guided high-intensity focused ultrasound (MR-HIFU) hyperthermia in cancer treatment. In addition, he is interested in high-performance hybrid control solutions for linear systems.

Bert Maljaars received the M.Sc. degree (cum laude) and the Ph.D. degree in mechanical engineering from the Eindhoven University of Technology (TU/e), Eindhoven, the Netherlands, in 2012 and 2017, respectively. His Ph.D. work focused on model predictive control and actuator management in nuclear fusion devices. Afterward, he was a Post-Doctoral Researcher with the Control Systems Technology Group, TU/e, working on model predictive control and state estimation for hyperthermia applications. In 2018, he joined the Research and Technology Development Group, SKF, Houten, the Netherlands, where he is working in the areas of signal processing and state estimation.

Luukas C. Sebeke studied physics at the University of Stuttgart, Stuttgart, Germany. He is currently pursuing the Ph.D. degree with the Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, the Netherlands. He works as a Scientist at the University Hospital of Cologne, Cologne, Germany. His research interests revolve around the clinical translation of novel applications of magnetic-resonance-guided high-intensity focused ultrasound. The thermal ablation of inoperable pancreatic cancer and mild hyperthermia as an adjuvant treatment to chemotherapy are applications of particular interest.
Edwin Heijman is currently a Senior Scientist with Philips Research, Eindhoven, the Netherlands, an MR Physicist and a Specialist in MR-guided high-intensity focused ultrasound (MR-HIFU) therapy. As a member of prof. H. Grüll’s department, he is involved in preclinical experiments for local drug delivery, ultrasound-induced ablation/hyperthermia of the pancreas, and the development of MR sequences and HIFU technology. He has a particular interest in pediatric MRI and actively involving patients during their examinations. He also works as an expatriate at the Academical Hospital Cologne, Cologne, Germany, for translating new MR-HIFU applications toward clinical applications.

Holger Grüll received the Ph.D. degree in physical chemistry from the University of Cologne, Cologne, Germany, in 1996. A post-doctoral stay followed at the Ben-Gurion University of Be’er Sheva, Be’er Sheva, Israel, with research on nanoparticles and wetting phenomena. In 2000, he started at Philips Research, Eindhoven, the Netherlands, in molecular diagnostics working on nanoparticles for integrated biosensor technologies. Since 2004, he has been responsible for the research projects on molecular imaging and therapy at Philips. Currently, the major research activity is centered around therapeutic interventions using high-intensity focused ultrasound under MRI guidance. In 2007, he was appointed as a part-time Professor at the Biomedical NMR Group, Chair for Molecular Imaging and Image-guided Interventions, Eindhoven University of Technology, Eindhoven, the Netherlands. His group is focusing on molecular and multimodal imaging applications with new types of tracers and contrast agents, drug delivery, as well as image-guided therapeutic interventions. In 2016, he was appointed as a full-time Professor at the Department of Radiology, Chair for Experimental Imaging and Image-guided Therapy, University Hospital of Cologne, Cologne, Germany.

Dr. Grüll received a Fellowship of the Alexander von Humboldt Society for a post-doctoral at the National Institute of Standards and Technology, USA, working on thin films and biomimetic membranes after his Ph.D. study.

W. P. Maurice H. Heemels (Fellow, IEEE) received the M.Sc. degree (summa cum laude) in mathematics and the Ph.D. degree (summa cum laude) in control theory from the Eindhoven University of Technology (TU/e), Eindhoven, the Netherlands, in 1995 and 1999, respectively.

From 2000 to 2004, he was with the Electrical Engineering Department, TU/e, as an Assistant Professor. From 2004 to 2006, he was with the Embedded Systems Institute (ESI) as a Research fellow. Since 2006, he has been with the Department of Mechanical Engineering, TU/e, where he is currently a Full Professor. He held visiting professor positions at the Swiss Federal Institute of Technology (ETH), Zürich, Switzerland, in 2001; University of California at Santa Barbara, Santa Barbara, CA, USA, in 2008; and the University of Lorraine, Nancy, France, in 2020. His current research interests include hybrid and cyber-physical systems, networked and event-triggered control systems, and model predictive control.

Dr. Heemels served on the Editorial Board of Automatica, Nonlinear Analysis: Hybrid Systems, Annual Reviews in Control, and the IEEE Transactions on Automatic Control. He was a recipient of a personal VICI grant awarded by NWO (Dutch Research Council). He is also the Chair of the IFAC Technical Committee on Networked Systems for the term 2017–2023. He was a recipient of the 2019 IEEE L-CSS Outstanding Paper Award and was elected to the IEEE-CSS Board of Governors for the term 2021–2023.