Optimal Thermal Actuation for Mitigation of Heat-Induced Wafer Deformation

Daniël W. M. Veldman, Rob H. B. Fey, Hans Zwart, *Member, IEEE*, Marc M. J. van de Wal, Joris D. B. J. van den Boom, and Henk Nijmeijer, *Fellow, IEEE*

Abstract—**An important step in the production of integrated circuits is the projection of the pattern of electronic connections on a silicon wafer. The light used to project the pattern moves over the wafer and induces a local temperature increase. The resulting thermal expansion of the wafer leads to a significant reduction in the imaging quality of next-generation wafer scanners. Thermal actuators that move together with the projection light can be used to improve the imaging quality. Because the placement of these actuators largely determines the performance of the resulting control system, a method to support the design of an effective thermal actuator layout is presented. The proposed method computes the smallest actuation heat load that consists of a single spatial shape and respects certain constraints on wafer deformations. A gradient-based optimization algorithm is presented to compute the optimal actuation heat load. The proposed method is applied to a wafer heating problem for which the resulting shapes of the actuation heat load have a clear physical interpretation.**

Index Terms—**Optimal control, quadratic state constraints, thermal actuation, wafer heating.**

I. INTRODUCTION

PHOTOLITHOGRAPHY is a crucial production step in the production of integrated circuits (ICs). During the photolithography process, a light source projects a pattern of electronic connections onto a silicon wafer coated with a photoresist. The pattern is not projected on the whole wafer at once but only in a small area, called the slit, that moves over the surface of the wafer (see [1]). The light used to project the pattern causes a local temperature increase and thermal expansion of the wafer, which leads to a deteriorated imaging

quality. With the critical dimensions of the projected pattern approaching the subnanometer range, this process has a significant impact on the quality of the produced ICs (see [2]–[4]).

The deterioration of the imaging quality due to wafer heating can potentially be mitigated by moving thermal actuators that are placed above the wafer. The design of a thermal actuator layout is an important but nontrivial task which is critical for the performance of the resulting control system. One difficulty is that the designed thermal actuator layout should be able to reduce the wafer deformations in the slit below a certain level by using only a small amount of heating power. Furthermore, many types of thermal actuators can only heat or cool. Therefore, deciding which areas should be heated and which should be cooled is an important design decision.

The design of such a thermal actuator layout can be considered as an input selection problem which has been considered in many publications (see [5]–[8]). Because there is *a priori* knowledge about the heat load generated by the exposing light, methods that can use this information are most natural for this problem. The first of these methods has been proposed by Al-Sulaiman and Zaman [9] and by Cao *et al.* [10]. For every input set, both methods evaluate a quadratic cost function similar to the one used in the linear-quadratic tracking (LQT) problem (see [11]). An advantage of this approach is that input constraints can be included.

The main problem with the approach in [9] and [10] is that the physics of the problem is essentially governed by partial differential equations (PDEs) which means that there is a very large number of possible actuator layouts, even after spatial discretization by, for example, the finite-element (FE) method. Therefore, evaluating the cost function for every input set is computationally intractable. More efficient algorithms can be obtained by changing the problem formulation. An example of this are the results of Stadler [12] who showed for (time independent) elliptic PDEs that adding an L^1-control cost to the classical LQT cost promotes sparsity of the resulting optimal control, i.e., it results in a (time independent) optimal control that is zero in large parts of the domain and, thus, gives a good indication of effective actuator locations. Herzog *et al.* [13] later extended this idea to the control of parabolic PDEs where an additional L^1-type cost promotes the directional sparsity of the resulting control, i.e., it results in time and spatially dependent controls that are zero for all time in large parts of the domain. Kunisch *et al.* [14] further modified the problem formulation slightly to enable the design of point actuators for parabolic PDEs. This framework was extended by Boulanger

and Trautmann [15] to design point actuators for the 1-D Korteweg-de-Vries-Burgers equation.

A problem setting closely related to the results in [12] and [13] is considered in the works of Privat *et al.* [16] and Kalise *et al.* [17]. In both articles, an optimal actuation restriction problem is studied, i.e., the question "if actuation can only be applied in a certain fraction of the considered spatial domain, actuation in which area of the domain is most effective?" is answered. In [16], an analytic solution for 1-D parabolic equations with probabilistic initial conditions is derived, and in [17], a gradient-based optimization algorithm is proposed and applied to 2-D problems. It is worth mentioning that the approach in [16] can also be used to find an optimal shape for the actuation heat load. In contrast to the optimal actuation restriction problem, actuation can now be applied in the whole design domain but is required to have a fixed shape of which the intensity varies over time. However, the shapes obtained from the method in [16] are designed under probabilistic initial conditions, which means that the available information about the disturbance in the wafer heating problem cannot be used in this approach. It is worth mentioning that the optimization of a single-shape actuation heat can also be considered as a combined plant and control design problem (see [18] and [19]). In this setting, the actuation heat load is viewed as part of the plant, and only the intensity is considered as a control input.

None of the above-mentioned PDE-oriented approaches addresses how the designed actuation heat load should be realized by physical actuators. In particular, the following difficulties arise when these approaches are applied to the wafer heating problem. First of all, most existing thermal actuators are not suitable to realize controls with high (directional) sparsity that are concentrated near a few points in the spatial domain. This is even more problematic because most of these approaches allow the spatial shape of the control to vary over time inside the (small) regions where the control is concentrated. An actuator layout that can create such controls would, thus, require a high density of small thermal actuators in these regions, which is even harder to realize in reality. Second, there is the problem that thermal actuators can typically only heat or cool, which makes deciding where heaters and where coolers are placed an important aspect of the design of a thermal actuator layout. This problem is not addressed by the methods mentioned earlier. Third, the relation between the weightings in the cost function and the achieved imaging quality is not easily determined. In practice, the designer needs to meet a certain performance, i.e., the deformation on the wafer surface needs to be below a certain tolerance. However, tuning the weights in the LQT cost function such that this goal is achieved is not straightforward and will require many iterative designs.

Because of the first difficulty, the design of controls with directional sparsity is not considered in this article. Instead, the focus is on the computation of the optimal actuation heat load that consists of a single shape which leads to controls that are distributed over larger areas in space, and therefore, easier to realize by thermal actuators. This problem formulation also has the advantage that, at least when the corresponding intensity does not change sign, the shape of the actuation heat load can give a good indication of where heaters and coolers should be placed. Therefore, this approach also presents a way to address the second difficulty. The third difficulty is addressed by formulating the required imaging quality as a constraint. This leads to a state-constrained optimization problem and, to the best of our knowledge, such problems have always been studied without state constraints. It should be noted that the presented approach will not describe, just as the all other previously mentioned PDE-oriented approaches, how the designed actuation heat load should be realized by thermal actuators. However, for the considered wafer heating problem, the shape of the actuation heat load computed by our approach gives a good indication about the choice and placement of actuators that could approximately realize this shape.

In this article, a method to compute the shape and corresponding (scalar) intensity of the smallest actuation heat load preserving certain input and state constraints is presented. A gradient-based optimization algorithm is used to find the optimal shape and intensity simultaneously. The method is applied to a 2-D thermomechanical wafer heating model. The obtained shape of the actuation heat load clearly indicates where heaters and coolers should be placed and have a clear physical interpretation.

The remainder of this article is structured as follows. In Section II, the wafer heating physics is described, and the optimization problem for the actuation heat load is formulated. In Section III, the FE discretization and optimization procedure are discussed. In Section IV, the resulting actuation heat loads are presented and compared with free-shape solutions of the optimal control problem that are not required to consist of a single shape. Finally, in Section V, the conclusions are formulated and discussed.

II. MODELING AND PROBLEM FORMULATION

A. Wafer Heating Model

The wafer is a thin silicon disk, typically with a radius of 300 mm and a thickness of 0.7 mm. When the wafer is exposed to the projection light, it is placed on a water-cooled supporting structure which is assumed to have a constant temperature T_0. Because the wafer is thin, the temperature variations along the thickness are negligible and the temperature field in the wafer can be considered to be a function of the in-plane Cartesian coordinates (x, y) and time t only. The temperature increase in the wafer T w.r.t. the temperature of the supporting structure T_0 is the solution of the 2-D heat equation

$$\rho c H \frac{\partial T}{\partial t} = kH \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) - h_c T + Q \qquad (1)$$

where ρ [kg/m^3], c [J/kg/K], k [W/m/K], and H [m] are the mass density, specific heat capacity, thermal conductivity, and thickness of the wafer, respectively, h_c [W/m^2/K] is the thermal conductance between the wafer and the supporting structure, and $Q = Q(x, y, t)$ [W/m^2] is the heat load that results from the projecting light and from actuation, that is,
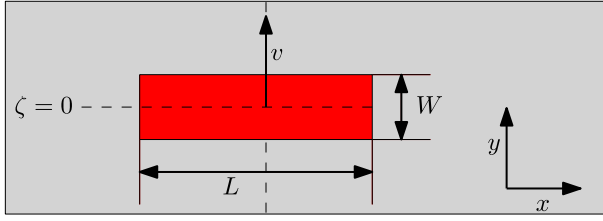
$$Q = Q_{\mathrm{exp}} + Q_{\mathrm{act}} \qquad (2)$$

Fig. 1.   Heat load (red) that is applied to the wafer (gray).

where $Q_{\exp} = Q_{\exp}(x, y, t)$ is the heat load resulting from the light that projects the pattern of electronic connections on the wafer and $Q_{\mathrm{act}} = Q_{\mathrm{act}}(x, y, t)$ is the actuation heat load. Note that convective and radiative heat transfers are negligible compared with the heat conduction to the supporting structure. It is assumed that the wafer temperature is initially equal to the temperature of the supporting structure T_0, i.e., the initial condition is $T(x, y, 0) = 0$. The spatial domain $(x, y) \in \mathbb{R}^2$ is considered to be infinite.

The heat load $Q_{\exp}$ is induced by the light that projects the pattern of electronic connections on the wafer and has a power $P_{\exp}$ [W] which is uniformly applied over the slit $\Omega_{\mathrm{slit}} \subset \mathbb{R}^2$ (the red area in Fig.1 with length L and width W). During the exposure of the wafer, the light source consecutively scans about 100 rectangular areas on the wafer, which are called fields. The scanning of a single field is considered. During the time interval $t \in (0, t_e)$ in which a single field is scanned, the slit moves with a constant velocity v in the positive y-direction

$$Q_{\exp}(x, y, t) = B_{\exp}(x, y - vt)u_{\exp}(t) \tag{3}$$

where $B_{\exp}(x, y - vt)$ [1/m^2] and $u_{\exp}(t)$ [W] are the shape and intensity of the expose load, respectively. In particular, the shape $B_{\exp}(x, y - vt)$ equals $1/LW$ when $x \in [-L/2, L/2]$ and $y - vt \in [-W/2, W/2]$ and zero otherwise, and the intensity $u_{\exp}(t) = P_{\exp}$ is constant. Note that the assumption of an infinite spatial domain limits our analysis to fields that are not close to the wafer edge. This assumption is valid for the majority of fields on the wafer (see [1]). The actuation heat load Q_{act} will be discussed in Section II-B.

Because the considered domain is infinite and the applied heat load is moving, it is convenient to consider a moving coordinate system $(x, \zeta, t) = (x, y - vt, t)$ in which the shape of $Q_{\exp}$ is fixed. Let $T^{(y)}(x, y, t)$ and $Q^{(y)}(x, y, t)$ denote the temperature field and applied heat load expressed in (x, y, t)-coordinates as in (1). The temperature field $T^{(\zeta)}(x, \zeta, t)$ and applied heat load $Q^{(\zeta)}(x, \zeta, t)$ expressed in (x, ζ, t)-coordinates are then equal to $T^{(y)}(x, \zeta + vt, t)$ and $Q^{(y)}(x, \zeta + vt, t)$, respectively. It can be shown that $T(x, \zeta, t) = T^{(\zeta)}(x, \zeta, t)$ satisfies (see [20, Ch. 11])

$$\rho c H \left(\frac{\partial T}{\partial t} - v \frac{\partial T}{\partial \zeta} \right) = kH \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial \zeta^2} \right) - h_c T + Q \tag{4}$$

with $Q(x, \zeta, t) = Q^{(\zeta)}(x, \zeta, t)$. In particular, the heat load induced by the expose load $Q_{\exp}^{(y)}(x, y, t)$ in (3) becomes

$$Q_{\exp}^{(\zeta)}(x, \zeta, t) = B_{\exp}(x, \zeta)u_{\exp}(t). \tag{5}$$

In the remainder of this article, only the (x, ζ, t)-coordinate system will be used and the used coordinate system will no longer be indicated, $T = T(x, \zeta, t) = T^{(\zeta)}(x, \zeta, t)$ and $Q = Q(x, \zeta, t) = Q^{(\zeta)}(x, \zeta, t)$. Note that the origin of the (x, ζ)-coordinate system is at the center of the slit (see Fig.1).

The mechanical model used to predict the resulting in-plane displacement field is based on linear strain-displacement relations and the plane-stress relations for an isotropic material. It is assumed that inertia effects are negligible, which is a standard assumption in thermomechanical models (see [21]). The resulting model in ζ-coordinates takes the form

$$\frac{EH}{1 - \nu^2} \left(\frac{\partial^2 d_x}{\partial x^2} + \frac{1 - \nu}{2} \frac{\partial^2 d_x}{\partial \zeta^2} + \frac{1 + \nu}{2} \frac{\partial^2 d_\zeta}{\partial x \partial \zeta} \right) - k_s d_x$$
$$= \frac{\alpha EH}{1 - \nu} \frac{\partial T}{\partial x} \tag{6}$$

$$\frac{EH}{1 - \nu^2} \left(\frac{\partial^2 d_\zeta}{\partial \zeta^2} + \frac{1 - \nu}{2} \frac{\partial^2 d_\zeta}{\partial x^2} + \frac{1 + \nu}{2} \frac{\partial^2 d_x}{\partial x \partial \zeta} \right) - k_s d_\zeta$$
$$= \frac{\alpha EH}{1 - \nu} \frac{\partial T}{\partial \zeta} \tag{7}$$

where $d_x = d_x(x, \zeta, t)$ [m] and $d_\zeta = d_\zeta(x, \zeta, t)$ [m] are the displacement field components in the x- and ζ-directions, respectively, E [N/m^2], ν [−], and α [1/K] are the Young's modulus, Poisson's ratio, and coefficient of thermal expansion of the wafer, respectively, and k_s [N/m^3] represents the shear stiffness of the supporting structure per unit area.

The applied actuation heat load should compensate the light-induced heat load $Q_{\exp}$ such that a sufficiently good imaging quality is attained. This is achieved when the deformation in the slit is below a certain threshold δ_{slit}. It is, thus, required that for all $(x, \zeta) \in \Omega_{\mathrm{slit}}$

$$d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t) \leq \delta_{\mathrm{slit}}^2. \tag{8}$$

The applied actuation heat load should also prevent slip between the wafer and the supporting structure, which means that the displacement in the whole wafer surface should be below the threshold δ_{slip}. It is thus required that for all $(x, \zeta) \in \mathbb{R}^2$

$$d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t) \leq \delta_{\mathrm{slip}}^2. \tag{9}$$

It will be convenient to write (8) and (9) as one inequality

$$d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t) \leq d_{\max}^2(x, \zeta) \tag{10}$$

where $d_{\max}(x, \zeta) = \min\{\delta_{\mathrm{slit}}, \delta_{\mathrm{slip}}\}$ when $(x, \zeta) \in \Omega_{\mathrm{slit}}$ and $d_{\max}(x, \zeta) = \delta_{\mathrm{slip}}$ otherwise.

Finally, the constraint (10) should be achieved by the smallest possible actuation heat load. In particular, it is required that the actuation heat load has minimal (squared) L^2-norm

$$J_0 = \int_0^{t_e} \iint_{\mathbb{R}^2} Q_{\mathrm{act}}^2(x, \zeta, t) \, dx \, d\zeta \, dt. \tag{11}$$

In short, the cost functional J_0 in (11) should be minimized over the actuation heat loads $Q_{\mathrm{act}}(x, \zeta, t)$ that result in a displacement field for which the constraint (10) is satisfied.

B. Actuation Heat Load

The heat load Q_{act} should be realized by a layout of thermal actuators. The case where thermal actuators are placed above the wafer is considered. Because of the current design of wafer scanners where the wafer is placed on a stage that moves underneath the expose light (see [22]), it is most natural to assume that these actuators are fixed to the machine frame. This means that their absolute speed is zero and that their relative speed w.r.t. the wafer is the same as the relative speed of the expose load. To get insight in an effective placement of actuators, it is required that the actuation heat load has a fixed shape in the moving (x, ζ)-coordinate system, that is,

$$Q_{\text{act}}(x, \zeta, t) = B(x, \zeta)u(t) \tag{12}$$

where $B(x, \zeta)$ [1/m²] and $u(t)$ [W] are the shape and intensity of the actuation heat load, respectively. It is not possible to apply thermal actuation in the area where the pattern is currently projected because the required thermal actuators will block the projecting light source. Since the light source projects the pattern in the area $\Omega_{\text{slit}} \subset \mathbb{R}^2$, we, thus, require

$$B(x, \zeta) = 0, \quad \text{for } (x, \zeta) \in \Omega_{\text{slit}}. \tag{13}$$

Note that this constraint prevents to have $Q_{\text{act}} = -Q_{\text{exp}}$, i.e., the situation where the actuation heat load cancels the expose load, which would lead to $d_x(x, \zeta, t) = d_\zeta(x, \zeta, t) = 0$. The results in Section IV will demonstrate that there are values of δ_{slit} and δ_{slip} for which there is no actuation heat load such that (10) and (13) are satisfied. Furthermore, an important aspect of the design of a thermal actuator layout is that thermal actuators can typically only heat or cool. By requiring that the intensity $u(t)$ is nonnegative, that is,

$$u(t) \geq 0 \tag{14}$$

it is clear that areas where the designed shape of the actuation heat load $B(x, \zeta)$ is positive should be heated and that areas where $B(x, \zeta)$ is negative should be cooled.

Note that the representation $(B(x, \zeta), u(t))$ of $Q_{\text{act}}(x, \zeta, t)$ in (12) is clearly nonunique: if $(B_0(x, \zeta, t), u_0(t))$ is a representation satisfying (14), then the same actuation heat load $B_0(x, \zeta)u_0(t)$ can be represented by $(\beta B_0(x, \zeta), u_0(t)/\beta)$ for any $\beta > 0$. This nonuniqueness can be removed by normalizing the shape of the actuation heat load $B(x, \zeta)$ such that is has unit L^1-norm, that is,

$$\iint_{\mathbb{R}^2} |B(x, \zeta)| \, dx \, d\zeta = 1. \tag{15}$$

Note that this normalization is consistent with $B(x, \zeta)$ having the unit [1/m²] and $u(t)$ having the unit [W]. In addition, note that this normalization and the constraint on the sign of $u(t)$ in (14) gives $u(t)$ the interpretation of total applied actuation heat, that is,

$$u(t) = \iint_{\mathbb{R}^2} |Q_{\text{act}}(x, \zeta, t)| \, dx \, d\zeta. \tag{16}$$

It is now clear that the constraint on $u(t)$ in (14) and normalization of $B(x, \zeta)$ in (15) make the representation of a single-shape actuation heat load $B(x, \zeta)u(t)$ unique. Note that this (obviously) does not imply that the optimal

single-shape actuation heat load will be unique. To simplify the optimization algorithm, the normalization (15) will not be enforced during the optimization process but is only applied afterward to obtain results that easier to interpret and compare.

Finding a solution to the obtained optimization problem is not trivial because of the nonlinear constraint on the resulting displacements in (10) and because of the combined optimization of the shape and intensity of the actuation heat load in (12). The first problem will be addressed by the two-step solution procedure described in Section II-C. The second problem will be addressed by the optimization algorithm in Section III.

C. Two-Step Solution Procedure

Because the nonlinear state constraint as in (10) leads to a difficult (and possibly infeasible) optimization problem, a two-step optimization procedure is proposed, in which two considerably simpler optimization problems that only involve linear input constraints need to be solved. In the first step, it is attempted to find an actuation heat load Q_{act} for which the constraint (10) is satisfied. Note that depending on the values of δ_{slit} and δ_{slip} such a solution might not exist and that it is not trivial to find such a solution. In the second step, the admissible solution found in the first step is used as a starting point for the minimization of the cost functional J_0 in (11) subject to the constraint (10).

1) Finding an Admissible Solution: To find an admissible solution, i.e., an actuation heat load that respects the constraint (10), the following penalty functional is considered:

$$J_1 = \int_0^{t_e} \iint_{\mathbb{R}^2} \left[d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t) - d_{\text{max}}^2(x, \zeta)\right]^+$$
$$\times \, dx \, d\zeta \, dt \tag{17}$$

where the function $[\,\cdot\,]^+ : \mathbb{R} \to \mathbb{R}$ is defined by

$$[a]^+ = \begin{cases} a, & \text{when } a \geq 0 \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

Note that the integrand only contributes to the value of J_1 when $d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t) - d_{\text{max}}^2(x, \zeta) > 0$, so when the maximally allowed displacement in (10) is exceeded. In addition, note that $J_1 \geq 0$ always and that if $J_1 = 0$, the maximally allowed displacement $d_{\text{max}}(x, \zeta)$ is not exceeded. An admissible solution can, thus, be found by minimizing J_1.

Note that J_1 is convex (but not strictly convex) in the displacement field components d_x and d_ζ, and that by linearity of (4), (6), and (7), J_1 is also convex in the applied heat load Q_{act}. This implies that varying $B(x, \zeta)$ while keeping $u(t)$ fixed and varying $u(t)$ while keeping $B(x, \zeta)$ fixed cannot decrease J_1 further than the value of J_1 at a local minimum (see [23]). However, the simultaneous minimization of J_1 over $B(x, \zeta)$ and $u(t)$ is nonconvex. It is, therefore, possible that an admissible solution exists even if a (local) minimum with $J_1 > 0$ is found.

2) Finding the Optimal Solution: In case an admissible actuation heat load $Q_{\text{act}}(x, \zeta, t) = B(x, \zeta)u(t)$ has been found, it has been established that the admissible set is nonempty. The problem that remains is finding the point in the admissible set for which the L^2-norm in (11) is minimal. Since a point in the admissible set has already been found, it is natural to use this as a starting point for our search. The non-linear state constraint (10) is preserved during the iterations by the barrier function method (see [24]). The barrier functional is defined as $J_2 = \infty$ when the constraint (10) is not satisfied for any $(x, \zeta) \in \mathbb{R}^2$ and $0 \le t \le t_e$ and otherwise as

$$
\begin{aligned}
J_2 &= \int_0^{t_e} \iint_{\mathbb{R}^2} \left[-1 + \left(\frac{d_{\max}^2(x, \zeta)}{d_{\max}^2(x, \zeta) - d_x^2(x, \zeta, t) - d_\zeta^2(x, \zeta, t)} \right)^p \right] \\
&\quad \times dx \ d\zeta \ dt
\end{aligned}
\tag{19}
$$

where the power $p > 1$ can be tuned. The term "-1" assures that $J_2 = 0$ when $d_x(x, \zeta, t) = d_\zeta(x, \zeta, t) = 0$. The constraint (10) is preserved during iterations by minimizing

$$
J_0 + w J_2
\tag{20}
$$

where $w > 0$ is a weight that can be tuned. The tuning of the power p and the weight w is discussed in Appendix A.

It follows that $J_0 + w J_2$ is strictly convex in the actuation heat load Q_{act}. Minimizing $J_0 + w J_2$ for a fixed shape of the actuation heat load $B(x, \zeta)$ yields, therefore, a unique optimal intensity $u(t)$, and minimizing $J_0 + w J_2$ for a fixed intensity $u(t)$ yields a unique shape of the actuation heat load $B(x, \zeta)$ (see [25, Proposition 10.6]). Whether the optimal single-shape actuation heat load $B(x, \zeta)u(t)$ is unique has not been proven, but the numerical results in Section IV suggest this is the case.

III. Spatial Discretization and Optimization

A. Finite-Element Method

The FE method (see [26]) is used to solve the thermal model (4) and the thermomechanical model (6) and (7), which enables the computation of the cost functionals J_0, J_1, and J_2 in (11), (17), and (19), respectively. Note that the considered problem in Fig. 1 is symmetric in $x = 0$. Therefore, only the domain $x \ge 0$ is considered and symmetric boundary conditions ($\partial T / \partial x = 0$) are applied at $x = 0$.

For the thermal model (4), the temperature field $T(x, \zeta, t)$ is expressed as a linear combination of FE shape functions as

$$
T(x, \zeta, t) = \mathbf{N}(x, \zeta)\boldsymbol{\theta}(t)
\tag{21}
$$

where $\mathbf{N}(x, \zeta)$ is the (row) vector of (linear) FE shape functions, and $\boldsymbol{\theta}(t)$ is the (column) vector of nodal temperatures. The actuation heat load Q_{act} of the form (12) is expressed as

$$
Q_{\text{act}}(x, \zeta, t) = \mathbf{N}(x, \zeta)\mathbf{B}u(t)
\tag{22}
$$

where $\mathbf{B}$ is the (column) vector with the nodal values of the shape $B(x, \zeta)$ of the actuation heat load in (12). Note that the shape functions $\mathbf{N}(x, \zeta)$ will describe the temperature field in a bounded domain, so that this step involves the truncation of the infinite domain $x \ge 0$ to a bounded domain Ω which

should be chosen large enough to accurately approximate the solution on the infinite domain.

For the thermal FE model, all edges of the bounded domain Ω are perfectly insulated, i.e., $\partial T / \partial n = 0$ on the edge $\partial \Omega$, where n is the normal to the edge. A Galerkin discretization of (4) with this boundary condition that also incorporates the constraint (13) takes the form

$$
\mathbf{E}\dot{\boldsymbol{\theta}}(t) = \mathbf{A}\boldsymbol{\theta}(t) + \mathbf{B}_{\text{exp}}u_{\text{exp}}(t) + \mathbf{E}_{\mathbf{B}}\mathbf{B}u(t)
\tag{23}
$$

with initial condition $\boldsymbol{\theta}(0) = \mathbf{0}$ and

$$
\mathbf{E}_0 = \iint_\Omega \mathbf{N}^\top \mathbf{N} \ dx \ d\zeta, \quad \mathbf{E} = \rho c H \mathbf{E}_0
\tag{24}
$$

$$
\mathbf{B}_{\text{exp}} = \iint_\Omega \mathbf{N}^\top B_{\text{exp}} \ dx \ d\zeta
\tag{25}
$$

$$
\mathbf{E}_{\mathbf{B}} = \iint_{\Omega \backslash \Omega_{\text{slit}}} \mathbf{N}^\top \mathbf{N} \ dx \ d\zeta
\tag{26}
$$

where the dependence of $\mathbf{N}(x, \zeta)$ and $B_{\text{exp}}(x, \zeta)$ on x and ζ has been dropped. Here, $\mathbf{E}$ is the heat capacity matrix. The heat conduction matrix $\mathbf{A}$ is given in Appendix B. Note that the constraint (13) is incorporated in the matrix $\mathbf{E}_{\mathbf{B}}$.

The mechanical FE model uses the same shape functions as the thermal FE model. The displacement field components $d_x(x, \zeta, t)$ and $d_\zeta(x, \zeta, t)$ are, thus, approximated as

$$
d_x(x, \zeta, t) = \mathbf{N}(x, \zeta)\mathbf{d}_x(t), \quad d_\zeta(x, \zeta, t) = \mathbf{N}(x, \zeta)\mathbf{d}_\zeta(t)
\tag{27}
$$

where $\mathbf{d}_x(t)$ and $\mathbf{d}_\zeta(t)$ are the (column) vectors of nodal displacement field components in the x- and ζ-directions, respectively. Application of the Galerkin method to (6) and (7) based on the expressions for $T(x, \zeta, t)$ in (21) and the expressions for $d_x(x, \zeta, t)$ and $d_\zeta(x, \zeta, t)$ in (27) (with free boundary conditions, i.e., no external force is applied at the edges) yields

$$
\mathbf{K}\mathbf{d}(t) = \mathbf{L}\boldsymbol{\theta}(t)
\tag{28}
$$

with $\mathbf{K}$ the stiffness matrix, $\mathbf{L}$ the thermal load matrix, and

$$
\mathbf{d}(t) = \begin{bmatrix} \mathbf{d}_x(t) \\ \mathbf{d}_\zeta(t) \end{bmatrix}.
\tag{29}
$$

Explicit expressions for $\mathbf{K}$ and $\mathbf{L}$ in terms of the FE shape functions $\mathbf{N}(x, \zeta)$ are given in Appendix B.

At the edge $x = 0$, a symmetric boundary condition is needed, which means that the displacement in the x-direction is zero and the displacement in the ζ-direction is free. In the FE model, this means that the entries of $\mathbf{d}_x(t)$ in (27) that correspond to nodes on the line $x = 0$ are zero. The mechanical degrees of freedom (DOFs) $\mathbf{d}(t)$ in (29) can, thus, be partitioned into free DOFs $\mathbf{d}_{\text{f}}(t)$ and constrained DOFs $\mathbf{d}_{\text{c}}(t)$, where the constrained DOFs are zero, i.e., $\mathbf{d}_{\text{c}}(t) = \mathbf{0}$. It follows that there exists a Boolean matrix $\mathbf{S}$ such that

$$
\mathbf{d}_{\text{f}}(t) = \mathbf{S}\mathbf{d}(t), \quad \mathbf{d}(t) = \mathbf{S}^\top \mathbf{d}_{\text{f}}(t).
\tag{30}
$$

Inserting the second identity in (28) and multiplying the resulting equation by $\mathbf{S}$ yields an equation for the free DOFs

$$
\mathbf{S}\mathbf{K}\mathbf{S}^\top \mathbf{d}_{\text{f}}(t) = \mathbf{S}\mathbf{L}\boldsymbol{\theta}(t).
\tag{31}
$$

Solving this equation for $\mathbf{d}_f(t)$ and inserting the result back into the second identity in (30) yields an expression for the nodal displacements in which the symmetric boundary condition is considered

$$\mathbf{d}(t) = \begin{bmatrix} \mathbf{d}_x(t) \\ \mathbf{d}_\zeta(t) \end{bmatrix} = \mathbf{S}^\top (\mathbf{S}\mathbf{K}\mathbf{S}^\top)^{-1}\mathbf{S}\mathbf{L}\boldsymbol{\theta}(t). \quad (32)$$

The discretization of the cost functional J_0 in (11) is made based on the expression for $Q_{\text{act}}(x, \zeta, t)$ in (22), which yields

$$J_0 = \int_0^{t_e} u(t)\mathbf{B}^\top \mathbf{E}_0 \mathbf{B} u(t) \ dt \quad (33)$$

with $\mathbf{E}_0$ as in (24). The nonquadratic cost functionals J_1 and J_2 are discretized by nodal interpolation. This means that a scalar nonlinear function $f(x, \zeta)$ on the domain $(x, \zeta) \in \mathbb{R}^2$ is approximated in terms of the FE shape functions $\mathbf{N}(x, \zeta)$ by $\mathbf{N}(x, \zeta)\mathbf{f}$, where $\mathbf{f}$ is a vector containing the nodal values of $f(x, \zeta)$ at the nodes of the FE mesh. The integral of $f(x, \zeta)$ over $(x, \zeta) \in \mathbb{R}^2$ is then approximated as

$$\iint_{\mathbb{R}^2} f(x, \zeta) \ dx \ d\zeta$$
$$\approx \iint_\Omega \mathbf{N}(x, \zeta)\mathbf{f} \ dx \ d\zeta$$
$$= \iint_\Omega \mathbf{1}^\top \mathbf{N}^\top(x, \zeta)\mathbf{N}(x, \zeta)\mathbf{f} \ dx \ d\zeta = \mathbf{1}^\top \mathbf{E}_0 \mathbf{f} \quad (34)$$

where $\mathbf{1}$ is a column vector of ones. For the second identity, it is assumed that $\mathbf{N}(x, \zeta)\mathbf{1} = 1$ for all $(x, \zeta) \in \Omega$. This assumption is satisfied for standard Lagrangian FE shape functions $\mathbf{N}(x, \zeta)$ (see [26]). The nonquadratic cost functionals J_1 and J_2 in (17) and (19) are, thus, discretized as

$$J_1 = \int_0^{t_e} \mathbf{1}^\top \mathbf{E}_0 \left[\mathbf{d}_x^2(t) + \mathbf{d}_\zeta^2(t) - \mathbf{d}_{\max}^2 \right]^+ dt \quad (35)$$

$$J_2 = \int_0^{t_e} \mathbf{1}^\top \mathbf{E}_0 \left(\left[\frac{\mathbf{d}_{\max}^2}{\mathbf{d}_{\max}^2 - \mathbf{d}_x^2(t) - \mathbf{d}_\zeta^2(t)} \right]^p - \mathbf{1} \right) dt \quad (36)$$

where the operations $[\cdot]^+$, $(\cdot)^2$, $[\cdot]^p$, and $\cdot/\cdot$ are applied componentwise, and $\mathbf{d}_{\max}$ is a vector containing the values of $d_{\max}(x, \zeta)$ in the nodes of the FE model.

B. Sensitivity Analysis

As discussed in Section II-C, two optimization problems need to be solved. In the first optimization problem, the cost functional is $J = J_1$ and in the second optimization problem the cost functional is $J = J_0 + wJ_2$. In this section, the computation of the gradients of these cost functionals w.r.t. the discretized shape of the actuation heat load $\mathbf{B}$ and the intensity $u(t)$ is discussed. These will be used in the gradient-based optimization procedure in Section III-C.

The computation of the gradient w.r.t. the intensity $u(t)$ is based on the following well-known result. Consider a cost functional of the form

$$J = J(\mathbf{B}, u) = \int_0^{t_e} (\bar{f}(\boldsymbol{\theta}(t)) + g(\mathbf{B}, u(t))) \ dt \quad (37)$$

where the state $\boldsymbol{\theta}(t)$ depends on $\mathbf{B}$ and $u(t)$ through (23). The gradient $\nabla_u J(\mathbf{B}, u)$ is defined by the property that for any variation $\tilde{u}(t)$ of $u(t)$, it holds that

$$\langle \nabla_u J(\mathbf{B}, u), \tilde{u} \rangle_u = \lim_{\varepsilon \to 0} \frac{J(\mathbf{B}, u + \varepsilon \tilde{u}) - J(\mathbf{B}, u)}{\varepsilon} \quad (38)$$

where $\langle \cdot \cdot \rangle_u$ is the inner product on the space of intensities

$$\langle u_1(t), u_2(t) \rangle_u = \int_0^{t_e} u_1(t)u_2(t) \ dt. \quad (39)$$

The gradient $\nabla_u J(\mathbf{B}, u)$ can be computed based on the adjoint state $\boldsymbol{\varphi}(t)$ (see [11] and [27])

$$(\nabla_u J(\mathbf{B}, u))(t) = \mathbf{B}^\top \mathbf{E}_{\mathbf{B}} \boldsymbol{\varphi}(t) + \frac{\partial g}{\partial u}(\mathbf{B}, u(t)) \quad (40)$$

where $\partial g/\partial u$ denotes the partial derivative of g w.r.t. its second argument and $\boldsymbol{\varphi}(t)$ is the solution of

$$-\mathbf{E}^\top \dot{\boldsymbol{\varphi}}(t) = \mathbf{A}^\top \boldsymbol{\varphi}(t) + \left(\frac{\partial \bar{f}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}(t)) \right)^\top, \quad \boldsymbol{\varphi}(t_e) = \mathbf{0}. \quad (41)$$

Note that the gradient $\nabla_u J(\mathbf{B}_0, u_0)$ is now computed in the following steps: 1) compute $\boldsymbol{\theta}_0(t)$ as the solution of (23) with $\mathbf{B} = \mathbf{B}_0$ and $u(t) = u_0(t)$; 2) compute $\boldsymbol{\varphi}_0(t)$ from (41) with $\boldsymbol{\theta}(t) = \boldsymbol{\theta}_0(t)$, $\mathbf{B} = \mathbf{B}_0$, and $u(t) = u_0(t)$; and 3) compute $\nabla J(\mathbf{B}_0, u_0)$ from (40) with $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}_0(t)$, $\boldsymbol{\theta}(t) = \boldsymbol{\theta}_0(t)$, $\mathbf{B} = \mathbf{B}_0$, and $u(t) = u_0(t)$. In addition, note that the solution of (41) is computed by integrating backward in time starting from the final condition $\boldsymbol{\varphi}(t_e) = \mathbf{0}$.

The cost functionals $J = J_1$ and $J = J_0 + wJ_2$ with J_0, J_1, and J_2 as in (33), (35), and (36) are written in the form

$$J = J(\mathbf{B}, u) = \int_0^{t_e} (f(\mathbf{d}(t)) + g(\mathbf{B}, u(t))) \ dt \quad (42)$$

where $\mathbf{d}(t)$ is the vector of nodal displacements resulting from $\mathbf{B}$ and $u(t)$, i.e., $\mathbf{d}(t)$ is computed by solving $\boldsymbol{\theta}(t)$ from (23) and then $\mathbf{d}(t)$ from (32). Note that (42) can be written in the form (37) by setting

$$\bar{f}(\boldsymbol{\theta}(t)) = f(\mathbf{S}^\top (\mathbf{S}\mathbf{K}\mathbf{S}^\top)^{-1}\mathbf{S}\mathbf{L}\boldsymbol{\theta}(t)). \quad (43)$$

In particular, it follows that

$$\frac{\partial \bar{f}}{\partial \boldsymbol{\theta}} = \frac{\partial f}{\partial \mathbf{d}} \frac{\partial \mathbf{d}}{\partial \boldsymbol{\theta}} = \frac{\partial f}{\partial \mathbf{d}} \mathbf{S}^\top (\mathbf{S}\mathbf{K}\mathbf{S}^\top)^{-1}\mathbf{S}\mathbf{L}. \quad (44)$$

Using (44), (41) can now be expressed in terms of the function f instead of $\bar{f}$ so that the gradient of the cost functional (42) can also be computed based on the adjoint state. Explicit expressions for f, g, $\partial f/\partial \mathbf{d}$, and $\partial g/\partial u$ when $J = J_1$ or $J = J_0 + wJ_2$ are given in Appendix C.

Similarly, the gradient $\nabla_{\mathbf{B}} J(\mathbf{B}, u)$ is defined by the property that for any variation $\tilde{\mathbf{B}}$ of $\mathbf{B}$, it holds that

$$\langle \nabla_{\mathbf{B}} J(\mathbf{B}, u), \tilde{\mathbf{B}} \rangle_B = \lim_{\varepsilon \to 0} \frac{J(\mathbf{B} + \varepsilon \tilde{\mathbf{B}}, u) - J(\mathbf{B}, u)}{\varepsilon} \quad (45)$$

where $\langle \cdot \cdot \rangle_B$ is the inner product on the space of actuation heat load shapes (which are parametrized by their nodal values according to (22)) defined by

$$\langle \mathbf{B}_1, \mathbf{B}_2 \rangle_B = \mathbf{B}_1^\top \mathbf{E}_0 \mathbf{B}_2. \quad (46)$$

In Appendix D, it is shown that the gradient of a cost functional J w.r.t. $\mathbf{B}$ of the form (37) can be expressed in terms of the adjoint state $\varphi(t)$ that satisfies (41) as

$$\nabla_{\mathbf{B}}J(\mathbf{B}, u) = \mathbf{E}_0^{-1}\mathbf{E}_{\mathbf{B}}\int_0^{t_e}\varphi(t)u(t)\ dt$$

$$+\mathbf{E}_0^{-1}\int_0^{t_e}\left(\frac{\partial g}{\partial \mathbf{B}}(\mathbf{B}, u(t))\right)^{\top}dt. \quad (47)$$

The procedure to compute $\nabla_{\mathbf{B}}J(\mathbf{B}, u)$ is similar to the procedure to compute $\nabla_u J(\mathbf{B}, u)$ described earlier. Using (44), this result can again be applied for $J = J_1$ and $J = J_0 + wJ_2$ which are of the form (42). Explicit formulas for $\partial g/\partial \mathbf{B}$ are given in Appendix C.

C. Optimization Algorithm

With the actuation heat load that consists of a single shape as in (22), both the shape of the discretized actuation heat load $\mathbf{B}$ and the intensity $u(t)$ need to be optimized. Since J_1 and $J_0 + wJ_2$ are convex in u, minimizing $J(\mathbf{B}, u)$ over u for fixed $\mathbf{B}$ yields a unique minimal value of J, although the minimizer u is not necessarily unique (see [23]). The function $\hat{J}(\mathbf{B}) = \min_u J(\mathbf{B}, u)$ is, thus, well defined. A possible approach is to minimize $\hat{J}(\mathbf{B})$ over $\mathbf{B}$. One can easily verify that $\hat{J}(\mathbf{B}) = \hat{J}(\beta\mathbf{B})$ for any $\beta > 0$. This scaling freedom can be removed by normalizing $\mathbf{B}$, e.g., using (15) or by requiring that $\|\mathbf{B}\|_B^2 = \langle\mathbf{B}, \mathbf{B}\rangle_B = 1$. In the latter case, $\hat{J}(\mathbf{B})$ is in fact minimized over the Stiefel manifold, on which there are typically multiple local minima (see [28] and [29]). Another disadvantage of this approach is that every evaluation of $\hat{J}(\mathbf{B})$ requires the solution of an optimal control problem, which is costly.

We, therefore, use a different approach in which the shape $\mathbf{B}$ and intensity $u(t)$ of the actuation heat load are updated alternately. In this way, it is prevented that a lot of time is spent on the computation of the optimal intensity $u(t)$ for a suboptimal shape of the actuation heat load $\mathbf{B}$, which typically happens when the minimization of $\hat{J}(\mathbf{B})$ is considered. The resulting algorithm takes the following form.

1: Initial guess $\mathbf{B}^{(0)}$ and $u^{(0)} \neq 0$
2: $k = 0$
3: **while** $k <$ `max_iters` **do**
4:     $\mathbf{B}^{(k+1)} = \text{UpdateB}(\mathbf{B}^{(k)}, u^{(k)})$
5:     $u^{(k+1)} = \text{UpdateU}(\mathbf{B}^{(k+1)}, u^{(k)})$
6:     **if** Converged($\mathbf{B}^{(k)}, \mathbf{B}^{(k+1)}, u^{(k)}, u^{(k+1)}$) **then**
7:         **return** $(\mathbf{B}^{(k+1)}, u^{(k+1)})$
8:     **end if**
9:     $k = k + 1$
10: **end while**

Here, $\mathbf{B}^{(k)}$ and $u^{(k)}(t)$ denote the kth iterate of $\mathbf{B}$ and $u(t)$, respectively, `max_iters` denotes the maximum number of iterations, "UpdateB" and "UpdateU" represent the update schemes for $\mathbf{B}$ and u, and "Converged" represents the method that checks for convergence. The methods "UpdateB, " "UpdateU, " and "Converged" will be elaborated in the following.

Note that it is possible to choose the initial guess $\mathbf{B}^{(0)} = \mathbf{0}$ but that the initial guess $u^{(0)}$ must be nonzero because

otherwise $\nabla_{\mathbf{B}}J(\mathbf{B}^{(0)}, u^{(0)}) = \mathbf{0}$, which leads to a problem in "UpdateB" in line 4. Swapping lines 4 and 5 such that first u is updated and then $\mathbf{B}$ is updated leads to an algorithm in which the initial guess $u^{(0)}$ can be zero and $\mathbf{B}^{(0)}$ must be nonzero. However, this form is not presented here because it is easier to generate an initial guess for $u(t)$, which depends on one variable, than for $\mathbf{B}$, which represents a function depending on the two spatial variables (x, ζ) and because the considered examples have shown good convergence starting from the initial guess $\mathbf{B}^{(0)} = \mathbf{0}$. It is recommended to use an initial guess $u^{(0)}$ that satisfies $u^{(0)}(t_e) = 0$ because $u(t_e) = 0$ must hold at the minimum of J.

The update procedure "UpdateB" in line 4 of the algorithm updates the current iterate $\mathbf{B}^{(k)}$ in the direction of the gradient

$$\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} - h_{\mathbf{B}}\nabla_{\mathbf{B}}J(\mathbf{B}^{(k)}, u^{(k)}). \quad (48)$$

The stepsize $h_{\mathbf{B}} > 0$ is estimated based on a quadratic approximation of $J(\mathbf{B}^{(k+1)}, u^{(k)})$ around $h_{\mathbf{B}} = 0$

$$J(\mathbf{B}^{(k+1)}, u^{(k)}) = J(\mathbf{B}^{(k)}, u^{(k)}) + hG_{\mathbf{B}}^{(k)} + \frac{1}{2}h^2 H_{\mathbf{B}}^{(k)} \quad (49)$$

with $G_{\mathbf{B}}^{(k)} = -\langle\nabla_{\mathbf{B}}J(\mathbf{B}^{(k)}, u^{(k)}), \nabla_{\mathbf{B}}J(\mathbf{B}^{(k)}, u^{(k)})\rangle_B$ and

$$H_{\mathbf{B}}^{(k)} = \frac{\partial^2}{\partial h^2}[J(\mathbf{B}^{(k)} + h\nabla_{\mathbf{B}}J(\mathbf{B}^{(k)}, u^{(k)}), u^{(k)})]_{h=0}. \quad (50)$$

The computation of $H_{\mathbf{B}}^{(k)}$ is discussed in Appendix E. The stepsize $h_{\mathbf{B}}$ can now be estimated by minimizing the RHS of (49), which leads to $h_{\mathbf{B},\text{opt}} = -G_{\mathbf{B}}^{(k)}/H_{\mathbf{B}}^{(k)}$. Since the considered cost functions are not quadratic, the expansion in (49) is an approximation and there is no guarantee that this choice indeed leads to a decrease in the cost function J. If an increase is observed, the step size is halved until a decrease in J is observed. Note that this will eventually happen because (away from a critical point of J where $\nabla_{\mathbf{B}}J = 0$) $G_{\mathbf{B}}^{(k)} < 0$ and the linear term in (49) will dominate for $h_{\mathbf{B}}$ small enough.

The update described by "UpdateU" in line 5 of the algorithm preserves the constraint (14) through the projected gradient method (see [30]). The current iterate $u^{(k)}(t)$ is, thus, updated according to

$$u^{(k+1)}(t) = \Pi_u(u^{(k)} - h_u\nabla_u J(\mathbf{B}^{(k+1)}, u^{(k)}))(t) \quad (51)$$

where $h_u > 0$ denotes the step size and Π_u denotes the projection onto the admissible set given by

$$(\Pi_u(u))(t) = \begin{cases} 0, & \text{when } u(t) \leq 0 \\ u(t), & \text{otherwise.} \end{cases} \quad (52)$$

Note that $u^{(k+1)}$ does not depend linearly on h_u. For example, when $u^{(k)}(t_1) = 1$ and $(\nabla_u J)(t_1) = 1$ on a certain time instant $t = t_1$, then $u^{(k+1)}(t_1) = 1 - h_u$ for $0 \leq h_u \leq 1$ and $u^{(k+1)}(t_1) = 0$ for $h_u \geq 1$. Determining the step size h_u is, thus, more involved than determining the step size $h_{\mathbf{B}}$.

To estimate h_u, the right derivative of (51) is considered

$$(\Pi \nabla_u J(\mathbf{B}^{(k+1)}, u^{(k)}))(t)$$

$$:= \lim_{h \downarrow 0} \frac{u^{(k)}(t) - u^{(k+1)}(t)}{h}$$

$$= \begin{cases} 0, & \text{when } u^{(k)}(t) = 0 \text{ and} \\ & \nabla_u J(\mathbf{B}^{(k+1)}, u^{(k)})(t) \geq 0 \\ \nabla_u J(\mathbf{B}^{(k+1)}, u^{(k)})(t), & \text{otherwise.} \end{cases}$$

$$(53)$$

At a certain time instant $t = t_1$, the right derivative of $(\Pi \nabla_u J)(t_1)$ is, thus, equal to the gradient $(\nabla_u J)(t_1)$ unless the current point lies on the border of the admissible set (i.e., $u^{(k)}(t_1) = 0$) and the update direction $-(\nabla_u J)(t_1)$ points outside the admissible set (i.e., $-(\nabla_u J)(t_1) \leq 0$). The second-order approximation of $J(\mathbf{B}^{(k+1)}, u^{(k+1)})$ for positive h_u near $h_u = 0$, thus, takes the form

$$J(\mathbf{B}^{(k+1)}, u^{(k+1)}) \approx J(\mathbf{B}^{(k+1)}, u^{(k)}) + h_u G_u^{(k)} + \frac{1}{2} h_u^2 H_u^{(k)}$$

$$(54)$$

with $G_u^{(k)} = -\langle \nabla_u J, \Pi \nabla_u J \rangle_u$ and

$$H_u^{(k)} = \left. \frac{\partial^2}{\partial h^2} (J(\mathbf{B}^{k+1}, u^{(k)} + h \Pi \nabla_u J)) \right|_{h=0} \quad (55)$$

where the gradient $\nabla_u J$ and the right derivative $\Pi \nabla_u J$ are evaluated in the point $(\mathbf{B}^{(k+1)}, u^{(k)})$. The computation of $H_u^{(k)}$ is further discussed in Appendix E. Similarly as for $h_\mathbf{B}$, the step size is estimated as $h_{u,\text{opt}} = -G_u^{(k)}/H_u^{(k)}$. If this step size does lead to an increase in the cost function, the step size is halved until a decrease in J is observed. This process will terminate eventually because the linear term in (54) will dominate the higher order terms and $G_u^{(k)} < 0$.

The criteria for convergence checked in line 6 of the algorithm are: 1) the relative change $\|\mathbf{B}^{(k+1)} - \mathbf{B}^{(k)}\|_B / \|\mathbf{B}^{(k)}\|_B$ is below `tol`; 2) the relative change $\|u^{(k+1)} - u^{(k)}\|_u / \|u^{(k)}\|_u$ is below `tol`; and 3) the relative decrease of the cost functional $(J(\mathbf{B}^{(k)}, u^{(k)}) - J(\mathbf{B}^{(k+1)}, u^{(k+1)}))/J(\mathbf{B}^{(k)}, u^{(k)})$ is below `tol`. Note that the same tolerance `tol` is used in all three conditions and that the norms $\|\cdot\|_B$ and $\|\cdot\|_u$ are induced by the inner products in (46) and (39), respectively.

Note that the shape of the actuation heat load $B(x, \zeta) = \mathbf{N}(x, \zeta)\mathbf{B}$ is not enforced to have unit L^1-norm during the update process. The normalization (15) can easily be achieved by rescaling $\mathbf{B}$ and $u(t)$ after the algorithm has terminated.

IV. RESULTS

Recall that the considered situation in Fig. 1 is symmetric in $x = 0$ so that only the domain $x \geq 0$ needs to be considered in the FE model. This infinite domain is truncated to $(x, \zeta) \in \Omega = [0, 4L] \times [-3vt_e, 3vt_e]$, which is chosen such that the temperature increase and deformation at the edges of the domain are negligible. The FE model uses linear quadrilateral elements and has 6360 nodes and 6188 elements. The mesh is shown in Fig. 2. A rectangular grid is used with a mesh size of 2 mm in the x-direction and a mesh size of 0.7 mm in the ζ-direction near the area where the heat load
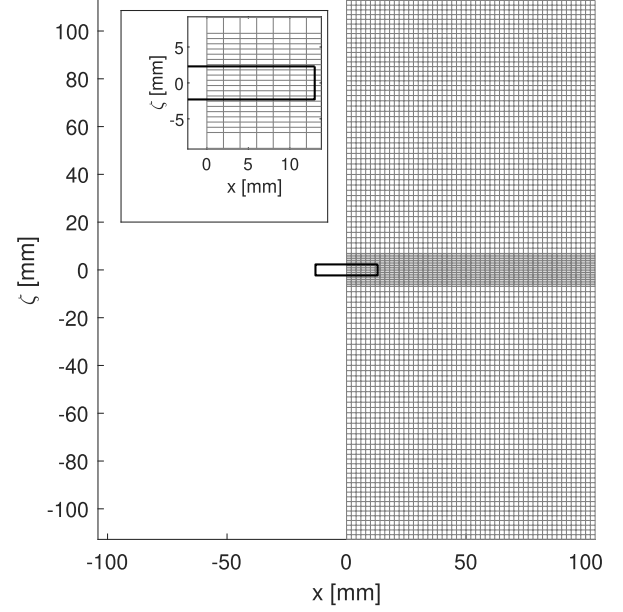

Fig. 2. Used mesh (gray lines) with the slit area in which the heat load $Q_{\exp}$ is applied (black rectangle).

TABLE I
PARAMETER VALUES

ρ	2329 [kg/m³]	v	0.276 [m/s]
c	705 [J/kg/K]	E	167 [GPa]
H	0.775 [mm]	ν	0.3 [-]
k	149 [W/K/m]	k_s	$1.21 \cdot 10^{12}$ [N/m³]
h_c	1500 [W/m²/K]	δ_{slit}	2 [nm]
$P_{\exp}$	3.2 [W]	δ_{slip}	3.67 [nm]
L	26 [mm]	p	3.5 [-]
W	4.6 [mm]	w	$2.1 \cdot 10^5$ [W²/m⁴]
t_e	0.136 [s]	`tol`	10^{-3} [-]

is applied and a mesh size of 2 mm outside this area. The time interval $[0, t_e]$ is discretized using $N = 200$ equidistant time points. The time discretization scheme of Apel and Flaig [31] is used, which leads to a discretely consistent computation of the gradients. Details can be found in Appendix F. Frequently, linear systems involving the stiffness matrix $\mathbf{SKS}^\top$ and the matrix $\mathbf{E}/\tau - \mathbf{A}/2$, where τ is the grid spacing of the time grid, need to be solved. To speed up this process, these systems are solved using a precomputed lower–upper (LU) factorization.

The used parameter values are given in Table I. Note that an actuation heat load will be designed that achieves a maximal deformation in the slit of $\delta_{\text{slit}} = 2$ nm, which is a reduction of a factor 2 compared with the maximal deformation of 4.1 nm that occurs without actuation. In addition, observe that $\delta_{\text{slip}} = 3.67$ nm is exceeded without actuation, which means actuation is needed to prevent slip. It should be noted that no admissible solution can be found when δ_{slit} or δ_{slip} are chosen too small. For example, for $\delta_{\text{slit}} = 0.5$ nm and $\delta_{\text{slip}} = 3.67$ nm, no admissible actuation heat load can be found.

A. Single-Shape Actuation Heat Load

Fig. 3 shows the design procedure described in the previous section. For ease of interpretation, the shape of the actuation heat load computed in the domain $x \geq 0$ is mirrored in $x = 0$ in Figs. 3(a)–(c), although the domain $x < 0$ is not considered
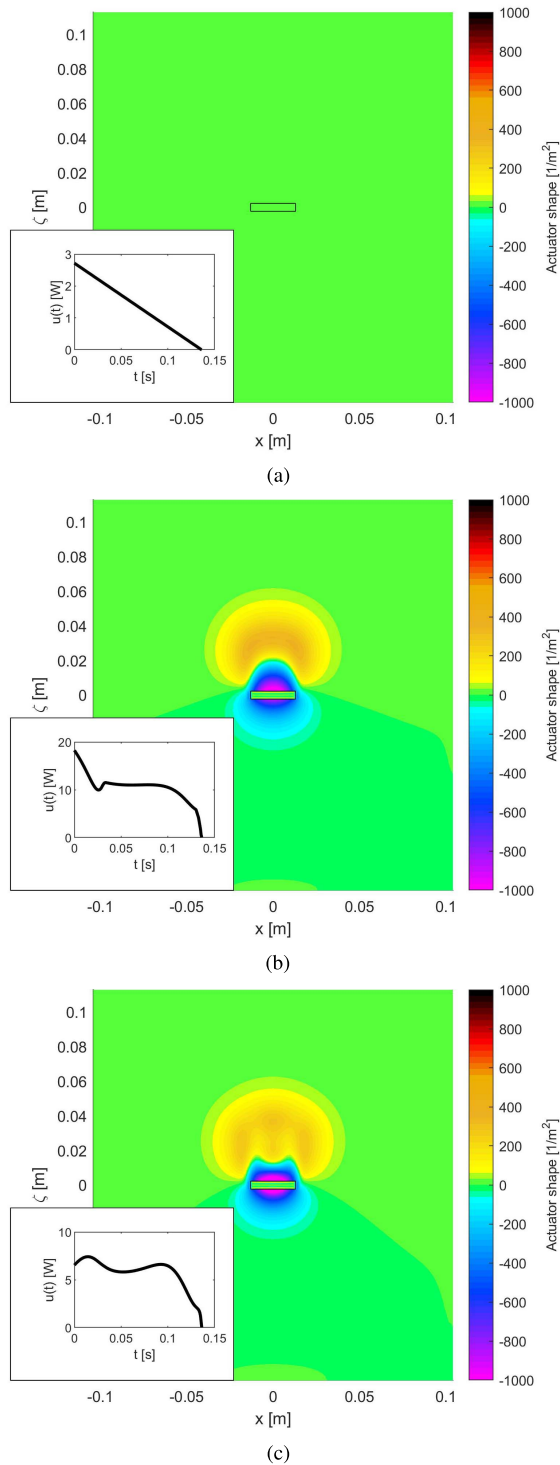
(a)

(b)

(c)

Fig. 3. Shapes $B(x,\zeta)$ and intensities $u(t)$ of the actuation heat load (a) that are used as initial guess in the optimization, (b) that are obtained after searching for an admissible solution, and (c) that are obtained after minimizing the actuation effort. The black rectangle indicates Ω_{slit}.

in the FE model. Starting from the initial guess in Fig. 3(a), where the shape of the actuation heat load $B(x,\zeta) = 0$, and the intensity is chosen rather arbitrarily as $u(t) = t_e - t$ [W], minimizing J_1 in (17) leads to the admissible design in Fig. 3(b). This design is then used to initialize the optimization procedure for $J_0 + wJ_2$, which leads to the design in Fig. 3(c). Note that the normalization $\|B\|_{L^1} = 1$

is applied in Fig. 3 (except for Fig. 3(a) where $B(x,\zeta) = 0$). Looking at the intensities $u(t)$ in Fig. 3(b) and (c) the second actuation step significantly reduces the applied heat. In particular, the value of $J_0 = 2389$ for the design in Fig. 3(b) is reduced to $J_0 = 717$ for the design in Fig. 3(c).

The designed shape of the optimal actuation heat load in Fig. 3(c) can be understood as follows. Note that having $Q_{\text{act}} = -Q_{\text{exp}}$ would result in zero deformations because no net heat load is applied. However, the constraint (13) excludes this solution. The cooling around the slit in Fig. 3(c) attempts to have some of this effect while respecting the constraint (13). A more surprising aspect of the shape in Fig. 3(c) may be the heating applied in the area where $\zeta > 0$. To understand this, recall that the expose load moves in the positive y-direction, which means that during scanning the heat applied by the exposing light is accumulating in the area where $\zeta < 0$. The thermal expansion due to this heating pushes the slit into the positive ζ-direction. The actuation heat load applied in the area where $\zeta > 0$ now creates thermal expansion in front of the slit, which pushes the slit back into the negative ζ-direction, thus reducing the total deformation in the slit.

Table II lists the values of $J_0 + wJ_2$, J_0, and

$$\|Q_{\text{act}}\|_{L^1} = \int_0^{t_e} \iint_{\mathbb{R}^2} |Q_{\text{act}}(x,\zeta,t)| \; dx \; d\zeta \; dt \qquad (56)$$

for the computed optimal actuation heat load found after minimizing $J_0 + wJ_2$. Note that $\|Q_{\text{act}}\|_{L^1}$ has the interpretation of the total applied actuation energy in [J]. The table also shows how many iterations (i.e., # updates of $\mathbf{B}$ + # updates of $u(t)$) and the computational time that are required to compute the optimal actuation heat load. As can be seen, finding an admissible solution, i.e., finding an actuation heat load for which $J_1 = 0$, requires only a few iterations, whereas the minimization of $J_0 + wJ_2$ requires more. Note that the number of iterations needed for the minimization of J_1 will increase as the constraint (10) becomes more stringent.

Fig. 4 shows the designed shapes $B(x,\zeta)$ and intensities $u(t)$ that are obtained when the initial intensity $u^{(0)}(t) = 20(t_e - t)$ and the initial shape $B^{(0)}(x,\zeta)$ is chosen as $-1/(8LW)$ for $(x,\zeta) \in [-3L/2, 3L/2] \times [-3W/2, 3W/2] \setminus \Omega_{\text{slit}}$ and zero otherwise. This initial guess is motivated by the idea that cooling around the slit is a quite effective method to reduce the deformation in the slit. The fact that the shapes and intensities of the actuation heat loads in Fig. 3(b) and (c) are different illustrates that there are indeed many admissible solutions that satisfy the constraints (10) and (13). The particular solution found after minimizing J_1, thus, strongly depends on the initial guess that is used. Note that the optimal designs in Fig. 3(c) are very similar, which suggests that the single-shape optimal actuation heat load may be unique. This is also indicated by the results in Table II.

The algorithm from Section III-C can be extended such that input constraints on the shape of the actuation heat load $B(x,\zeta)$ can be incorporated. The updates for $B(x,\zeta)$ are then also computed using the projected gradient method, similarly as the updates of $u(t)$ outlined in Section III-C. This enables the design of an actuation heat load that can only cool (i.e., $B(x,\zeta) \leq 0$) or heat (i.e., $B(x,\zeta) \geq 0$).
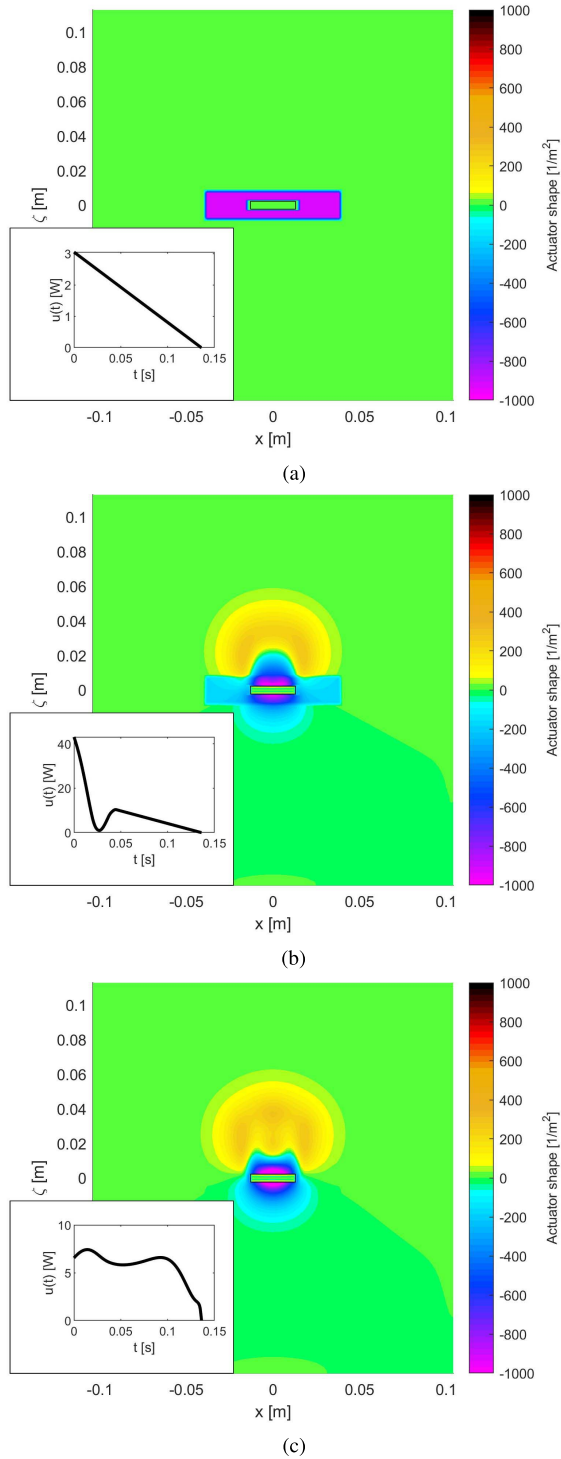
(a)

(b)

(c)

Fig. 4. Shapes $B(x,\zeta)$ and intensities $u(t)$ of the actuation heat load obtained starting from a different initial guess as in Fig. 3. (a) Initial guess. (b) Admissible solution. (c) Optimal solution. The black rectangle indicates Ω_{slit}.

The optimal designs computed with these additional constraints are shown in Figs. 5 and 6. The initial guess in Fig. 3(a) was also used as a starting point for these results. The actuation heat load in Fig. 5 seems to aim at counteracting the heat load induced by the expose light, i.e., it attempts to approximate the situation $Q_{\mathrm{act}} = -Q_{\mathrm{exp}}$ as well as possible. It is remarkable to see that the heating in Fig. 6 is applied all
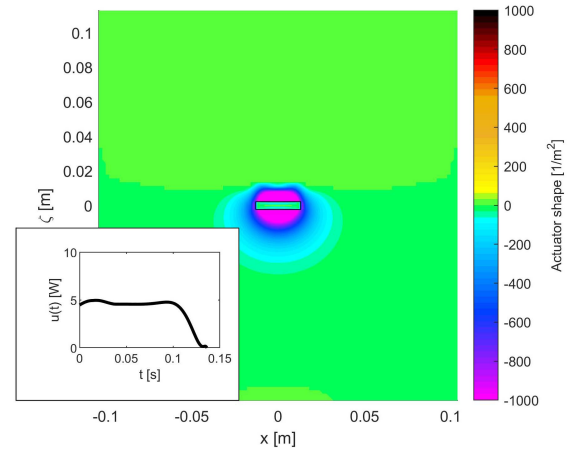

Fig. 5. Shape $B(x,\zeta)$ and intensity $u(t)$ computed with cooler constraints $B(x,\zeta) \leq 0$ and $u(t) \geq 0$.
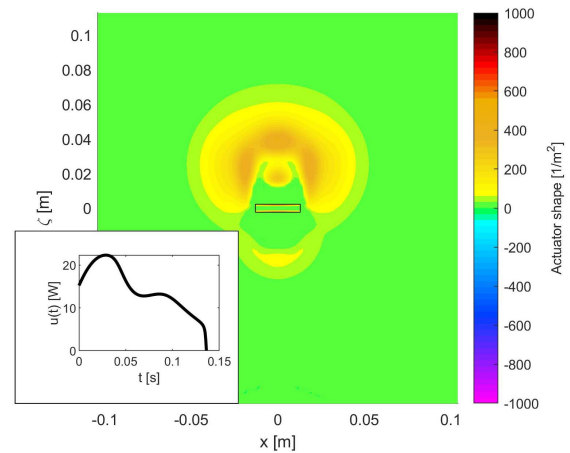

Fig. 6. Shape $B(x,\zeta)$ and intensity $u(t)$ computed with heater constraints $B(x,\zeta) \geq 0$ and $u(t) \geq 0$.

around the slit (in particular also for $\zeta < 0$). Such heating leads to thermal expansion of the material around the slit and compression of the material inside the slit and can thus counteract the thermal expansion due to the expose light inside the slit. Table II indicates that the minimal values of $J_0 + wJ_2$ and J_0 obtained with these additional constraints are indeed higher than the minimum value for the designs obtained without these constraints. Note that the actuation heat load with cooler constraints has a lower L^1-norm than the unconstrained actuation heat load. This reflects that the shape $B(x,\zeta)$ designed with cooler constraints is concentrated in a smaller area than the other designs which leads to a smaller L^1-norm, as was already observed in the works of Stadler [12] and Herzog *et al.* [13].

B. Comparison With Free-Shape Optimal Controls

The obtained single-shape actuation heat loads of the form (12) will be compared with a free-shape actuation heat load

$$Q_{\mathrm{act}}(x,\zeta,t) = U(x,\zeta,t). \tag{57}$$

Similar to the constraint (13) for the single-shape actuation heat load, it is required that no heat can be applied inside the

TABLE II

COMPUTATION OF THE SINGLE-SHAPE ACTUATION HEAT LOAD $Q_{\text{act}}(x, \zeta, t) = B(x, \zeta)u(t)$

	$J_0 + wJ_2$	J_0	$\|Q_{\text{act}}\|_{L^1}$	minimization of J_1		minimization of $J_0 + wJ_2$	
				# iterations	duration [s]	# iterations	duration [s]
Starting from initial guess 1 (see Fig. 3)	803	717	0.794	3	4	133	350
Starting from initial guess 2 (see Fig. 4)	803	717	0.793	4	7	203	541
Cooler constraints (see Fig. 5)	1167	1072	0.555	5	12	223	700
Heater constraints (see Fig. 6)	3198	2640	2	9	26	247	732

TABLE III

COMPUTATION OF FREE-SHAPE OPTIMAL CONTROL $Q_{\text{act}}(x, \zeta, t) = U(x, \zeta, t)$

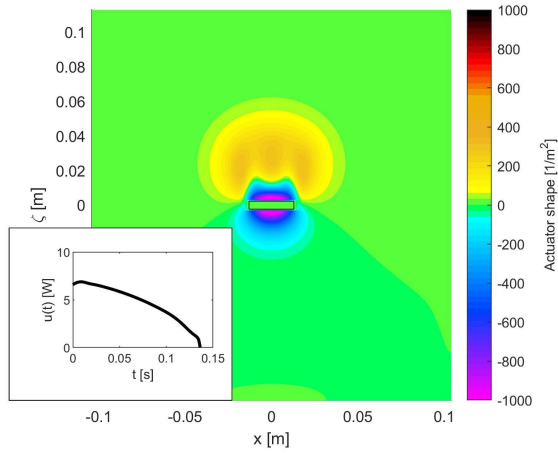|	$J_0 + wJ_2$	J_0	$\|Q_{\text{act}}\|_{L^1}$	minimization of J_1		minimization of $J_0 + wJ_2$	
				# iterations	duration [s]	# iterations	duration [s]
Starting from initial guess 1 (see Fig. 7)	703	602	0.687	4	9	57	159
Starting from initial guess 2 (see Fig. 8)	703	602	0.686	6	12	60	161
Cooler constraints (see Fig. 9)	1139	1028	0.555	7	25	100	375
Heater constraints (see Fig. 10)	2308	1838	1.610	9	40	120	442

Fig. 7. First POD mode of the actuation heat load $U(x, \zeta, t)$ computed from the initial guess $U^{(0)}(x, \zeta, t) = 0$.
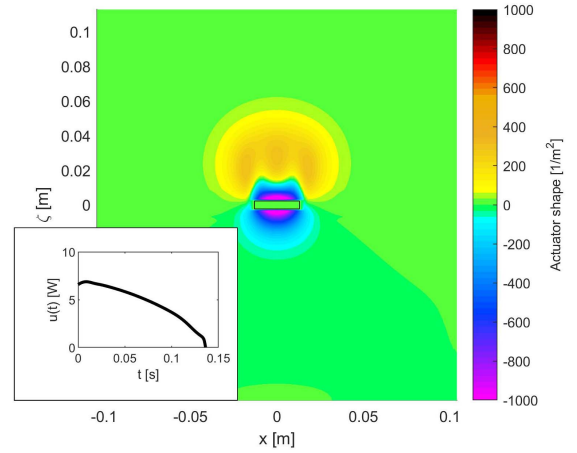

Fig. 8. First POD mode of the actuation heat load $U(x, \zeta, t)$ computed from the initial guess $U^{(0)}(x, \zeta, t) = B(x, \zeta)u(t)$ with $B(x, \zeta)$ and $u(t)$ as in Fig. 4(a).

area Ω_{slit} where the light source projects the pattern, that is,

$$U(x, \zeta, t) = 0, \quad \text{for } (x, \zeta) \in \Omega_{\text{slit}}, 0 \leq t \leq t_e. \quad (58)$$

The actuation heat load in (57) leads to an optimal control problem with a strictly convex cost functional J_0, a convex state constraint (10), and a convex input constraint (58). For this problem it can be shown that if the admissible set is nonempty, the solution to this problem is unique (this follows for example from [25, Proposition 10.6]). This problem is solved similarly as for the single-shape actuation heat load. First, the cost functional J_1 in (17) is minimized to find an admissible solution which is used as starting point for the minimization of $J_0 + wJ_2$. The main difference is that now the actuation heat load in (57) is used to compute the temperature field in (4) and cost functional J_0 in (11). The minimization is again implemented using the projected gradient method (see [27] or [30]) which enables the preservation of the constraint (58) throughout the iterations. The gradient is computed from (40) based on the adjoint state. The step size is controlled based on Hessian information, similar to the procedure for the single-shape actuation heat load described in Section III-C.

The free-shape optimal controls have been computed for the same four cases that have been considered for the single-shape

actuation heat load in Section IV-A. In particular, the controls have been computed starting from two different initial guesses $(U^{(0)}(x, \zeta, t) = 0$ and $U^{(0)}(x, \zeta, t) = B(x, \zeta)u(t)$ with $B(x, \zeta)$ and $u(t)$ as in Fig. 4(a)), with cooler constraints $(U(x, \zeta, t) \leq 0)$, and with heater constraints $(U(x, \zeta, t) \geq 0)$. Naturally, the obtained minimal values of $J_0 + wJ_2$ for the free-shape actuation heat load in Table III are lower than the values obtained for the single-shape actuation heat load in Table II.

To give an impression of the computed controls, a proper orthogonal decomposition (POD) (see [32]) of the optimal control $U(x, \zeta, t)$ is made (the POD was computed w.r.t. the $\langle \cdot \cdot \rangle_B$-inner product from (46)). The first POD modes $B_{\text{POD}}(x, \zeta)$ obtained for the four considered cases are shown in Figs. 7–10. The inserts in Figs. 7–10 also show $u_{\text{POD}}(t) = \langle U(x, \zeta, t), B_{\text{POD}}(x, \zeta) \rangle_B$. Since the optimal control $U(x, \zeta, t)$ is unique, the controls $U(x, \zeta, t)$ obtained for the first two cases should be the same. Indeed, the first POD modes in Figs. 7 and 8 cannot be distinguished. This demonstrates the accuracy of the used algorithm.

The obtained shapes of the actuation heat load in Figs. 7–10 are similar to the obtained shapes in Figs. 3(c), 4(c), 5, and 6, respectively, that were obtained by the optimization of the single shape actuation heat load. However, the intensities are
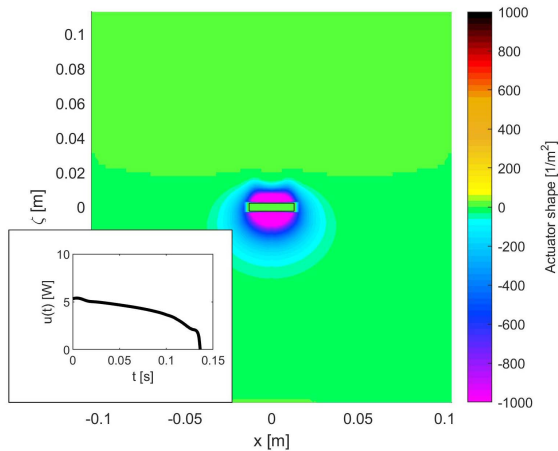
Fig. 9. First POD mode of the actuation heat load $U(x, \zeta, t)$ designed with cooler constraints $U(x, \zeta, t) \leq 0$.
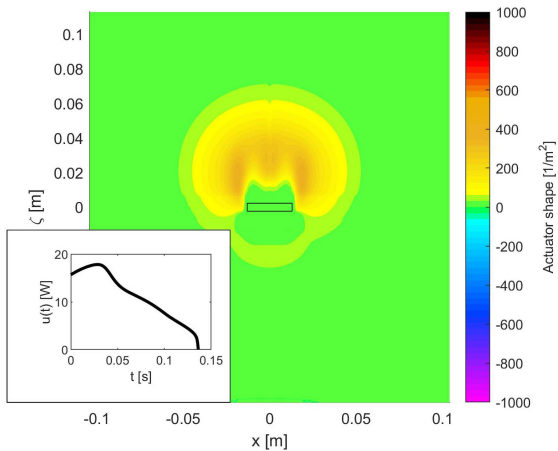

Fig. 10. First POD mode of the actuation heat load $U(x, \zeta, t)$ designed with heater constraints $U(x, \zeta, t) \geq 0$.

clearly distinct. It is remarkable that the intensities $u_{\mathrm{POD}}(t)$ never change sign, i.e., they satisfy the constraint (14), and that the spatial shapes in Figs. 9 and 10 also satisfy the input constraints $B(x, \zeta) \leq 0$ and $B(x, \zeta) \geq 0$, respectively. This is not guaranteed by the POD method.

Note that $B_{\mathrm{POD}}(x, \zeta) u_{\mathrm{POD}}(t)$ is the best least-squares approximation of $U(x, \zeta, t)$ that consists of a single shape. However, for w small enough, applying $Q_{\mathrm{act}}(x, \zeta, t) = B_{\mathrm{POD}}(x, \zeta) u_{\mathrm{POD}}(t)$ will always lead to a violation of the constraint (10). This can be seen as follows. For $w \to 0$, the control $U(x, \zeta, t)$ obtained by minimizing $J_0 + wJ_2$ approaches the control of smallest L^2-norm that satisfies the constraint (10). Since the L^2-norm of $B_{\mathrm{POD}}(x, \zeta) u_{\mathrm{POD}}(t)$ will always be smaller than the L^2-norm of $U(x, \zeta, t)$, $B_{\mathrm{POD}}(x, \zeta) u_{\mathrm{POD}}(t)$ will not satisfy (10) for w small enough.

The computed POD shapes offer a partial explanation for the observed uniqueness of the single-shape actuation heat load $B(x, \zeta) u(t)$ in Section IV-A. Inspection of the numerical results shows that the singular value corresponding to the first POD mode is more than three times larger than the second singular value for all four considered cases. This indicates that there is a single shape most important in $U(x, \zeta, t)$. It can,

therefore, be expected that the single-shape actuation heat load will have a shape similar to the dominant shape in $U(x, \zeta, t)$.

Table III also shows that the free-shape optimal controls $U(x, \zeta, t)$ are computed faster than the single-shape actuation heat loads. This suggests that the single-shape actuation heat load might be computed faster by using $B_{\mathrm{POD}}(x, \zeta) u_{\mathrm{POD}}(t)$ as a starting point for the optimization, but this not the case.

V. CONCLUSION AND DISCUSSIONS

An approach to compute the shape and intensity of the smallest actuation heat load that sufficiently mitigates wafer deformation has been proposed. The method consists of two steps. The first step is a search for an actuation heat load that sufficiently reduces the wafer deformations. In the second step, the actuation effort is minimized while keeping the wafer deformation sufficiently small. The method has been used to design an optimal single-shape actuation heat load in a 2-D wafer heating model for the scanning of a single field.

It has been shown that different initializations of the optimization algorithm converge to the same single-shape actuation heat load. This suggests that the optimal single-shape actuation heat load is unique. However, this claim has not been proven yet and is considered to be a topic of future research. This claim is supported by the computed free-shape actuation heat loads, of which the (typically unique) first POD mode resembles the obtained single-shape actuation heat loads.

The proposed method to design a single-shape actuation heat load has great flexibility and can be applied to many variations of the problem. For example, it has been demonstrated that additional input constraints requiring that the applied actuation heat load can only heat or cool can be included easily in the design. Some other recommendations for future research are the following. Here, an infinite domain $(x, y) \in \mathbb{R}^2$ was considered on which the effects of wafer edges are neglected. The formulation can be extended such that the effect of wafer edges can also be considered. Another important extension would be to consider a more realistic exposing pattern in which multiple fields are scanned in a meandering fashion.

A time-consuming step in the algorithm is solving the ordinary differential equations for the temperature field. The cost of these computations can potentially be reduced by applying model order reduction (MOR) techniques. Especially because the actuation heat load consists of a single shape, such techniques show great potential to reduce the computational cost. In particular, such MOR techniques might enable to compute the single-shape actuation heat load faster than the free-shape optimal control and might enable the application of the method to much larger FE models, which are typically needed for more complex expose patterns and 3-D wafer and wafer clamp models.

Another interesting problem is the design of an actuation heat load that consists of multiple shapes, e.g., an actuation heat load of of the form

$$Q_{\mathrm{act}}(x, \zeta, t) = B_1(x, \zeta) u_1(t) + B_2(x, \zeta) u_2(t). \quad (59)$$

However, such an extension comes with several additional problems. For example, it is no longer straightforward

how the representation of the applied actuation heat load should be made unique. To see this, denote $B(x, \zeta) = [B_1(x, \zeta), B_2(x, \zeta)]$ and $u(t) = [u_1(t), u_2(t)]^\top$ so that $Q_{\text{act}} = B(x, \zeta)u(t)$. Now it is easy to see that the representations (B, u) and $(BQ, Q^{-1}u)$ lead to the same actuation heat load for any invertible 2×2 matrix Q. It is, therefore, not trivial to find conditions that make the representation of a multishape actuation heat load unique.

APPENDIX A
PARAMETERS IN THE BARRIER FUNCTION J_2

In this section, feasible values for the power p and the weight w appearing in the cost functional $J_0 + wJ_2$ (with J_0 as in (11) and J_2 as in (19)) will be determined.

The power p should be chosen such that J_2 increases "fast enough" near the edge of the admissible set. In particular, it is desired that the integrand is approximately zero at points (x, ζ, t) for which there is enough margin in (10) and that the integrand is large for points (x, ζ, t) where the margin in (10) is small. This idea can be translated into the requirement that the integrand in (19) for points (x, ζ, t) at which

$$d_{\max}(x, \zeta) - \sqrt{d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t)} < \varepsilon_1 d_{\max}(x, \zeta) \quad (60)$$

is at least $M > 1$ times larger than the value of the integrand in (19) for points (x, ζ, t) at which

$$d_{\max}(x, \zeta) - \sqrt{d_x^2(x, \zeta, t) + d_\zeta^2(x, \zeta, t)} > \varepsilon_2 d_{\max}(x, \zeta) \quad (61)$$

where $0 < \varepsilon_1 < \varepsilon_2 < 1$. Parameters ε_1 and ε_2 are used to quantify a "small" and "large enough" margin relative to maximally allowed displacement at that point. Note that equations (60) and (61) are equivalent to

$$\frac{d_x^2 + d_y^2}{d_{\max}^2} > (1 - \varepsilon_1)^2, \quad \frac{d_x^2 + d_y^2}{d_{\max}^2} < (1 - \varepsilon_2)^2 \quad (62)$$

respectively, where the dependence on x, ζ, and t has been dropped. Using that the integrand in (19) is monotonically increasing in $d_x^2 + d_\zeta^2$, the requirement on p translates to

$$(1 - (1 - \varepsilon_1)^2)^{-p} - 1 > M((1 - (1 - \varepsilon_2)^2)^{-p} - 1). \quad (63)$$

Using that $M > 1$, it is easy to see that this equation is satisfied when

$$(1 - (1 - \varepsilon_1)^2)^{-p} > M(1 - (1 - \varepsilon_2)^2)^{-p}. \quad (64)$$

Some straightforward algebra now shows that this implies that

$$p > \frac{\log(M)}{\log(2\varepsilon_2 - \varepsilon_2^2) - \log(2\varepsilon_1 - \varepsilon_1^2)}. \quad (65)$$

For example, in Section IV, the power p is determined based on the requirement that points for which the margin in (10) is below 0.1 nm contribute 1000 times more to the value of J_2 than points for which there is more than 1-nm margin in (10). Using that $\min_{x, \zeta} d_{\max}(x, \zeta) = 2$ nm, this translates to $M = 1000$, $\varepsilon_1 = 0.05$, and $\varepsilon_2 = 0.5$, for which (65) becomes $p > 3.3858$. Indeed, this lower bound is slightly below the value of $p = 3.5$ used in the examples in Section IV.

The weight w should be chosen such that the barrier function J_2 starts to dominate the value of cost functional $J_0 + wJ_2$ when the margin in (10) becomes "small" (as in (60)). To find the value of w that realizes this, typical magnitudes of the integrands in (11) and (19) are needed. It is reasonable to expect that the applied actuator heat load Q_{act} will be of similar magnitude as the heat load induced by the projecting light Q_{exp}. Recall that Q_{exp} has a power P_{exp} that is uniformly distributed over the area $L \times W$ (see Fig. 1) so that Q_{exp} has a typical magnitude of P_{exp}/LW. The magnitude of the integrand in (11) is, thus, estimated as $(P_{\text{exp}}/LW)^2$. When the margin in (10) becomes "small," as in (60), the typical magnitude of the integrand in (19) is $(1 - (1 - \varepsilon_1)^2)^{-p}$. The weight w should, thus, be chosen such that

$$\left(\frac{P_{\text{exp}}}{LW}\right)^2 \approx w(1 - (1 - \varepsilon_1)^2)^{-p}. \quad (66)$$

For the considered parameter values this yields $w = 2.1 \cdot 10^5$, which is the value of w used in Section IV.

APPENDIX B
FINITE-ELEMENT MATRICES

The system matrix $\mathbf{A}$ in (23) is given by

$$\mathbf{A} = \iint_\Omega \left(\rho c H v \mathbf{N}^\top \frac{\partial \mathbf{N}}{\partial \zeta} - h_c \mathbf{N}^\top \mathbf{N} \right.$$
$$\left. - kH \left(\frac{\partial \mathbf{N}^\top}{\partial x} \frac{\partial \mathbf{N}}{\partial x} + \frac{\partial \mathbf{N}^\top}{\partial \zeta} \frac{\partial \mathbf{N}}{\partial \zeta} \right) \right) dx \, d\zeta. \quad (67)$$

The stiffness matrix $\mathbf{K}$ and the thermal load matrix $\mathbf{L}$ in (28) are given by

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{xx} & \mathbf{K}_{x\zeta} \\ \mathbf{K}_{\zeta x} & \mathbf{K}_{\zeta\zeta} \end{bmatrix}, \quad \mathbf{L} = \begin{bmatrix} \mathbf{L}_x \\ \mathbf{L}_\zeta \end{bmatrix} \quad (68)$$

with

$$\mathbf{K}_{xx} = \iint_\Omega \left(\frac{EH}{1 - \nu^2} \frac{\partial \mathbf{N}^\top}{\partial x} \frac{\partial \mathbf{N}}{\partial x} + \frac{EH}{2(1 + \nu)} \frac{\partial \mathbf{N}^\top}{\partial \zeta} \frac{\partial \mathbf{N}}{\partial \zeta} \right.$$
$$\left. + k_s \mathbf{N}^\top \mathbf{N} \right) dx \, d\zeta \quad (69)$$

$$\mathbf{K}_{x\zeta} = \mathbf{K}_{\zeta x}^\top = \iint_\Omega \frac{EH}{1 - \nu^2} \left(\nu \frac{\partial \mathbf{N}^\top}{\partial x} \frac{\partial \mathbf{N}}{\partial \zeta} \frac{\partial \mathbf{N}^\top}{\partial \zeta} \frac{\partial \mathbf{N}}{\partial x} \right)$$
$$\times dx \, d\zeta \quad (70)$$

$$\mathbf{K}_{\zeta\zeta} = \iint_\Omega \left(\frac{EH}{1 - \nu^2} \frac{\partial \mathbf{N}^\top}{\partial \zeta} \frac{\partial \mathbf{N}}{\partial \zeta} + \frac{EH}{2(1 + \nu)} \frac{\partial \mathbf{N}^\top}{\partial x} \frac{\partial \mathbf{N}}{\partial x} \right.$$
$$\left. + k_s \mathbf{N}^\top \mathbf{N} \right) dx \, d\zeta \quad (71)$$

$$\mathbf{L}_x = \frac{\alpha E H}{1 - \nu} \iint_\Omega \frac{\partial \mathbf{N}^\top}{\partial x} \mathbf{N} \, dx \, d\zeta \quad (72)$$

$$\mathbf{L}_\zeta = \frac{\alpha E H}{1 - \nu} \iint_\Omega \frac{\partial \mathbf{N}^\top}{\partial \zeta} \mathbf{N} \, dx \, d\zeta. \quad (73)$$

Note that $\mathbf{K}$ is invertible for $k_s > 0$.

APPENDIX C
EXPLICIT EXPRESSIONS FOR THE
DERIVATIVES OF f AND g

The cost functional $J = J_1$ with J_1 as in (35) can be written in the form (42) by setting

$$f(\mathbf{d}(t)) = \mathbf{1}^\top \mathbf{E}_0 \left[\mathbf{d}_x^2(t) + \mathbf{d}_\zeta^2(t) - \mathbf{d}_{\max}^2 \right]^+ \tag{74}$$

and $g(\mathbf{B}, u(t)) = 0$. The derivative of f w.r.t. $\mathbf{d} = [\mathbf{d}_x^\top, \mathbf{d}_\zeta^\top]^\top$ is

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{d}(t)) = 2 \begin{bmatrix} \mathbf{d}_x(t) \circ \mathbf{1}^+(t) \circ \mathbf{E}_0 \mathbf{1} \\ \mathbf{d}_\zeta(t) \circ \mathbf{1}^+(t) \circ \mathbf{E}_0 \mathbf{1} \end{bmatrix}^\top \tag{75}$$

where $\circ$ denotes the componentwise (Hadamard) product of vectors and the ith component of $\mathbf{1}^+(t)$ is

$$\mathbf{1}_i^+(t) = \begin{cases} 1, & \text{when } \mathbf{d}_{x,i}^2(t) + \mathbf{d}_{\zeta,i}^2(t) - \mathbf{d}_{\max,i}^2 \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{76}$$

where $\mathbf{d}_{x,i}(t)$, $\mathbf{d}_{\zeta,i}(t)$, and $\mathbf{d}_{\max,i}(t)$ are the ith components of $\mathbf{d}_x(t)$, $\mathbf{d}_\zeta(t)$, and $\mathbf{d}_{\max}(t)$, respectively.

Differentiating (75) again to $\mathbf{d}$ yields the Hessian

$$\frac{\partial^2 f}{\partial \mathbf{d}^2}(\mathbf{d}(t))$$
$$= 2 \begin{bmatrix} \text{diag}(\mathbf{1}^+(t) \circ \mathbf{E}_0 \mathbf{1}) & 0 \\ 0 & \text{diag}(\mathbf{1}^+(t) \circ \mathbf{E}_0 \mathbf{1}) \end{bmatrix} \tag{77}$$

where $\text{diag}(\mathbf{v})$ denotes a matrix with the components of the vector $\mathbf{v}$ on the diagonal.

The cost functional $J = J_0 + w J_2$ with J_0 and J_2 as in (33) and (36) can be written in the form (42) by setting

$$f(\mathbf{d}(t)) = w \mathbf{1}^\top \mathbf{E}_0 \left[\left[\frac{\mathbf{d}_{\max}^2}{\mathbf{d}_{\max}^2 - \mathbf{d}_x^2(t) - \mathbf{d}_\zeta^2(t)} \right]^p - \mathbf{1} \right] \tag{78}$$

$$g(\mathbf{B}, u(t)) = u^\top(t) \mathbf{B}^\top \mathbf{E}_0 \mathbf{B} u(t). \tag{79}$$

The derivative of f w.r.t. $\mathbf{d} = [\mathbf{d}_x^\top, \mathbf{d}_\zeta^\top]^\top$ is

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{d}(t)) = 2w \begin{bmatrix} \mathbf{d}_x(t) \circ \mathbf{a}(t) \circ \mathbf{E}_0 \mathbf{1} \\ \mathbf{d}_\zeta(t) \circ \mathbf{a}(t) \circ \mathbf{E}_0 \mathbf{1} \end{bmatrix}^\top \tag{80}$$

with

$$\mathbf{a}(t) = p \left(\frac{\mathbf{d}_{\max}^2}{\mathbf{d}_{\max}^2 - \mathbf{d}_x^2(t) - \mathbf{d}_\zeta^2(t)} \right)^{p+1} \circ \frac{1}{\mathbf{d}_{\max}^2} \tag{81}$$

where $\circ$ again denotes the componentwise Hadamard product of vectors and the operations $\cdot / \cdot$, and $(\cdot)^2$ and $(\cdot)^p$ are again applied componentwise.

To obtain the Hessian of f note that

$$\frac{\partial^2 f}{\partial \mathbf{d}^2}(\mathbf{d}(t)) = w \begin{bmatrix} \dfrac{\partial^2 f}{\partial \mathbf{d}_x^2}(\mathbf{d}(t)) & \dfrac{\partial^2 f}{\partial \mathbf{d}_x \partial \mathbf{d}_\zeta}(\mathbf{d}(t)) \\ \dfrac{\partial^2 f}{\partial \mathbf{d}_x \partial \mathbf{d}_\zeta}(\mathbf{d}(t)) & \dfrac{\partial^2 f}{\partial \mathbf{d}_\zeta^2}(\mathbf{d}(t)) \end{bmatrix}. \tag{82}$$

By differentiating the top part of (80) w.r.t. $\mathbf{d}_x$, we see $\partial^2 J_2 / \partial \mathbf{d}_x^2$ is a diagonal matrix with diagonal equal to

$$2 \left(\mathbf{a}(t) + 2(p+1)\mathbf{a}(t) \circ \frac{\mathbf{d}_x^2(t)}{\mathbf{b}(t)} \right) \circ \mathbf{E}_0 \mathbf{1} \tag{83}$$

where $\mathbf{b}(t) = \mathbf{d}_{\max}^2 - \mathbf{d}_x^2(t) - \mathbf{d}_\zeta^2(t)$. Similarly, the diagonal entries of $\partial^2 g / \partial \mathbf{d}_x \partial \mathbf{d}_\zeta$ and $\partial^2 g / \partial \mathbf{d}_\zeta^2$ are

$$2 \left(2(p+1)\mathbf{a}(t) \circ \frac{\mathbf{d}_x(t) \circ \mathbf{d}_\zeta(t)}{\mathbf{b}(t)} \right) \circ \mathbf{E}_0 \mathbf{1} \tag{84}$$

$$2 \left(\mathbf{a}(t) + 2(p+1)\mathbf{a}(t) \circ \frac{\mathbf{d}_\zeta^2(t)}{\mathbf{b}(t)} \right) \circ \mathbf{E}_0 \mathbf{1}. \tag{85}$$

The derivatives and Hessians of g in (79) evaluated in the point $(\mathbf{B}, u(t))$ are

$$\frac{\partial g}{\partial \mathbf{B}} = 2 \, u^2(t) \mathbf{B}^\top \mathbf{E}_0, \quad \frac{\partial^2 g}{\partial \mathbf{B}^2} = 2 \, u^2(t) \mathbf{E}_0 \tag{86}$$

$$\frac{\partial g}{\partial u} = 2 \, u(t) \mathbf{B}^\top \mathbf{E}_0 \mathbf{B}, \quad \frac{\partial^2 g}{\partial u^2} = 2 \mathbf{B}^\top \mathbf{E}_0 \mathbf{B}. \tag{87}$$

APPENDIX D
GRADIENT W.R.T. $\mathbf{B}$

To derive the formula (47) for the gradient $\nabla_{\mathbf{B}} J(\mathbf{B}, u)$ of the cost functional J in (37), consider a perturbation $\tilde{\mathbf{B}}$ of $\mathbf{B}$. From (23), it follows that changing $\mathbf{B}$ to $\mathbf{B} + \tilde{\mathbf{B}}$ changes the state $\boldsymbol{\theta}(t)$ to $\boldsymbol{\theta}(t) + \tilde{\boldsymbol{\theta}}(t)$, where $\tilde{\boldsymbol{\theta}}(t)$ is the solution of

$$\mathbf{E}\dot{\tilde{\boldsymbol{\theta}}}(t) = \mathbf{A}\tilde{\boldsymbol{\theta}}(t) + \mathbf{E}_{\mathbf{B}}\tilde{\mathbf{B}}u(t), \quad \tilde{\boldsymbol{\theta}}(0) = \mathbf{0}. \tag{88}$$

Using the definition of the gradient (45) and the form of the cost function (37), it follows that

$$\langle \nabla_{\mathbf{B}} J(\mathbf{B}, u), \tilde{\mathbf{B}} \rangle_B$$
$$= \lim_{\varepsilon \to 0} \frac{J(\mathbf{B} + \varepsilon \tilde{\mathbf{B}}, u) - J(\mathbf{B}, u)}{\varepsilon}$$
$$= \lim_{\varepsilon \to 0} \int_0^{t_e} \frac{\bar{f}(\boldsymbol{\theta}(t) + \varepsilon \tilde{\boldsymbol{\theta}}(t)) - \bar{f}(\boldsymbol{\theta}(t))}{\varepsilon} \, dt$$
$$+ \int_0^{t_e} \frac{g(\mathbf{B} + \varepsilon \tilde{\mathbf{B}}, u(t)) - g(\mathbf{B}, u(t))}{\varepsilon} \, dt$$
$$= \int_0^{t_e} \left(\frac{\partial \bar{f}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}(t))\tilde{\boldsymbol{\theta}}(t) + \frac{\partial g}{\partial \mathbf{B}}(\mathbf{B}, u(t))\tilde{\mathbf{B}} \right) dt. \tag{89}$$

With $\boldsymbol{\varphi}(t)$ and the adjoint state as in (41), it follows that

$$0 = \boldsymbol{\varphi}^\top(t)\mathbf{E}\tilde{\boldsymbol{\theta}}(t) \Big|_{t=0}^{t_e}$$
$$= \int_0^{t_e} (\dot{\boldsymbol{\varphi}}^\top(t)\mathbf{E}\tilde{\boldsymbol{\theta}}(t) + \boldsymbol{\varphi}^\top(t)\mathbf{E}\dot{\tilde{\boldsymbol{\theta}}}(t)) \, dt$$
$$= -\int_0^{t_e} \frac{\partial \bar{f}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}(t))\tilde{\boldsymbol{\theta}}(t) \, dt + \int_0^{t_e} \boldsymbol{\varphi}^\top(t)\mathbf{E}_{\mathbf{B}}\tilde{\mathbf{B}}u(t) \, dt \tag{90}$$

where the last identity follows after substitution of (41) and (88) for $\mathbf{E}^\top \dot{\boldsymbol{\varphi}}(t)$ and $\mathbf{E}\dot{\tilde{\boldsymbol{\theta}}}(t)$, respectively. Using (90) to replace the first term in (89), it follows

$$\langle \nabla_{\mathbf{B}} J(\mathbf{B}, u), \tilde{\mathbf{B}} \rangle_B$$
$$= \int_0^{t_e} \left(\boldsymbol{\varphi}^\top(t)\mathbf{E}_{\mathbf{B}}\tilde{\mathbf{B}}u(t) + \frac{\partial g}{\partial \mathbf{B}}(\mathbf{B}, u(t))\tilde{\mathbf{B}} \right) dt$$
$$= \left\langle \mathbf{E}_0^{-1} \int_0^{t_e} \left[\mathbf{E}_{\mathbf{B}}\boldsymbol{\varphi}(t)u(t) + \left[\frac{\partial g}{\partial \mathbf{B}}(\mathbf{B}, u(t)) \right]^\top \right] dt, \tilde{\mathbf{B}} \right\rangle_B \tag{91}$$

where the last identity follows because $u(t)$ is scalar and from the definition of the inner product $\langle \cdot \cdot \rangle_B$ in (46).

Since (91) holds for all perturbations $\tilde{\mathbf{B}}$, (47) follows.

APPENDIX E
COMPUTATION OF THE HESSIANS

To compute $H_{\mathbf{B}}$, the second derivative of the function $h \mapsto J(\mathbf{B} + h\tilde{\mathbf{B}}, u)$ is computed as

$$
\begin{aligned}
H_{\mathbf{B}} &= \frac{\partial^2}{\partial h^2}(J(\mathbf{B} + h\tilde{\mathbf{B}}, u))\bigg|_{h=0} \\
&= \frac{\partial^2}{\partial h^2} \int_0^{t_e} (\bar{f}(\boldsymbol{\theta}(t) + h\tilde{\boldsymbol{\theta}}(t)) + g(\mathbf{B} + h\tilde{\mathbf{B}}, u(t)))\ dt\bigg|_{h=0} \\
&= \int_0^{t_e} \left(\tilde{\boldsymbol{\theta}}^\top(t) \frac{\partial^2 \bar{f}}{\partial \boldsymbol{\theta}^2}(\boldsymbol{\theta}(t))\tilde{\boldsymbol{\theta}}(t) + \tilde{\mathbf{B}}^\top \frac{\partial^2 g}{\partial \mathbf{B}^2}(\mathbf{B}, u(t))\tilde{\mathbf{B}} \right) dt
\end{aligned}
$$
(92)

where $\tilde{\boldsymbol{\theta}}(t)$ is the solution of (88), i.e., $\tilde{\boldsymbol{\theta}}(t)$ is the change in the state due to the perturbation $\tilde{\mathbf{B}}$. The Hessian $H_{\mathbf{B}}$ in the point $(\mathbf{B}_0, u_0)$ for the direction $\tilde{\mathbf{B}}$ can, thus, be computed in the following steps: 1) compute $\boldsymbol{\theta}(t)$ as the solution of (23) with $\mathbf{B} = \mathbf{B}_0$ and $u(t) = u_0(t)$; 2) compute $\boldsymbol{\theta}_0(t)$ from (88) with $u(t) = u_0(t)$; and 3) evaluate the integral in (92) with $\mathbf{B} = \mathbf{B}_0$, $u(t) = u_0(t)$, $\boldsymbol{\theta}(t) = \boldsymbol{\theta}_0(t)$, and $\tilde{\boldsymbol{\theta}}(t) = \tilde{\boldsymbol{\theta}}_0(t)$.

To compute $H_{\mathbf{B}}$ for a cost functional of the form (42), observe that it follows from (43) that

$$
\frac{\partial^2 \bar{f}}{\partial \boldsymbol{\theta}^2}(\boldsymbol{\theta}(t)) = \mathbf{M}^\top \frac{\partial^2 f}{\partial \mathbf{d}^2}(\boldsymbol{\theta}(t))\mathbf{M}
$$
(93)

where $\mathbf{M} = \mathbf{S}^\top \left(\mathbf{S}\mathbf{K}\mathbf{S}^\top\right)^{-1} \mathbf{S}\mathbf{L}$. Explicit expressions for $\partial^2 f/\partial \mathbf{d}^2$ are given in Appendix C.

The Hessian of a cost function J of the form (37) in the direction $\hat{u}(t)$ is

$$
\begin{aligned}
H_u &= \frac{\partial^2}{\partial h^2}(J(\mathbf{B}, u + h\hat{u}))\bigg|_{h=0} \\
&= \int_0^{t_e} \left(\hat{\boldsymbol{\theta}}^\top(t) \frac{\partial^2 \bar{f}}{\partial \boldsymbol{\theta}^2}(\boldsymbol{\theta}(t))\hat{\boldsymbol{\theta}}(t) + \hat{u}^2(t) \frac{\partial^2 g}{\partial u^2}(\mathbf{B}, u(t)) \right) dt
\end{aligned}
$$
(94)

where $\hat{\boldsymbol{\theta}}(t)$ is the solution of

$$
\mathbf{E}\dot{\hat{\boldsymbol{\theta}}}(t) = \mathbf{A}\hat{\boldsymbol{\theta}}(t) + \mathbf{E}_{\mathbf{B}}\mathbf{B}\hat{u}(t), \quad \hat{\boldsymbol{\theta}}(0) = \mathbf{0}.
$$
(95)

The procedure to compute H_u is, thus, similar to the procedure to compute $H_{\mathbf{B}}$ discussed earlier.

APPENDIX F
TIME DISCRETIZATION

This appendix summarizes the time discretization of the cost functional (37), state equation (23), adjoint state equation (41), and gradients (40) and (47) following the method in [31] that leads to discretely consistent gradients. We consider a uniform grid of $N + 1$ grid points $t_k = k\tau$ for $k = 0, 1, \ldots, N$ where $\tau = t_e/N$. The state equation (23) is discretized with the Crank–Nicolson method, which leads to

$$
\mathbf{E}\frac{\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_k}{\tau} = \mathbf{A}\frac{\boldsymbol{\theta}_{k+1} + \boldsymbol{\theta}_k}{2} + \mathbf{B}_{\exp}u_{\exp,k+1/2}
$$
$$
+ \mathbf{E}_{\mathbf{B}}\mathbf{B}u_{k+1/2}, \quad \boldsymbol{\theta}_0 = \mathbf{0} \quad (96)
$$

where $\boldsymbol{\theta}_k = \boldsymbol{\theta}(t_k)$, $u_{\exp,k+1/2} = (u_{\exp}(t_k) + u_{\exp}(t_{k+1}))/2$, and $u_{k+1/2} = (u(t_k) + u(t_{k+1}))/2$. The cost functional (37) is discretized using the trapezoid rule for the state-dependent part and the midpoint rule for the part dependent on $\mathbf{B}$ and $u(t)$

$$
J^\tau = \tau \sum_{k=0}^{N-1} \left[\frac{\bar{f}(\boldsymbol{\theta}_{k+1}) + \bar{f}(\boldsymbol{\theta}_k)}{2} + g(\mathbf{B}, u_{k+1/2}) \right].
$$
(97)

The discretized adjoint state variables $\boldsymbol{\varphi}_{k+1/2}$ ($k = 0, 1, \ldots, N-1$) are computed starting from the solution of

$$
\mathbf{E}^\top \frac{\boldsymbol{\varphi}_{N-1/2}}{\tau} = \mathbf{A}^\top \frac{\boldsymbol{\varphi}_{N-1/2}}{2} + \frac{1}{2}\left(\frac{\partial \bar{f}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_N)\right)^\top
$$
(98)

after which $\boldsymbol{\varphi}_{k-1/2}$ ($k = N-1, \ldots, 2, 1$) can be solved from

$$
\mathbf{E}^\top \frac{\boldsymbol{\varphi}_{k-1/2} - \boldsymbol{\varphi}_{k+1/2}}{\tau}
$$
$$
= \mathbf{A}^\top \frac{\boldsymbol{\varphi}_{k-1/2} + \boldsymbol{\varphi}_{k+1/2}}{2} + \left(\frac{\partial \bar{f}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_k)\right)^\top.
$$
(99)

The gradients of J^τ w.r.t. to $u_{k+1/2}$ and $\mathbf{B}$ w.r.t. the inner products $\tau \sum_{k=0}^{N-1} u_{k+1/2}^{(1)} u_{k+1/2}^{(2)}$ and $\mathbf{B}_1^\top \mathbf{E}_0 \mathbf{B}_2$ are now

$$
\begin{aligned}
&(\nabla_u J^\tau)_{k+1/2} \\
&= \mathbf{B}^\top \mathbf{E}_{\mathbf{B}}\boldsymbol{\varphi}_{k+1/2} + \frac{\partial g}{\partial u}(\mathbf{B}, u_{k+1/2})
\end{aligned}
$$
(100)

$$
\begin{aligned}
&\nabla_{\mathbf{B}} J^\tau \\
&= \tau \mathbf{E}_0^{-1} \sum_{k=0}^{N-1} \left[\mathbf{E}_{\mathbf{B}}\boldsymbol{\varphi}_{k+1/2}u_{k+1/2} + \frac{\partial g}{\partial \mathbf{B}}(\mathbf{B}, u_{k+1/2}) \right].
\end{aligned}
$$
(101)

REFERENCES

[1] D. W. M. Veldman, R. H. B. Fey, H. J. Zwart, M. M. J. van de Wal, J. D. B. J. van den Boom, and H. Nijmeijer, "Semi-analytic approximation of the temperature field resulting from moving heat loads," *Int. J. Heat Mass Transf.*, vol. 122, pp. 128–137, Jul. 2018.

[2] L. Subramany *et al.*, "Analysis of wafer heating in 14 nm DUV layers," *Proc. SPIE*, vol. 9778, Apr. 2016, Art. no. 97780U.

[3] N. Aung *et al.*, "Overlay control for 7 nm technology node and beyond," *Proc. SPIE*, vol. 10587, Mar. 2018, Art. no. 105870A.

[4] D. van den Hurk, S. Weiland, and K. van Berkel, "Modeling and localized feedforward control of thermal deformations induced by a moving heat load," in *Proc. SICE Int. Symp. Control Syst. (SICE ISCS)*, Mar. 2018, pp. 171–178.

[5] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, 2nd ed. Chichester, U.K.: Wiley, 2007.

[6] M. Van de Wal and B. de Jager, "A review of methods for input/output selection," *Automatica*, vol. 37, no. 4, pp. 487–510, Apr. 2001.

[7] J. A. Taylor, N. Luangsomboon, and D. Fooladivanda, "Allocating sensors and actuators via optimal estimation and control," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 3, pp. 1060–1067, Jun. 2016.

[8] Y. Lou and P. D. Christofides, "Optimal actuator/sensor placement for nonlinear control of the Kuramoto-Sivashinsky equation," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 5, pp. 737–745, Sep. 2003.

[9] F. Al-Sulaiman and S. Zaman, "Actuator placement in lumped parameter systems subjected to disturbance," *Comput. Struct.*, vol. 52, no. 1, pp. 41–47, Jul. 1994.

[10] Y. Cao, D. Biss, and J. Perkins, "Assessment of input-output controllability in the presence of control constraints," *Comput. Chem. Eng.*, vol. 20, no. 4, pp. 337–346, Apr. 1996.

[11] D. S. Naidu, *Optimal Control Systems*. Boca Raton, FL, USA: CRC Press, 2002.

[12] G. Stadler, "Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices," *Comput. Optim. Appl.*, vol. 44, no. 2, pp. 159–181, 2009.

[13] R. Herzog, G. Stadler, and G. Wachsmuth, "Directional sparsity in optimal control of partial differential equations," *SIAM J. Control Optim.*, vol. 50, no. 2, pp. 943–963, 2012.

[14] K. Kunisch, K. Pieper, and B. Vexler, "Measure valued directional sparsity for parabolic optimal control problems," *SIAM J. Control Optim.*, vol. 52, no. 5, pp. 3078–3108, 2014.

[15] A.-C. Boulanger and P. Trautmann, "Sparse optimal control of the KdV-Burgers equation on a bounded domain," *SIAM J. Control Optim.*, vol. 55, no. 6, pp. 3673–3706, 2017.

[16] Y. Privat, E. Trélat, and E. Zuazua, "Actuator design for parabolic distributed parameter systems with the moment method," *SIAM J. Control Optim.*, vol. 55, no. 2, pp. 1128–1152, 2017.

[17] D. Kalise, K. Kunisch, and K. Sturm, "Optimal actuator design based on shape calculus," *Math. Models Methods Appl. Sci.*, vol. 28, no. 13, pp. 2667–2717, 2018.

[18] H. K. Fathy, J. A. Reyer, P. Y. Papalambros, and A. G. Ulsoy, "On the coupling between the plant and controller optimization problems," in *Proc. Amer. Control Conf. (ACC)*, vol. 3, Jun. 2001, pp. 1864–1869.

[19] D. R. Herber and J. T. Allison, "Nested and simultaneous solution strategies for general combined plant and control design problems," *J. Mech. Des.*, vol. 141, no. 1, 2019, Art. no. 011402.

[20] D. W. Hahn and M. N. Özisik, *Heat Conduction*, 3rd ed. Hoboken, NJ, USA: Wiley, 2012.

[21] Y.-C. Fung, P. Tong, and X. Chen, *Classical and Computational Solid Mechanics* (Advanced Series in Engineering Science), vol. 1. Singapore: World Scientific, 2001.

[22] B. Rice, "Extreme ultraviolet (EUV) lithography," in *Nanolithography: The Art of Fabricating Nanoelectronic and Nanophotonic Devices and Systems*, M. Feldman, Ed. Philadelphia, PA, USA: Woodhead, 2014, pp. 42–79.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[24] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms Convex Programming* (Studies in Applied Mathematics), vol. 13. Philadelphia, PA, USA: SIAM, 1994.

[25] M. Minoux, *Mathematical Programming: Theory and Algorithms*. Chichester, U.K.: Wiley, 1986.

[26] O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, *The Finite Element Method: Its Basis and Fundamentals*, 7th ed. Oxford, U.K.: Butterworth, 2013.

[27] A. Borzì and V. Schulz, *Computational Optimization of Systems Governed by Partial Differential Equations* (Computational Science & Engineering), vol. 8. Philadelphia, PA, USA: SIAM, 2012.

[28] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization Algorithms on Matrix Manifolds*. Princeton, NJ, USA: Princeton Univ. Press, 2008.

[29] J. H. Manton, R. Mahony, and Y. Hua, "The geometry of weighted low-rank approximations," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 500–514, Feb. 2003.

[30] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich, *Optimization With PDE Constraints* (Mathematical Modelling: Theory and Applications), vol. 23. New York, NY, USA: Springer, 2009.

[31] T. Apel and T. G. Flaig, "Crank–Nicolson schemes for optimal control problems with evolution equations," *SIAM J. Numer. Anal.*, vol. 50, no. 3, pp. 1484–1512, 2012.

[32] R. Pinnau, "Model reduction via proper orthogonal decomposition," in *Model Order Reduction: Theory, Research Aspects and Applications* (Mathematics in Industry), vol. 13. Berlin, Germany: Springer, 2008, pp. 95–109.

Daniël W. M. Veldman received the M.Sc. degree in mechanical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2015, and the M.Sc. degree in mathematical sciences from Utrecht University, Utrecht, The Netherlands, in 2016. He is currently pursuing the Ph.D. degree within the Dynamics and Control Group, Department of Mechanical Engineering, Eindhoven University of Technology. His current research interests include the modeling, analysis, and control of nonlinear and thermomechanical systems.

Rob H. B. Fey received the M.Sc. degree in mechanical engineering and the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1987 and 1992, respectively.

From 1992 to 2002, he was a Senior Scientist with the Structural Dynamics Department, The Netherlands Organization for Applied Scientific Research, TNO, Delft, The Netherlands. Since 2002, he has been with the Dynamics and Control Group, Department of Mechanical Engineering, Eindhoven University of Technology, where he is currently an Associate Professor of structural dynamics. He has coauthored 53 refereed journal articles, 15 chapters in books, and 39 conference articles. His current research interests include the modeling, analysis, and validation of the dynamic behavior of complex, advanced structures and systems, MEMS, wafer stages, passive damping devices, and microturbines.

Dr. Fey was a recipient of the Shell Study Prize for his Ph.D. thesis. He is an Associate Editor of the *Journal of Vibration and Control*.

Hans Zwart (M'10) received the M.Sc. degree in mathematics from the University of Groningen, Groningen, The Netherlands, in 1984, and the Ph.D. degree in 1988. Since 1988, he has been with the Department of Applied Mathematics, University of Twente, Enschede, The Netherlands, where he is currently a Full Professor. Since 2011, he has also been holding a one-day-per-week appointment within the Department of Mechanical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands. He has authored a standard reference in the field of distributed parameter systems together with R. Curtain. His current research interest includes the control and analysis of distributed parameter systems, with an emphasis on the subclass of port-Hamiltonian systems.

Marc M. J. van de Wal received the M.Sc. and Ph.D. degrees in mechanical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 1993 and 1998, respectively. In 1998, he joined the Philips Center for Industrial Technology, Eindhoven. In 2010, he joined the Research Department, ASML, Veldhoven, The Netherlands. His current research interests include advancements in control system design for motion and thermal control applications.

Joris D. B. J. van den Boom received the M.Sc. degree in mechanical engineering from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2003. Since 2009, he has been with ASML, Veldhoven, The Netherlands, where he is currently a Thermal (System) Architect within the Machine Conditioning Cluster of the Business Function, Development and Engineering. His current research interests include the design, analysis, and control of various thermal and thermomechanical systems.

Henk Nijmeijer (F'00) was born in 1955. Since 2015, he has been the Scientific Director of the Dutch Institute of Systems and Control, Delft, The Netherlands. He is currently a Full Professor with Eindhoven University of Technology, Eindhoven, The Netherlands, and the Chairman of the Dynamics and Control Group. He has authored or coauthored several journal and conference articles, and several books.

Dr. Nijmeijer was or is on the editorial board of numerous journals. He has been an IFAC Council Member since 2011. He is a member of the Mexican Academy of Sciences. He is an Editor of *Communications in Nonlinear Science and Numerical Simulation*. He was a recipient of the IEE Heaviside Premium in 1990. He was a recipient of the 2015 IEEE Control Systems Technology Award. He was appointed an Honorary Knight of the "Golden Feedback Loop" Norwegian University of Science and Technology (NTNU) in 2011.