UNIVERSITÀ DEGLI STUDI DI TRIESTE
Dipartimento di Ingegneria e Architettura
XXXVI Ciclo del Dottorato Industriale in
Ingegneria Industriale e dell'Informazione

Advanced Optimization Techniques Applied to the Simulation of Laser Welding Process in Household Ovens

Ph.D. Student
Piero Favaretti

Ph.D. Program Coordinator
Prof. Fulvio Babich

Thesis Supervisor
Prof.ssa Lucia Parussini

Thesis Co-Supervisor
Bastian Grass

Anno Accademico 2023-2024

Abstract

Simulating the laser welding process inherently carries both high complexity and computational cost. Even more, when aiming to optimize this process, such as finding the combination of process variables that minimizes residual deformation, the computational expense required to individually analyze multiple configurations quickly becomes unsustainable.
In this doctoral research project, the overall computational cost has been reduced by acting on both the FEM modeling and optimization techniques. Initially, the model was simplified at mesoscopic level by disregarding microscopic dynamics and focusing on accurately representing the melt pool, particularly the laser penetration depth. This was achieved by introducing the concept of passive elements, specific FEM elements with thermomechanical properties dependent on the material state. Once the evaporation temperature is reached, these elements instantaneously transmit incoming energy to the underlying elements, facilitating keyhole formation.
Concerning the optimization process, to minimize computational cost, metamodels were employed. Specifically, two highly efficient machine learning techniques were adopted. Firstly, the optimum is iteratively determined by refining the metamodel only in the most promising regions, following the principles of Bayesian optimization. Secondly, a multifidelity approach was utilized for model training, involving the collaboration of high-fidelity and low-fidelity data to efficiently find the optimum solution. Regarding the multifidelity approach, the performance of both nested and non-nested infilling strategies has been investigated in relation to the data correlation coefficient.

Contents

List of Figures

List of Tables

List of Symbols

$\dot{E}_{in}$	incoming power into the system
$\dot{E}_{out}$	outgoing power from the system
$\dot{E}_{st}$	power stored in the system
Bi	Biot number
h	convective heat transfer coefficient
k	thermal conductivity
α	thermal diffusivity
ΔH_{evap}	evaporation enthalpy
ΔH_{melt}	melting enthalpy
q''	power per unit area
$\dot{q}$	internally generated power per unit volume
ρ	material density
c	specific heat
A_s	body outer surface
f_{G_volume}	3-D Gaussian power distribution
f_{G_plane}	2-D Gaussian power distribution
μ	mean value of Gaussian distribution
σ	standard deviation of Gaussian distribution
η	overall welding efficiency
ϕ	probability density function
Φ	cumulative distribution function
$\mathcal{GP}$	Gaussian process
$\mathcal{N}$	unit Gaussian process
$\mathbb{E}$	expectation
$m(x)$	mean function
$k(x, x')$	covariance function
K	covariance matrix
$f(x)$	real process
f_T	training outputs
f_*	Gaussian process (posterior) prediction (random variable)
$\bar{f}_*$	Gaussian process mean prediction
$\mathbb{V}$	Gaussian process variance prediction

l	fidelity level
$\rho_{(l-1)}$	correlation with level l-1 factor at level l
y	noisy observations
ϵ	noise
σ_n^2	variance of noisy observation
P	laser input power
pu	laser pulse duration
f	laser pulse frequency
v	laser speed

List of Acronyms

FEM	Finite Element Method
ACT	Ansys Customization Toolkit
$APDL$	Ansys Parametric Design Language
DOE	Design Of Experiment
PI	Probability of Improvement
$\mathbb{E}I$	Expected Improvement
AQF	AcQuisition Function
GP	Gaussian Process
PDF	Probability Density Function
CDF	Cumulative Distribution Function
MF	MultiFidelity
DCF	Data Correlation Factor
MOP	Multi-Objective Prognosis
$AMOP$	Adaptive Metamodel of Optimal Prognosis
HF	High Fidelity
LF	Low Fidelity

Chapter 1

Introduction - Project Genesis

The idea of pursuing a PhD dates back to my master's degree studies, but at that time, the desire to experience the working world prevailed. This decision led me, shortly after earning my master's degree, to leave Italy for Germany, where, after 5 years of groundwork in Baden-Württemberg and Bavaria, in 2018, I found a job as a simulation engineer at BSH GmbH, world leader in the production of household appliances.
Most of the projects I deal with on a daily basis are completed within a few weeks or at most a few months. However, there are issues that, due to their complexity, require appropriate time investments, specific methods, and advanced expertise. By their nature, these topics are usually incompatible with the fast-paced rhythms of the industry and are often addressed in an approximate manner or even postponed indefinitely.
In this organizational context, the pursuit of an industrial PhD fits perfectly, primarily as in-depth study of a complex subject and application of state-of-the-art technologies to a specific practical case. Framing a very complex and articulated problem in the form of a PhD has indeed allowed me,

first of all, to allocate the necessary internal resources for the project, as well as to draw on top-tier specific expertise provided by the university. From this idea, a collaboration between BSH and the University of Trieste was born, and the choice fell on a topic of great interest to the company, namely laser welding and its optimization.
But before delving into the project, here is some background information about the company I work for: BSH GmbH, acronym for Bosch Siemens Hausgeräte, is a German company based in Munich and specialized in the design and production of household appliances. It was established in 1967 through the collaboration between Robert Bosch GmbH and Siemens AG, and since 2015 BSH has been fully integrated into the Bosch Group. Currently, the company employs approximately 63,000 employees across 39 facilities and 3 continents. In 2022, it achieved a turnover of nearly 16 billion euros, making it the top-ranking company in Europe. The company's portfolio is extensive, and alongside the two internationally recognized flagship brands, Bosch and Siemens, there are also other smaller brands such as Neff and Gaggenau, as well as several local entities, like the historic German brand Constructa, which are now called "local heroes". The product portfolio is equally remarkable, encompassing not only household appliances but also service and ecosystem brands. Despite the household appliances sector being widely regarded as mature and traditional, the company has made significant investments in innovation in recent years, increasing its research and development expenditure to 5.3%. The drive for innovation is manifesting in various directions, ranging from digitalization to robotics, from sustainable packaging to Artificial Intelligence (AI). In the field of AI, the applications are countless. On the one hand, the company aims to utilize this new technology to offer customers new products and services, while on the other hand, it also seeks to enhance current products and manufacturing processes. This work

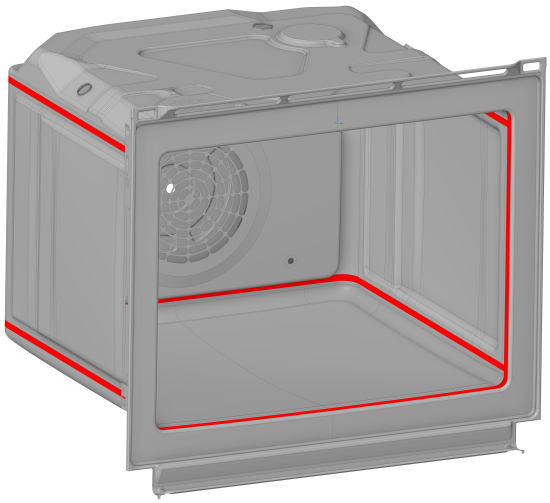
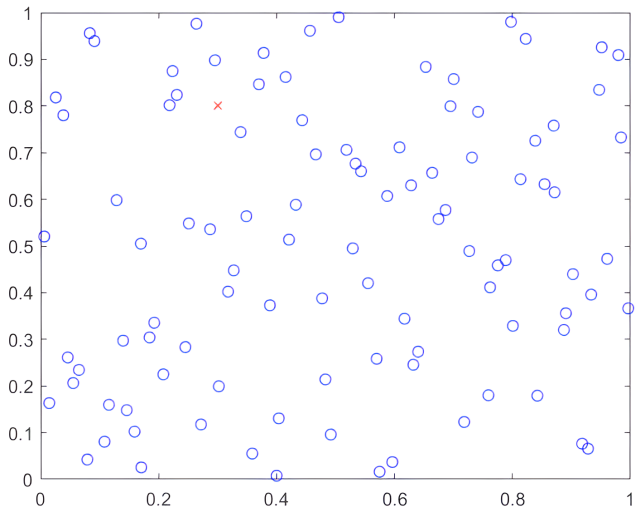


Figure 1.1: Example of oven cavity and, marked in red, the welded areas

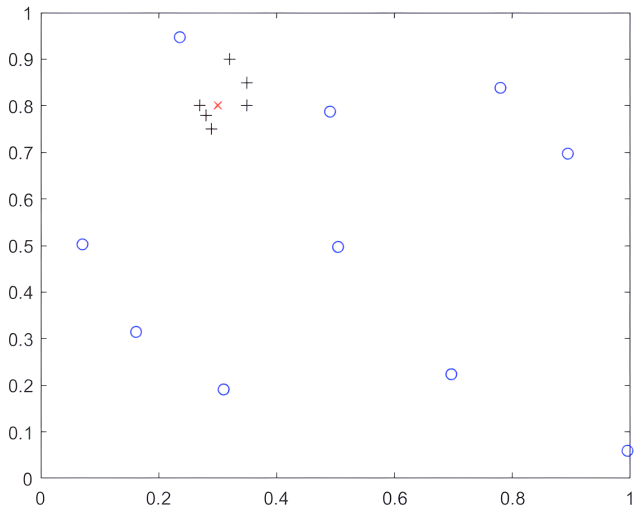
aligns with this second perspective.
In Traunreut, Bavaria, in the beautiful lakeside area led by the lake Chiemsee, BSH has a facility that was established in 1952. At this facility, the company develops ovens, cooktops, microwave ovens, and small household appliances. Ovens and cooktops are directly manufactured on-site. The core of an oven is the cavity, made of DC04 or DC06 steel and coated with enamel to protect it from oxidation, and it consists normally of four parts that are produced through stamping and welded together: top and bottom, initially stamped together as a combined component (known in German as Kombi-Teil) and then separated, along with the mantle and flange. The various parts are joined using linear welds in the overlapping areas (figure 1.1).
For quality and cleanliness reasons, as well as process efficiency and speed, laser welding has been used for several years now. However, laser welding also comes with

disadvantages or, more accurately, challenges. One of the primary challenges is its sensitivity to process parameters: laser welding involves very high power concentrated in small surface areas (in this specific case, several kW focused within a few tenths of a millimeter), and a slight variation in one of the characteristic parameters can have a significant effect on the final result. For this reason, the machines must be highly accurate, perfectly calibrated, and the parameters must be chosen with extreme care. Currently, the selection of welding parameters relies solely on the technicians' experience, a few simple theoretical, empirical, or semi-empirical formulas, and numerous experimental trials. This approach results in a simplistic process that is not easily adaptable when there are changes in boundary conditions or design parameters, such as plate thickness. Furthermore, due to the lack of information regarding the depth and diameter of the weld area at the interface, it becomes challenging to make assessments regarding the strength of the joint. Additional mechanical tests, such as shearing or peeling tests, would need to be conducted to evaluate its resistance. As further confirmation of what has been highlighted so far, during the quality control phase, welded components occasionally fail to meet the required standards and exhibit significant permanent deformations. These deformations not only pose aesthetic concerns but can also jeopardize the proper functionality of the oven. One exemplary case is the deformation of the mantle, also known as the side wall, which sometimes reaches a magnitude that hinders, if not prevents, the installation of the side racks that support the baking trays. Less visible but equally important is the issue of residual stresses: residual deformations are accompanied by residual stresses, which can be further accentuated during assembly due to increased clearance between the parts. In areas where these stresses are higher, during the regular operation of the oven cracks in the enamel layer could potentially appear, a condition that must be avoided for

aesthetic reasons and to prevent oxidation of the underlying steel layer.
Therefore, the primary objective of the welding process should be, in addition to ensuring a strong and durable joint, to minimize deformations and residual stresses. To this end, BSH initiated a multidisciplinary project in late 2019 in collaboration with the University of Trieste, framed as an Industrial Doctorate program, of which this work represents the final report. The project revolves around two main themes: the first is the modeling of the laser welding process using the Finite Element Method (FEM), while the second is the optimization of the process itself. The underlying idea of this activity is to delve into the latest optimization techniques based on machine learning and apply them to the specific case of the industrial laser welding process.
The complexity of the modeling phase lies not only in the nature of the physical problem to be analyzed but also in the need to simplify the model to a degree that makes it suitable for the optimization process, which of course requires a certain number of iterations. All of this without excessively compromising accuracy.
Regarding the optimization process, in order to minimize also in this case its computational cost, two specific approaches have been adopted: the Bayesian approach and the multifidelity approach. The first one involves iteratively adding points to the data used to train the surrogate model only in the area of interest, thus avoiding the need to use a large number of Design of Experiments distributed throughout the domain (figure 1.2). The multifidelity approach, on the other hand, involves combining data of different nature and quality for the creation of the surrogate model. This allows not only the utilization of all available resources but also the use of less valuable but faster sources of information to find the optimum. In the case of an optimization based on FEM simulations, for instance, low-fidelity data could



(a)



(b)

Figure 1.2: Conceptual illustration of unstructured optimization (a) and Bayesian optimization (b) in a biparametric domain. Red X represents the optimum, blu circles the initial sample points and the black crosses the iterative refinement

refer to the results obtained from models with a coarser mesh compared to high-fidelity data.
Maintaining the initially separate narrative, this introductory chapter will be followed by a chapter on FEM modeling of the laser welding process (Chapter 2) and a chapter dedicated to the optimization algorithms used to minimize the residual deformations (Chapter 3). Chapter 4 will then discuss the results of the optimization, and finally, in Chapter 5, conclusions will be drawn and potential future developments, both about FEM model and optimization strategies, will be discussed.

Chapter 2

Laser Welding Process Modeling

2.1 Physics of Laser Welding

In laser welding, the power intensity, expressed in W/mm^2, reaches such high levels that the vaporization temperature is achieved. As a result, a portion of the vaporized metal escapes from the system, representing a power loss. However, this allows for significantly deeper heat penetration compared to a low-intensity process. This is because the combined effects of material removal and vapor pressure on the liquid sidewalls create what is known as a "keyhole" - a microstructural hole in the metal layer that enables the laser beam to penetrate directly and indirectly through multiple reflections. Hence, laser welding exhibits high energy efficiency (figure 2.1).
From a physical standpoint, this process is highly complex. Not only does it involve all three phases of matter (solid, liquid, and vapor), but it also entails intricate dynamics of thermal and mass transfer, as well as interacting forces. The keyhole cavity contains metal vapor that partially absorbs the incoming laser light. The vapor's charged particles acquire kinetic energy from the beam pho-

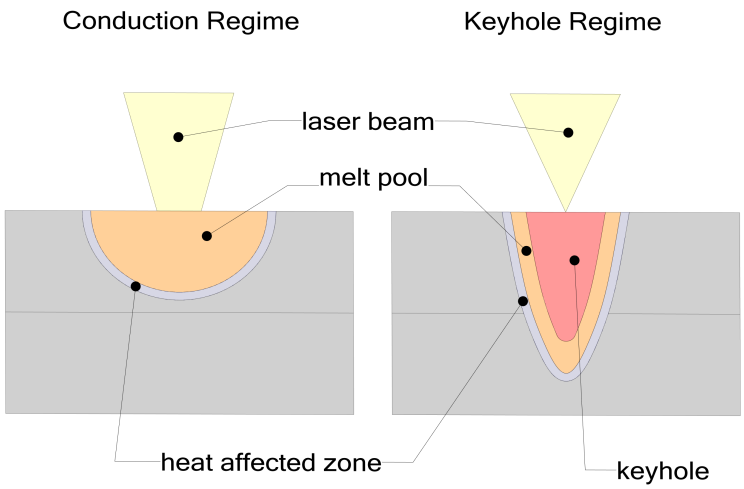


Figure 2.1: Schematic representation of low- and high-intensity laser beam welding regimes

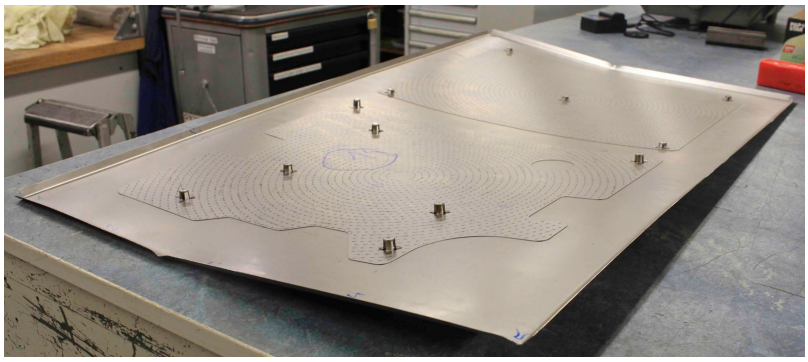
tons through a process called inverse Bremsstrahlung absorption (iB-absorption). When this energy gain becomes significant, it leads to beam damping, further ionization of the metal vapor, an increase in plasma temperature, and consequently, an increase in the iB-absorption coefficient. However, this effect remains as long as the plasma temperature stays below a critical value, as described by Zhou [38]. The presence of plasma within the keyhole plays a positive role in the welding process. It acts as a protective barrier, preventing the keyhole cavity from cooling due to the surrounding atmosphere. Additionally, plasma radiation enhances vaporization at the keyhole surface. Once the critical temperature is surpassed, damping becomes dominant, and vaporization diminishes.
The metal vapor can also coalesce into larger nanoparticles, forming fumes. Particles of comparable size or larger than the laser beam wavelength attenuate the laser power through Mie-scattering, while much smaller particles attenuate it through Rayleigh scattering. As the metal vapor

flows out of the cavity, it creates a plume, which is continually replenished by newly evaporated material. The plume can also contribute to beam damping, as described by Moscicki [29].
Due to temperature dependence, both the absorption and scattering coefficients are closely linked to the plasma properties and the temperature field. The angle of incidence of the laser beam ray with respect to the keyhole surface also affects the absorption process, as described by Bergstrom [6]. The geometry of the keyhole surface is influenced by the energy balance among the beam, plasma, and pool, as well as the force balance between plasma and pool (e.g., surface tension, Marangoni force, and recoil pressure). The interplay of these factors results in the complex and tightly coupled physical phenomena observed in keyhole laser welding. Fabbro [10] conducted experimental analysis of the keyhole structure at different speed regimes, highlighting the effects of ablation pressure, vapor plume, surface tension, gravity, and induced electromagnetic forces.

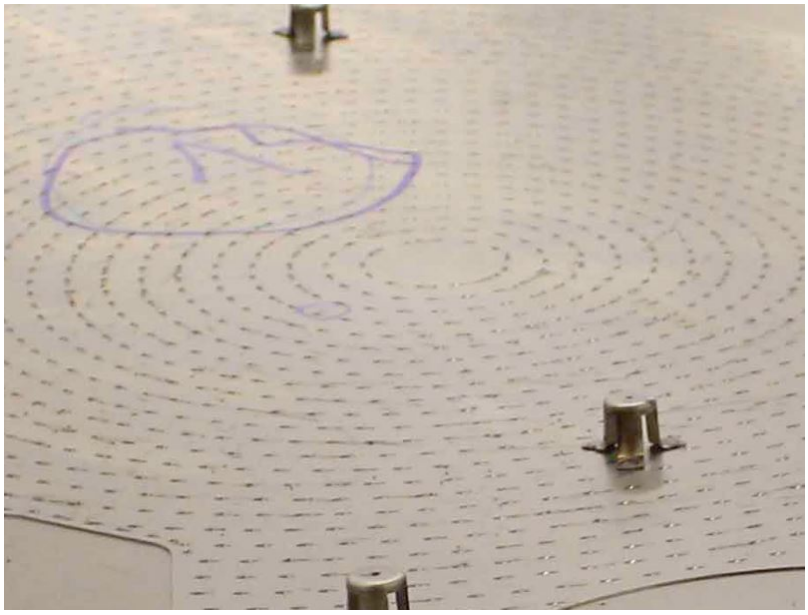
2.2 Numerical Simulation of Laser Welding

Several models have been proposed to analyze these phenomena. In [7] and [8] Courtois developed a 3-D model that describes both heat and fluid flow characteristics. It considers the multi-reflections of the laser beam in the keyhole, treating it under its wave form by solving Maxwell's equations. Zhang [36] employed the volume of fluid (VOF) method and analyzed the multi-reflections of the laser beam using a ray-tracing algorithm. Both methods share a high level of complexity and computational time. Alternatively, simplified methods are available as alternatives to the aforementioned self-consistent approaches, characterized by a very high level of detail. These simplified methods typically

involve simplifications in modeling the heat source and define the power as a function of penetration. For example, Hozoorbakhsh [19] proposed a heat source that combines a surface and an adaptive volumetric component, Shanmugam [32] used the double-ellipsoidal heat source concept based on Goldak studies [15], while Artinov [3] introduced a two-step thermal analysis where the moving heat flux results from a moving temperature distribution based on a fixed laser beam thermal analysis. If the focus of the study is solely on the keyhole depth and evaluating the size of the melted area, fundamental parameters for the welded joint, further simplifications can be adopted. For instance, Fabbro [11] proposed a model that describes the laser beam penetration as a function of a normalized aspect ratio ($R0$) and a characteristic speed ($V0$) related to the Peclet number. After solving the thermal fluid dynamical equations, a mechanical simulation must be added to evaluate residual stresses and deformations post-welding. This simulation is based on the temperature distribution calculated with the initial model. Huang [20] suggested employing a local solid model to evaluate temperature distribution and plastic strains, as well as a global shell model to assess the overall deformation resulting from the local plastic strains. Xu [35] successfully captured changes in convexity of ultra-thin 316 stainless steel plates welded with a pulsed laser process, considering the input power, angular deformation, and longitudinal shrinkage force. In this work, a thermal and elasto-plastic FEM model is proposed to simulate pulsed laser welding processes of sandwich layers, specifically focusing on residual deformations (figure 2.2). This model aims to serve industrial purposes and therefore needs to be both accurate and computationally efficient, while also accommodating complex geometries and welding paths. Integrating the microscale of the keyhole and the macroscale of the actual welded parts in the same model would exponentially increase computational time. To address this, the



(a) Case of two metal sheets welded together that lose planarity



(b) Detail of the complex welding path

Figure 2.2: Problem identification

keyhole description is simplified at the mesoscale level by introducing the concept of passive elements, which accurately predict the keyhole size based solely on the material properties of the components. Expanding on Tirand proposal [33] to increase thermal conductivity to simulate the Marangoni effect, in this case, the thermal conductivity is modified to also simulate metal vaporization and the presence of air inside the keyhole, which adds negligible thermal resistance.

2.3 Concept and Theory of Passive Elements

To simulate the material vaporization at high temperature and the formation of the keyhole, a simplified approach defined as the "passive element" concept is introduced. Instead of physically modeling the keyhole, such as using the *ekill* command in Ansys to deactivate elements above the vaporization temperature as suggested by Zhang [37], this method focuses on enhancing heat penetration through mathematical manipulation of the thermal material properties of the welded part. This approach involves modifying the thermal conductivity, significantly increasing it along the direction of penetration while setting the in-plane values to zero. By doing so, heat is rapidly transferred to the bottom of the keyhole, facilitating continuous penetration. Above the vaporization temperature, the thermal conductivity transitions from isotropic to orthotropic behavior. In this specific case a value of 10^{11} W/mK is assigned to the thermal conductivity, ensuring that under steady-state conditions, with a power of 3.5 kW applied to a radius of 0.2 mm, the temperature difference per unit length remains below 1°C (figure 2.3).
The elements transition into a state of thermal passivity, thus justifying their designation. To provide further clar-

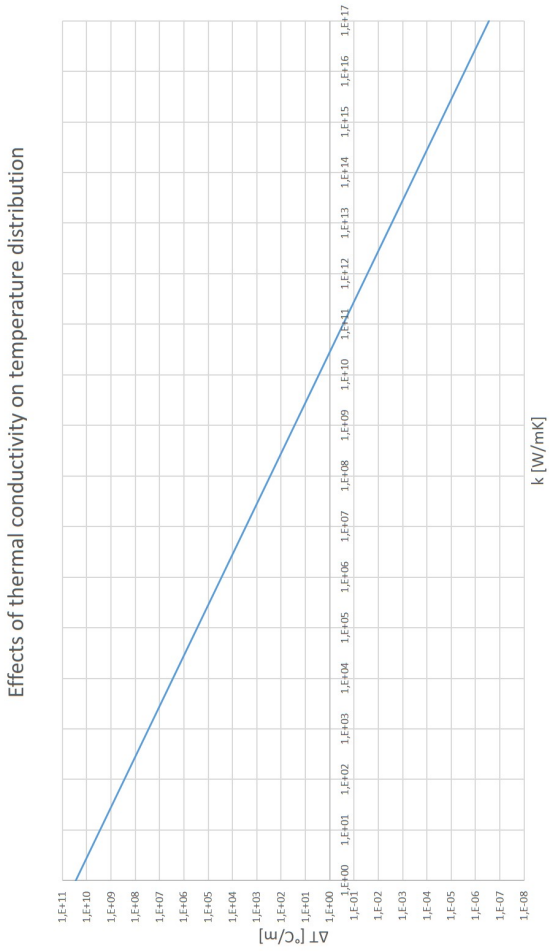


Figure 2.3: Influence of the thermal conductivity k on the stationary temperature distribution of an infinite thick DC04 steel plate with a heat source of 3.5 kW even distributed on a circular surface of radius 0.2 mm. This temperature difference introduces a numerical error due to the non-infinite k

ity, they are passive with respect to the energy above the vaporization temperature, while still retaining the energy associated with the solid and liquid phases. It is important to note that a portion of this energy is lost through the vaporization process and must be considered when calculating the overall process efficiency, as described in paragraph 5 and evaluated through experimental means. To assess the theoretical validity of this approach, it is essential to demonstrate that the assumption of steady-state conditions in the keyhole zone holds true at each time step. This assumption implies that the heat input from the outer surface is always equal to the heat output at the bottom of the keyhole, with no energy storage in the system ($\dot{E}_{in} = \dot{E}_{out}$). In simpler terms, the transient period must be shorter than the minimum time step, which, in this specific scenario, is 1 ms. Given the significantly high thermal conductivity above the evaporation temperature, the assumption of lumped capacitance is deemed acceptable. This approximation remains coherent for Bi values below 0.1, where Bi represents the Biot number [21]. The Biot number is defined as the quantity hl/k, a dimensionless parameter that plays a fundamental role in conduction problems. It provides a measure of the temperature drop in the solid (described by the conductivity k) relative to the temperature difference between the surface and the fluid (described by the heat transfer coefficient h) and can also be viewed as a ratio of thermal resistances. The core concept behind the lumped capacitance method is that the solid's temperature remains spatially uniform during the transient process. In other words, temperature gradients within the solid are considered negligible. From Fourier's law,

$$q_x^{''} = -k\frac{dT}{dx}, \tag{1}$$

while this assumption implies an infinitely high thermal conductivity, such a condition is practically unattainable. However, this condition is closely approximated if the resistance to conduction within the solid is much smaller than the resistance to heat transfer between the solid and its surroundings. By neglecting temperature gradients within the solid, the problem can no longer be analyzed solely through the general heat equation,

$$\frac{\delta T}{\delta t} = \alpha \nabla^2 T + \frac{\dot{q}}{\rho c}. \tag{2}$$

Instead, the transient temperature response is determined by formulating an overall energy balance on the solid, relating the rate of heat loss at the surface to the rate of change of internal energy,

$$-\dot{E}_{out} = \dot{E}_{st}, \tag{3}$$

$$-hA_s(T - T_\infty) = \rho V c \frac{dT}{dt}. \tag{4}$$

Introducing the temperature difference,

$$\theta \equiv T - T_\infty, \tag{5}$$

and recognizing that $d\theta/dt = dT/dt$ if T_∞ is constant, it follows that,

$$\frac{\rho V c}{hA_s} \frac{d\theta}{dt} = -\theta. \tag{6}$$

Separating variables and integrating from the initial conditions, for which $t = 0$ and $T(0) = T_i$, we then obtain,

$$\frac{\rho V c}{h A_s} \int_{\theta_i}^{\theta} \frac{d\theta}{\theta} = -\int_0^t dt, \tag{7}$$

with $\theta_i \equiv T_i - T_\infty$.

Solving the integral,

$$\frac{\rho V c}{h A_s} \ln \frac{\theta_i}{\theta} = t, \tag{8}$$

$$\frac{\theta}{\theta_i} = \frac{T - T_\infty}{T_i - T_\infty} = \exp\left[-\frac{h A_s}{\rho V c} t\right], \tag{9}$$

$(\rho V c)/(h A_s)$ is called thermal time constant and it is usually identified with τ. The larger is τ, the faster will be the transient (figure 2.4).

Nevertheless, it's important to note that the conductivity k does not appear in this therm. To evaluate the influence of k on the transient duration and to ensure the consistency of assuming steady-state conditions, the more general transient formulation in a semi-infinite solid must be considered. In this case, the problem becomes more complex, as the solution is a function of both space and time. To simplify the problem, the new variable η is defined as,

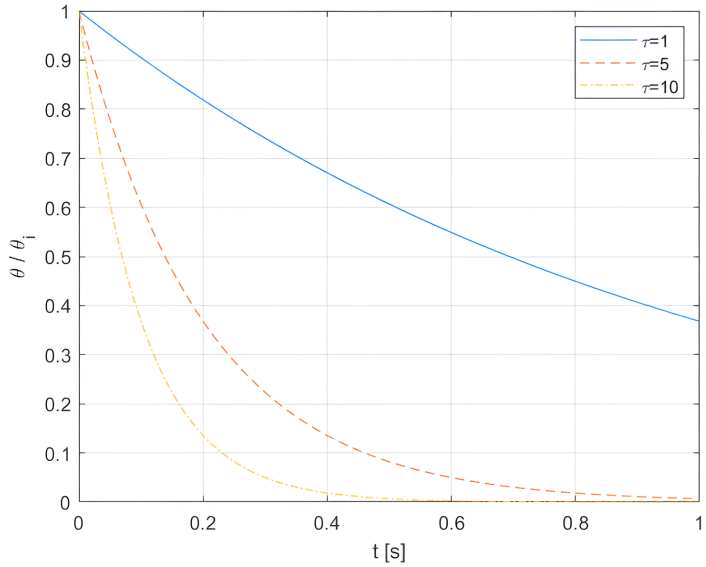


Figure 2.4: Transient temperature in body with uniform temperature distribution (Lumped Capacitance Method) in case of Dirichlet boundary conditions

$$\eta \equiv \frac{x}{\sqrt{4\alpha t}}. \tag{10}$$

Considering that,

$$\frac{\delta T}{\delta x} = \frac{dT}{d\eta}\frac{\delta \eta}{\delta x} = \frac{1}{\sqrt{4\alpha t}}\frac{dT}{d\eta}, \tag{11}$$

$$\frac{\delta^2 T}{\delta x^2} = \frac{d}{d\eta}\left[\frac{\delta T}{\delta x}\right]\frac{\delta \eta}{\delta x} = \frac{1}{4\alpha t}\frac{d^2 T}{d\eta^2}, \tag{12}$$

$$\frac{\delta T}{\delta t} = \frac{dT}{d\eta}\frac{\delta \eta}{\delta t} = \frac{x}{2t\sqrt{4\alpha t}}\frac{dT}{d\eta}, \tag{13}$$

then, for the Fourier equation for $\dot{q} = 0$,

$$\frac{\delta^2 T}{\delta x^2} = \frac{1}{\alpha}\frac{\delta T}{\delta t}, \tag{14}$$

becomes,

$$\frac{d^2 T}{d\eta^2} = -2\eta\frac{dT}{d\eta}, \tag{15}$$

$$\frac{d(dT/d\eta)}{(dT/d\eta)} = -2\eta d\eta. \tag{16}$$

Integrating twice,

$$\ln\left(dT/d\eta\right) = -\eta^2 + C_1, \tag{17}$$

$$\frac{dT}{d\eta} = C_1 \exp(-\eta^2), \tag{18}$$

$$T = C_1 \int_0^{\eta} \exp(-u^2) du + C_2, \tag{19}$$

with u as dummy variable. Applying now, for example, Dirichlet b.c. such that,

$$T(\eta = 0) = T_s, \quad \text{(i.c.)}$$

$$T(\eta \to \infty) = T_i, \quad \text{(b.c.)}$$

the expression for the dimensionless temperature in function of η is then,

$$\frac{T - T_s}{T_i - T_s} = \frac{2}{\sqrt{\pi}} \int_0^{\eta} \exp(-u^2) du \equiv \text{erf}(\eta), \tag{20}$$

with $C_1 = 2(T_i - T_s)/\sqrt{\pi}$ and $C_2 = T_s$. erf is the Gaussian error function. Knowing η it is now possible to evaluate one parameter between x and t but not both. The depth of the keyhole of course cannot exceed the total thickness of the metal sheets. In this specific case, two 0.5 mm metal sheets are joined together and thus the total thickness is 1 mm. Setting the density and specific heat of steel above the evaporation temperature to $\rho = 7287$ kg/m^3 and $c = 573$ J/kg/K, it is now possible to evaluate the time required to reach the steady state conditions[1]. The imposed temperature is reached only asymptotically but 99.9% is considered an acceptable approximation. Based on the given data, this

[1] These values have no physical meaning but are required for numerical consistency

temperature is reached after $t \approx 2e^{-6}$ s, corresponding to $\eta = 2.33$ (figure 2.5).

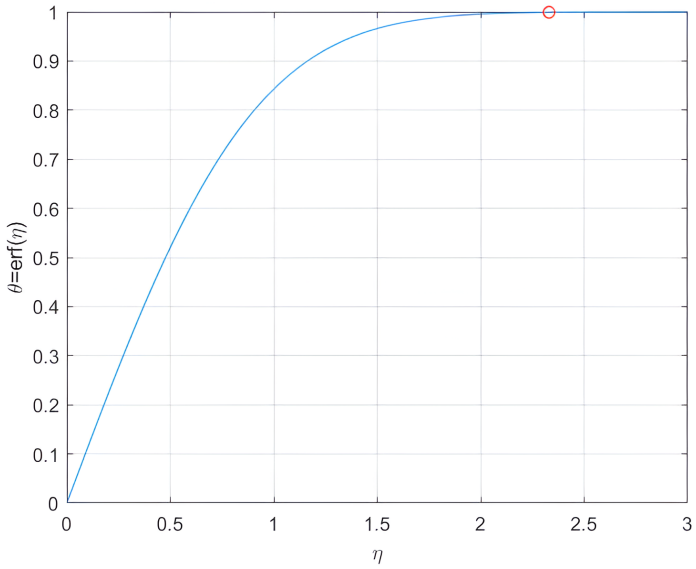


Figure 2.5: Temperature distribution in a semi-infinite solid in function of space and time. The marked spot represents the condition for $\theta = 99.9\%$

This value is two orders of magnitude smaller than the smallest time step and thus the assumption of steady state conditions in the keyhole zone at every time step is valid.

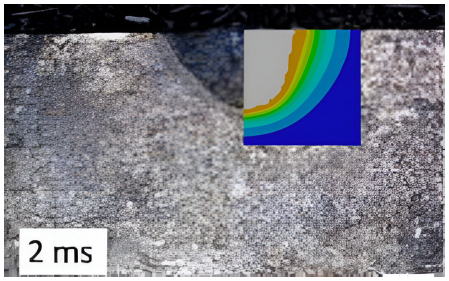
2.4 Experimental Validation of Passive Elements

The concept of passive elements is validated using experimental data sourced from the literature [26]. The problem being studied is axisymmetric and can be simulated using a 2-D model. The experimental data refers to the penetration

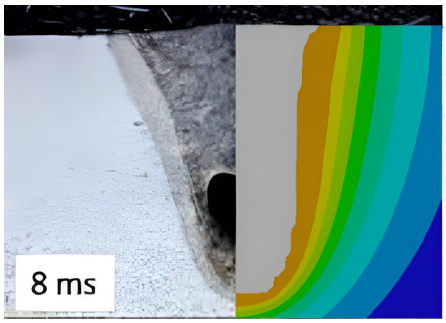
depth of a keyhole on a steel plate under three different laser powers (1000, 1250, and 1500 W) during a 25 ms transient period. The power source remains constant throughout the experiment. Assuming an overall process efficiency of 65%, which includes all power losses, there is a strong agreement in all three cases (as shown in figures 2.6 and 2.7).
The main power losses occur at laser level, for example due to transmission through a transparent or translucent material, on the welding surface due to reflection and scattering and due to plasma absorption. In summary, the passive elements, with knowledge of the overall efficiency of the welding process, effectively describe the formation of the keyhole. Given their simplicity, malleable definition, and time-saving nature, they represent a valid concept for analyzing welding processes in industrial applications.

2.5 Theory of Heat Source

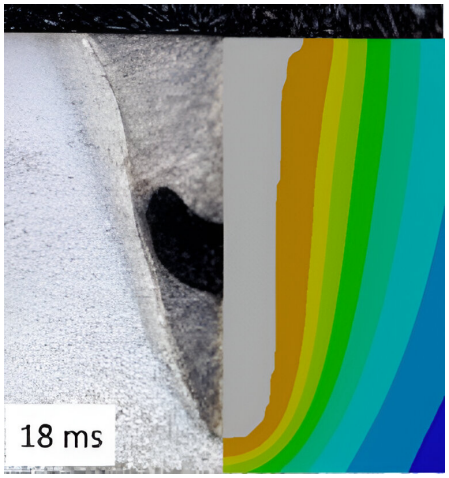
Regarding the laser beam, the heat source depends on both position and time. The position refers only to the in-plane position, not the coordinate along the layer thickness. This means that defocusing of the laser beam and reflections along the keyhole do not affect the total power or its distribution. Therefore, more complex approximations such as the 3-D conical heat source or the combination of conical and cylindrical heat sources as described by Dal [9] are not necessary in this case. By using passive elements that facilitate heat penetration, the heat source is simplified to a 2-D distribution applied solely to the upper face of the plate.
A generic heat source is mathematically described as a heat flux distribution, which is a function of space (x, y) and time. Dimensionally, this quantity represents power per unit area. Each welding technology has its own specific power distribution, which can be approximated using math-



(a) at 2 ms

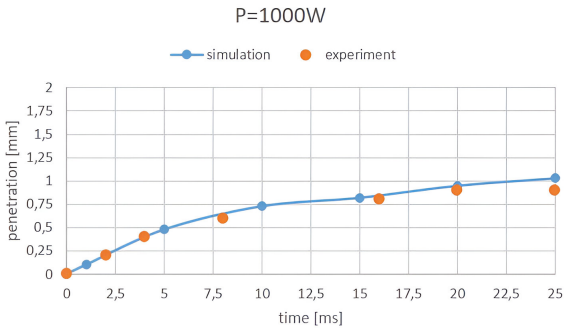


(b) at 8 ms

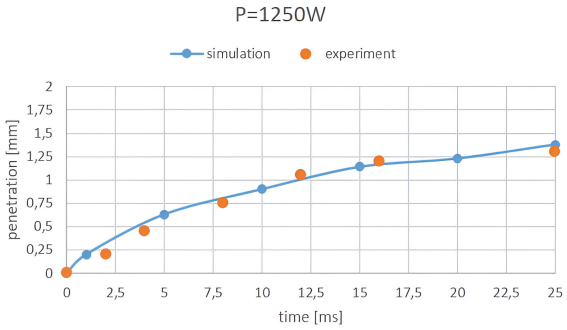


(c) at 18 ms

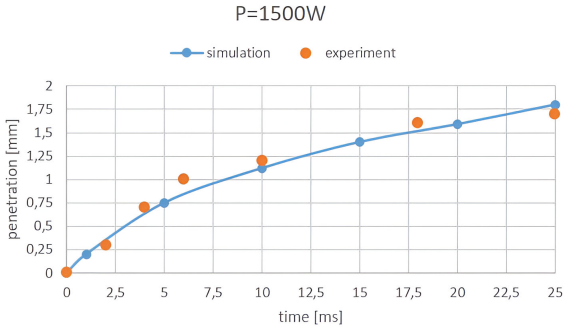
Figure 2.6: Comparison of experimental and numerical results with P = 1500 W



(a)



(b)



(c)

Figure 2.7: Comparison between simulated and experimental keyhole depth function of time for different powers

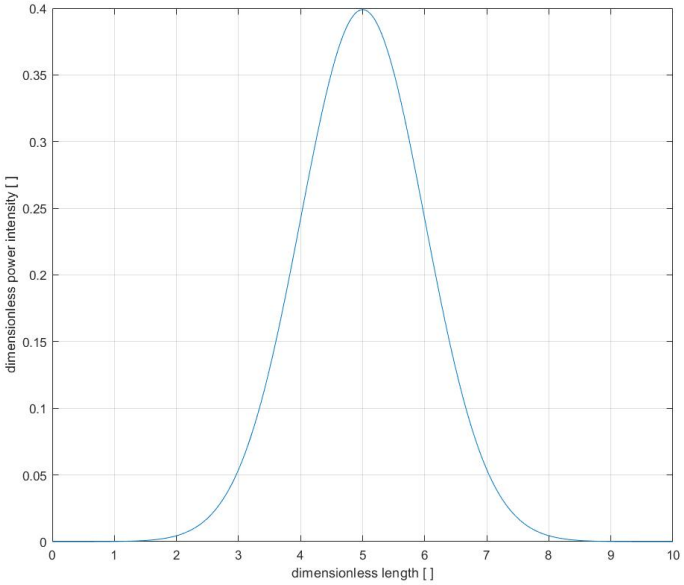


Figure 2.8: Plot of the 1-D Gaussian distribution defined in equation (21) with $\sigma = 1, \mu = 5, P = 1$

ematical functions or a combination of functions. Bradac [22] argues, for instance, that while the 3-D Gaussian distribution is more suitable for simulating laser and electron beam welding, arc welding is better described by a hemispherical surface or a double-ellipsoid source. In the case of a laser beam, the Gaussian distribution provides a good fit, characterized by a peak in the center and exponentially decreasing values with distance.
The general formulation of the Gaussian distribution on a plane (figure 2.8) is as follows,

$$f_{G\_plane} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{21}$$

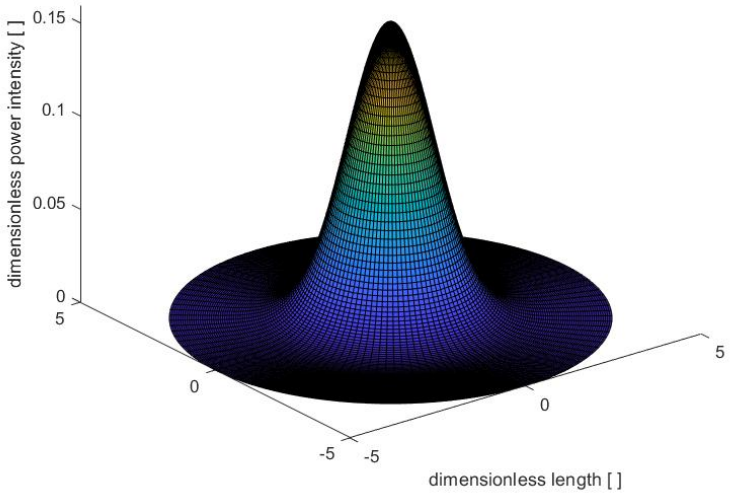


Figure 2.9: Plot of the 2-D Gaussian distribution defined in Equation (22) with $\sigma = 1, \mu = 0, P = 1$

The given passage describes the relationship between the standard deviation (σ), mean (μ), and total power (P). The standard deviation represents the distance from the mean that encompasses 34.15% of the total power. It is often linked to half of the beam's radius (r). This implies that the beam's radius is equivalent to two standard deviations. At this radius, the power amounts to 13.5% of the maximum value.
In the three-dimensional space (figure 2.9), the Gaussian function takes on the following general form,

$$f_{G\_volume} = Ae^{-\left(\frac{(x-\mu)^2}{2\sigma_x^2}+\frac{(y-\mu)^2}{2\sigma_y^2}\right)}. \tag{22}$$

If $\sigma_x = \sigma_y$ and $\mu = 0$, then,

$$f_{G\_volume} = Ae^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}. \tag{23}$$

Integrating in the domain $]-\infty, \infty[$, it can be demonstrated that,

$$\int\int Ae^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} dxdy = 2\pi A\sigma^2. \tag{24}$$

Considering that the integral of the heat flux is the total heat flow, this means,

$$2\pi A\sigma^2 = P \ \ [\mathrm{W}], \tag{25}$$

and thus,

$$A = \frac{P}{2\pi\sigma^2} \ \ [\mathrm{W/mm^2}]. \tag{26}$$

Substituting equation (26) in equation (23),

$$f_{G\_volume}(x, y) = \frac{P}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}, \tag{27}$$

valid for a steady laser beam with center in the origin and no losses. For a moving heat flux with a certain efficiency described by a parameter η, equation (27) becomes,

$$f_{G\_volume}(x, y, t) = \frac{\eta P}{2\pi\sigma^2} e^{-\left(\frac{(x-x_{0(t)})^2+(y-y_{0(t)})^2}{2\sigma^2}\right)}, \tag{28}$$

where x_0 and y_0 are the moving position of the laser beam center, function of time, and η is an overall efficiency that takes into account all losses and has been evaluated experimentally during the calibration process.

2.6 Heat Source Model

In the context of FEM analysis, a heat source in the Ansys ACT Extension "Moving Heat" can be defined as either a surface load or a volumetric load. The volumetric load provides additional control over heat penetration compared to the surface load by introducing the absorption coefficient as an extra parameter. Both approaches share certain limitations:

1. Only the Gaussian power distribution is available, meaning that not all types of heat sources can be accurately described using this tool.
2. Pulsed processes cannot be modeled using this extension.
3. The heating paths need to be defined in the CAD model.
4. The input parameters are not parameterizable, which means they cannot be optimized automatically.

For further details regarding individual input parameters, please refer to the original extension documentation. Both the methods proposed by Ansys and the new approach presented in this work can be categorized as simplified methods, as they focus on the macroscopic effects of heat transfer and consider thermal conduction as the dominant mechanism. The study by Geng [14] demonstrates that heat conduction is the primary heat transfer mechanism in weld pools, while melt convection becomes crucial at the microstructural scale.

The core principle of the presented method is to divide the entire space-time domain, through which the heat flux propagates, into numerous cells, creating a finely-gridded structure. At each point within this domain, the heat flux is calculated. The grid's resolution is determined by the distance between adjacent points. A higher resolution leads to a more accurate capture of the moving heat flux. For instance, if the moving heat source extends from the origin in the x- and y-directions over a distance of 0.1 m, starting at 0 s and ending after 1 s, and assuming a spatial resolution of 0.02 m and a temporal resolution of 0.2 s, the total number of degrees of freedom would be 216. The domain can be visualized as shown in Figure 2.10.

The presented method offers two significant advantages. Firstly, the domain and heat flux are defined only once during the pre-processing phase, ensuring consistency throughout the entire simulation. Secondly, all variables of the problem can be parameterized, allowing for optimization as each grid point serves as a degree of freedom for all variables. The required resolution of the grid depends on the specific characteristics of the problem. If the heat source has a small beam radius, a high spatial resolution is necessary. Similarly, if the process occurs rapidly, a high time resolution is required. Laser welding processes exhibit both of these conditions, which explains the need for high-resolution

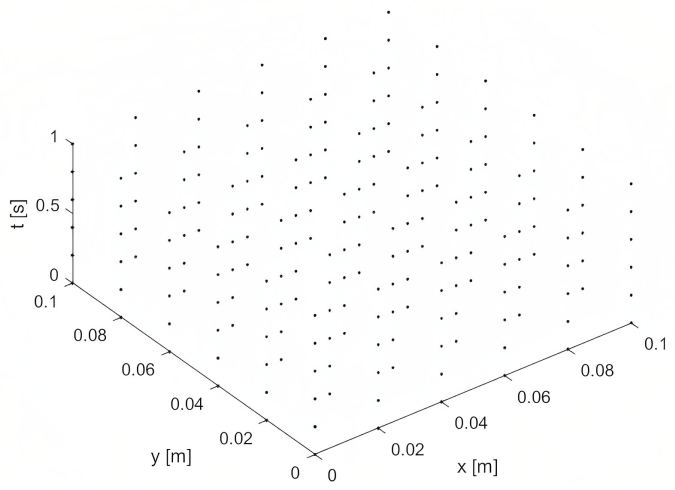


Figure 2.10: Evaluation points for a grid with space resolution of 0.02 m, time resolution of 0.2 s, space extension of 0.1 m and time extension of 1 s

domains in such cases.
Figure 2.11 illustrates a generic Gaussian heat flux distribution with a power of P = 1 W and a beam radius of r = 0.02 m (2σ) at three different resolutions. It is evident that a resolution of 0.02 m, which is equal to the beam radius, is insufficient to accurately capture the curve. The peak is completely missed, significantly affecting the total power introduced into the system (the area under the Gaussian curve). On the other hand, the resolution of 0.01 m appears acceptable as it delivers similar results to a resolution ten times higher.
The process of defining the heat flux involves two steps: the first step is carried out in APDL (ANSYS Parametric Design Language) and involves indexing the heat flux array in space and time based on the required resolution for the problem. The second step is implemented in Matlab, where the heat flux is calculated based on spatial and temporal

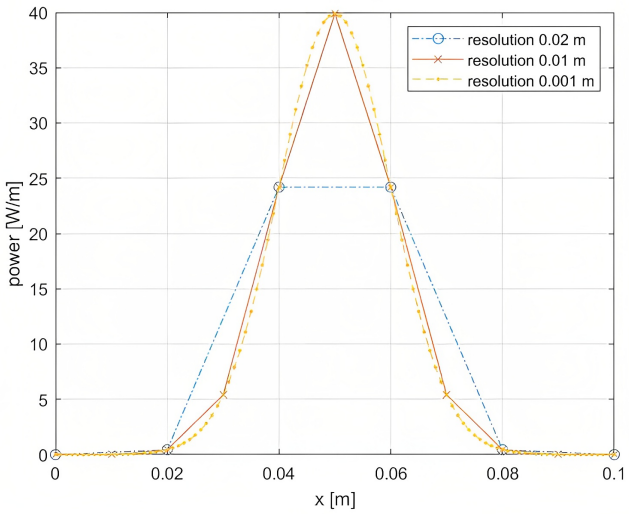


Figure 2.11: Accuracy of the Gaussian heat source distribution function of the resolution

variables. The interface between the two programming languages is established through a text file generated in Matlab, which contains APDL expressions for assigning the calculated heat flux to the corresponding indexes. These two steps are part of an iterative process designed for optimizing a laser welding process. The optimization is focused on minimizing residual deformation after welding by adjusting laser power, pulse duration, frequency, and velocity.
To reduce computational time, a metamodel trained using Kriging-based techniques is employed. Figure 4.1 illustrates the complete iterative optimization workflow. It begins with a set of DOEs that indicate the residual deformation for specific combinations of input parameters. An initial version of the metamodel is trained using these DOEs. The optimization is then performed on the metamodel, making predictions on deformation. The goal of the optimization is to find the minimum of the function and minimize residual deformation. However, since the function is not known

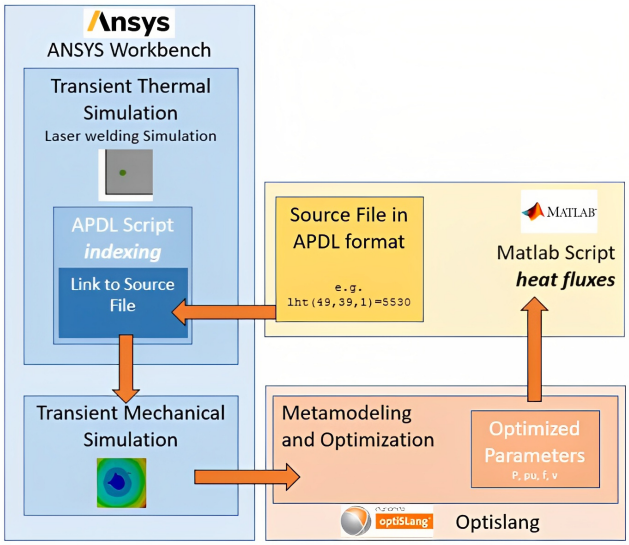


Figure 2.12: Graphic representation of the complete optimization workflow

across the entire domain, the optimization is not solely focused on finding the global minimum. Instead, the minimum is indirectly determined by maximizing the Expected Improvement ($\mathbb{E}$I) acquisition function, which considers areas of high uncertainty as well as the actual minimum. The optimal configuration obtained is then used to create a new power source distribution, and a new simulation is run. The set of DOEs is updated, and the optimization process continues until the $\mathbb{E}$I tolerance or a maximum number of iterations is reached. APDL and Matlab scripts are described in detail in Appendices A and B.

2.7 Validation of Heat Source Model

To validate the accuracy of the new heat flux formulation, the results obtained have been compared to the temperature field generated on a 1 mm steel plate using the Ansys ACT extension "ACT_MovingHeat_R170_v4.1" with the same heating source parameters. There is a slight difference in the definition of the Gaussian heat distribution between the two cases:
The ACT extension models the Gaussian heat flux source using the equation,

$$q = C_2 e^{-\frac{(x-x_0)^2+(y-y_0)^2}{C_1^2}}, \tag{29}$$

where q represents the heat flux on the desired surface, C_1 is the radius of the beam, C_2 is the source power intensity, and x_0 and y_0 are the instantaneous coordinates of the heat source center. On the other hand, the APDL formulation is given by,

$$q = \frac{P}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}}, \tag{30}$$

where P is the total power in watts and σ is the standard deviation, conventionally equal to half the radius of the beam. In the validation analysis, the values used in the ACT Moving Heat Flux extension are listed in Table 2.1.
The APDL script, instead of radius of beam and power intensity, requires in input total power and standard deviation. Based on Equations (29) and (30), these values are determined to be 1231 W and 1.4 mm, respectively. Figure 2.13 displays the temperature distribution across the entire surface at 1 s, while Figure 2.14 focuses specifically on the welding line (mid-line). The results show a high degree of

Laser parameter	value
velocity [mm/s]	80
radius of beam [mm]	2
power intensity [W/mm^2]	100
start time [s]	0
end time [s]	1

Table 2.1: Generic laser parameters used for the validation analysis

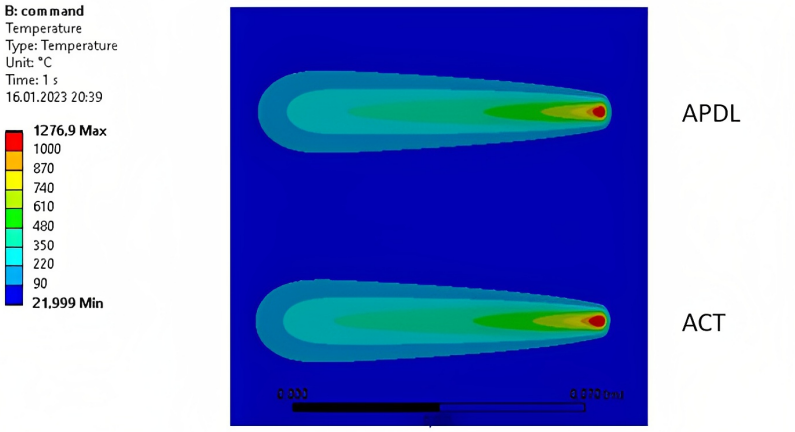


Figure 2.13: Contour plot of the top-surface temperature distribution at 1s

similarity between the two approaches, indicating that the new formulation is a valid technique for simulating moving heat fluxes. It can be utilized as a substitute for the ACT extension in more complex scenarios or for optimization purposes.

To ensure the correct application of input power to the model, an energy balance check was performed. A single welding spot was simulated using the full model, and the simulation continued until the temperature became uniform across the plate. In this particular case, there were no power losses, and the material properties were assumed to be constant and evaluated at room temperature. The input energy

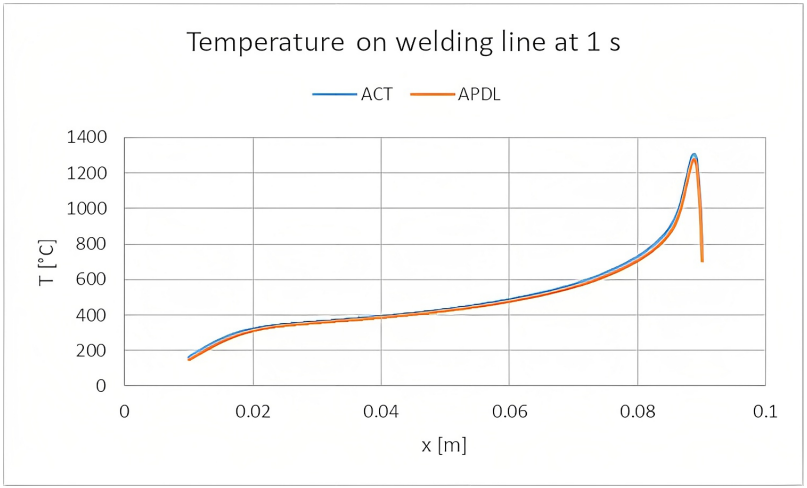


Figure 2.14: Plot of the mid-line temperature distribution at 1s

applied to the model had to match the increment of internal energy in the plate. After 1000 s, the average temperature reached 22.794°C. The total energy increment in the plate was calculated as follows:

$$\Delta E = \rho V c \Delta T = 7850 \cdot 0.1 \cdot 0.1 \cdot 0.001 \cdot 434 \cdot (22.794 - 22)$$
$$= 27.050 \text{ J/spot}$$

The total input energy, considering for this case a power (P) of 3.6 kW, a pulse duration (t) of 9 ms, and an efficiency (η) of 83.5% is equal to,

$$Q = P \cdot t \cdot \eta = 3600 \cdot 0.009 \cdot 0.835 = 27.054 \text{ J/spot}$$

The energy balance is perfectly satisfied.

2.8 Full Model

The FEM model presented follows a two-phase approach, consisting of a transient thermal simulation and a subsequent transient quasi-static mechanical simulation (Figure 2.15).
The thermal simulation serves as the initial phase, where the input corresponds to the laser power source, varying with both position and time. The output of this thermal simulation is the temperature field, which varies over time. The output obtained from the thermal simulation serves as the input for the subsequent mechanical simulation. In this phase, the mechanical behavior of the structure is analyzed under the influence of the temperature field. The primary output of the mechanical simulation is the residual deformations, specifically referred to as out-of-plane distortion in this particular case.
In the welding process, the melting temperature is considered to be the critical bonding temperature. When this temperature is reached, a bonded contact is established at the interface between the layers, and it remains active even as the temperature decreases. To ensure accurate welding and maintain the position of the parts, a magnetic field is generated during the process. This magnetic field serves the purpose of holding the parts in place and ensuring precise welding accuracy. Once the welding is completed, the magnetic field is deactivated, allowing the parts to be released. As a result, residual deformations become visible in the structure. To simulate this condition in the computational model, a contact interface is defined. During the welding phase, the contact is specified with frictional no separation properties, meaning there is no separation between the layers. However, during the cooling down phase, simple frictional properties are assigned to the contact interface, with a frictional coefficient of 0.2.
The validation of the full thermo-mechanical model was

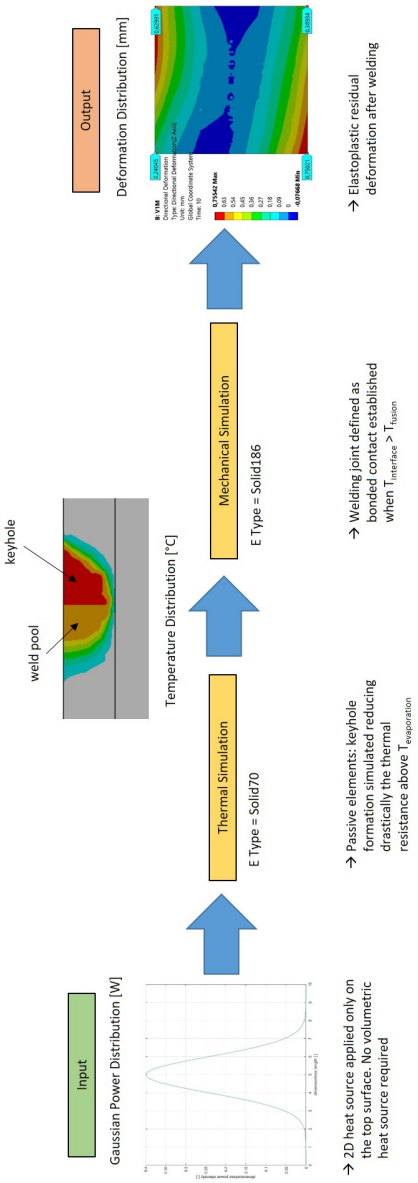
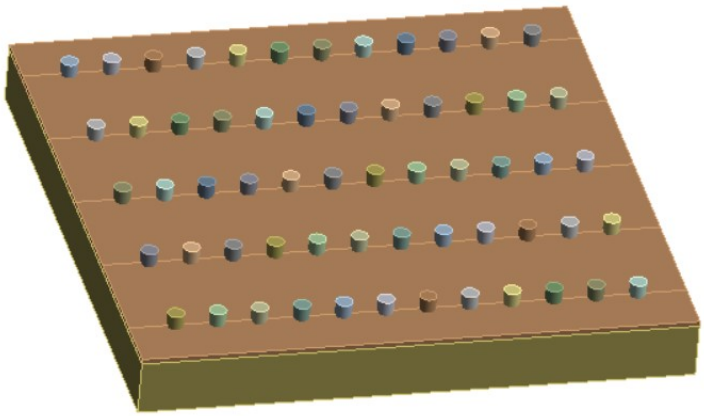


Figure 2.15: Schematic representation of the simulation flow and main characteristics and innovations of the model

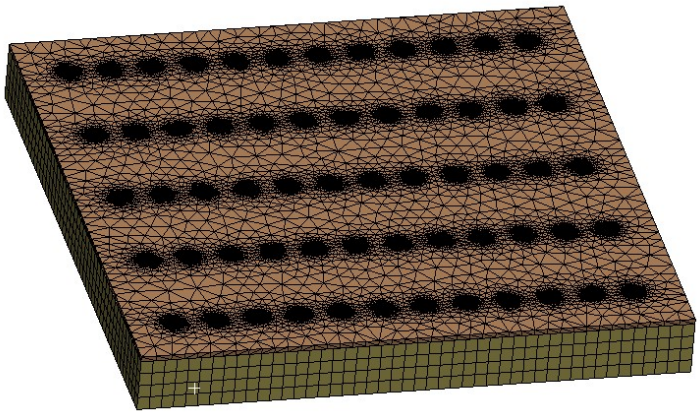
conducted using a test model comprising two square plates made of DC04 steel. Each plate had an edge length of 100 mm and a thickness of 0.5 mm. In the thermal simulation, the two welded plates were represented as a single layer. Additionally, a 10-mm thick steel plate was bonded at the bottom of the model solely for the purpose of simulating conductive thermal dissipation through the support. On the opposite side, convection heat transfer with a coefficient of 10 W/m^2K was applied. Towards the end of the welding process, this coefficient was increased by a factor of 4 to facilitate the cooling of the parts to room temperature. For the thin plates, the mesh was refined locally around the regions where pulsing power was applied (Figure 2.16).

2.9 Validation of Full Model

The described model was subjected to multiple simulations and experimental validations. The experiments were performed using a Trumpf Trulaser Cell 3000 laser machine, known for its dual-phase laser beam capability. This means that the total power can be distributed between two concentric areas: an outer area with a radius of 0.2 mm and an inner area with a radius of 0.05 mm. In this project, a power ratio of 70% outside and 30% inside was selected, and the Gaussian profile was adjusted accordingly to accommodate both power sources (Figure 2.17).
It is important to avoid excessive power concentrations in small areas to prevent instability in keyhole formation and ensure predictable welding performance. Residual deformations were accurately measured using a 3-D scanner after welding and compared with the simulation results. Various variables were considered in the model, including frequency, power, pulse duration, laser arm speed, and welding path. Some of these data were used to calibrate the model and estimate the overall process efficiency. The remaining data



(a) 2-layers configuration and dummy cylinders



(b) Mesh with local refinement

Figure 2.16: Geometry (a) and mesh (b) of the 3-D test model

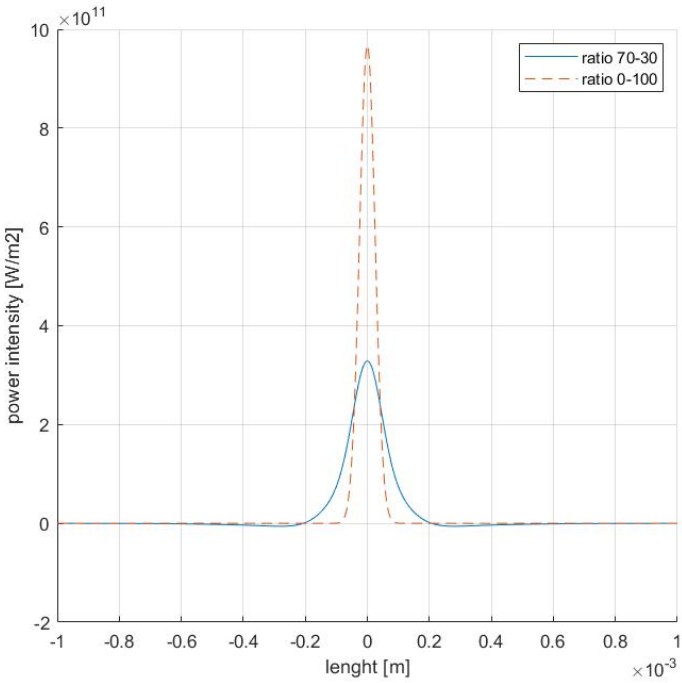


Figure 2.17: Comparison between ratio 70–30 and ratio 0–100 (entire energy concentrated in the inner circle) with P = 3800 W. The two distributions have the same underlying volume in the 3-D space

were utilized for model validation. The calibration process yielded an average efficiency of 83.5%. The final validation involved simulating the welding process with four different powers (3800 W, 3600 W, 3400 W, and 3200 W). For each configuration, the final deformation at the four corners relative to the center was evaluated and compared to experimental data. The other parameters were set as follows: pulse duration of 9 ms, frequency of 33.15 Hz, speed of 0.25 m/s, and a welding path consisting of five horizontal lines spaced at 7 mm from the left and right edges, 10 mm from the top and bottom edges, and 20 mm evenly spaced. DC04 steel plates were used for the experimental tests. The thermal and mechanical material properties were defined in the FEM model as specified in Table 2.2 and Figure 2.18.

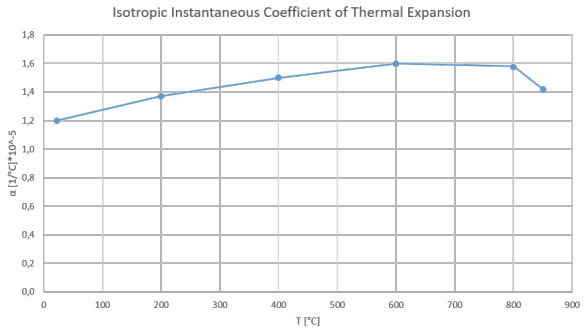
	Solid	**Liquid**	**Vapour**
T [K]	< 1798		> 3134
k_1 [W/mK]	60.5	33	0
k_2 [W/mK]	60.5	33	10^{11}
c [J/kgK]	434	573	573
ρ [kg/m^3]	7850	7287	7287
ΔH_{melt} [J/m^3]	1.85×10^9	–	
ΔH_{vap} [J/m^3]	–	4.44×10^{10}	

Table 2.2: Thermal Properties of steel DC04

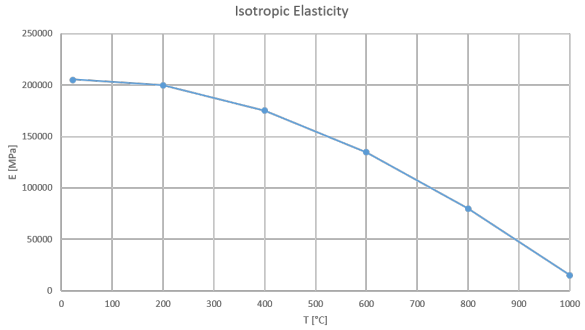
The results of the validation are visualized in Figure 2.19 and summarized in Table 2.3.
The average difference between the simulated and experimental results is 0.14 mm, or 0.08 mm when considering the positive or negative sign. With the exception of one corner corresponding to the initial spot, the simulation yields good results. Further investigations are being conducted to understand and address this issue. Figure 2.20 provides three snapshots of the thermal simulation.

		Configuration			
		3800W	**3600W**	**3400W**	**3200W**
Corner 1	Sim [mm]	0.75	0.71	0.68	0.62
	Exp [mm]	0.31	0.33	0.23	0.33
	Δd [mm]	0.44	0.38	0.45	0.29
Corner 2	Sim [mm]	0.39	0.37	0.37	0.42
	Exp [mm]	0.39	0.42	0.4	0.52
	Δd [mm]	0.00	-0.05	-0.03	-0.10
Corner 3	Sim [mm]	0.63	0.54	0.6	0.57
	Exp [mm]	0.55	0.49	0.62	0.84
	Δd [mm]	0.08	0.05	-0.02	-0.27
Corner 4	Sim [mm]	0.24	0.22	0.33	0.37
	Exp [mm]	0.24	0.28	0.28	0.36
	Δd [mm]	0.00	-0.06	0.05	0.01

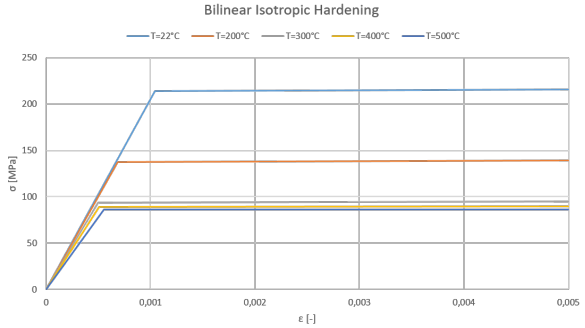
Table 2.3: Validation results: for every configuration is reported the simulated, the measured, and the differential deformation in z-direction (out-of-plane) using the plate center as reference



(a)

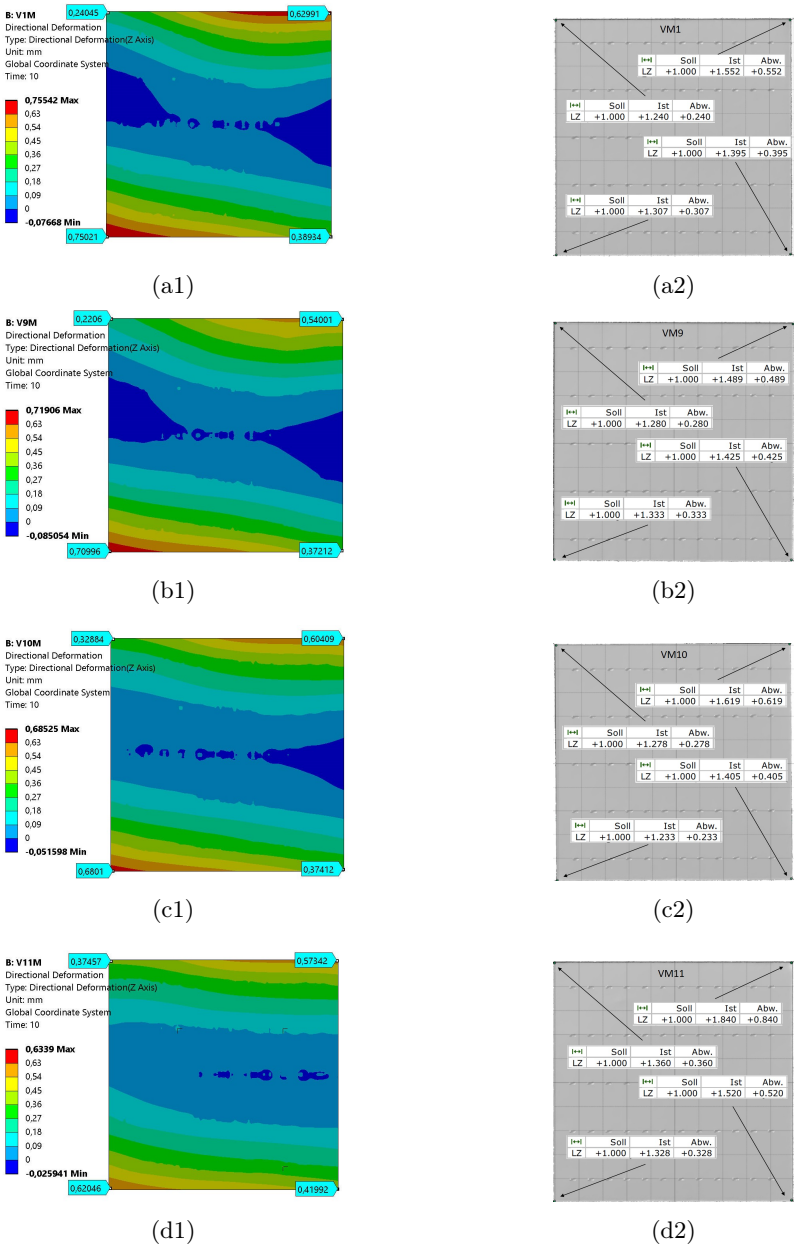


(b)



(c)

Figure 2.18: Mechanical properties of steel DC04 function of temperature: isotropic instantaneous coefficient of thermal expansion (a), isotropic elasticity (b), and bilinear isotropic hardening (c)



(a1) (a2)

(b1) (b2)

(c1) (c2)

(d1) (d2)

Figure 2.19: Comparison between simulation (1) and experimental (2) results at different powers: 3800 W (a), 3600 W (b), 3400 W (c), and 3200 W (d)

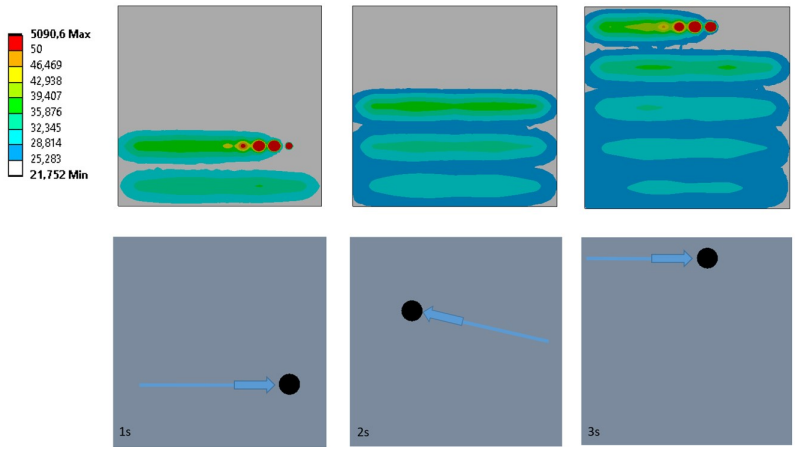


Figure 2.20: Temperature distribution in $°C$ at different time steps and corresponding position and direction of laser beam, indicated by a black spot and an arrow, respectively. Horizontal lines are the welding lines, along the oblique lines the laser beam moves from the end of a welding line to the next one

Chapter 3

Multifidelity Bayesian Optimization

3.1 Introduction to Metamodeling

Metamodeling, also known as surrogate modeling or response surface modeling, is a technique in machine learning that involves building a simplified model, or metamodel, to approximate the behavior of a complex, computationally expensive, or time-consuming system. In many real-world scenarios, the process of acquiring data from the target system can be resource-intensive, costly, or time-consuming. This could be due to various factors such as physical experiments, simulations, or expensive computations. Metamodeling addresses this challenge by creating a surrogate model that captures the key features and relationships of the target system, enabling faster and more cost-effective analysis and decision-making. The metamodel is typically built using a smaller set of training data, often obtained through a limited number of observations or simulations of the target system. The data consists of input-output pairs, where the inputs represent the system's variables or parameters, and the outputs represent the corresponding responses or outcomes. The metamodel then approximates

the mapping between the inputs and outputs, allowing for efficient evaluation and prediction without the need for extensive computations or simulations.
One of the primary benefits of metamodeling is its ability to accelerate the analysis and optimization of complex systems. By replacing the computationally expensive target system with a fast and simplified surrogate model, metamodeling enables rapid exploration of different scenarios, sensitivity analysis, and optimization algorithms. This is particularly valuable in engineering, manufacturing, and scientific domains where the underlying processes may involve intricate physics, or complex simulations. Metamodeling techniques can be broadly classified into two main categories: global and local metamodels. Global metamodels aim to approximate the target system's behavior across the entire input space, while local metamodels focus on specific regions of interest or areas where accurate predictions are crucial. Commonly used metamodeling techniques include polynomial regression, kriging (Gaussian process regression), radial basis functions, support vector regression, and artificial neural networks. Each technique has its own assumptions, advantages, and limitations, making it important to select an appropriate metamodeling approach based on the specific characteristics of the target system and the available training data.

3.2 Bayesian Optimization

Bayesian optimization is a technique for optimizing black-box functions that are expensive to evaluate. It combines the principles of Bayesian inference and optimization to efficiently explore and exploit the search space, ultimately finding the global optimum with a limited number of function evaluations. At its core, Bayesian optimization lever-

ages a probabilistic surrogate model, typically a Gaussian process, to model the unknown objective function. This surrogate model provides estimates of the objective function and its associated uncertainty at any point in the input space. By iteratively selecting the most promising points to evaluate based on an acquisition function, Bayesian optimization intelligently explores the search space, striking a balance between exploration and exploitation. The acquisition function guides the selection of the next point to evaluate by considering both the predicted function values and their uncertainties. It quantifies the potential improvement over the current best value and aims to strike a balance between exploring uncertain regions and exploiting promising areas. By sequentially updating the surrogate model with new evaluations and refining the acquisition function, Bayesian optimization progressively improves the estimation of the objective function and focuses on promising regions. This enables the efficient search for the global optimum, even in cases where the objective function is noisy, non-convex, or lacks gradient information.

3.3 Acquisition Function

In a Bayesian optimization problem, new samples are iteratively added based on specific criteria defined by the acquisition function (AQF). The AQF aims to balance the exploration and exploitation concepts in an efficient manner. Exploration involves selecting points with high posterior variance, while exploitation focuses on points with low mean values (for minimization problems). The literature on AQFs is extensive, and for this study, the Expected Improvement ($\mathbb{E}$I) approach proposed by Mockus [28] and later by Jones [23] has been implemented due to its robustness and simplicity. The $\mathbb{E}$I approach is similar to the

Probability of Improvement (PI) but considers the expectation instead. Computing the expectation involves solving an integral, but this computational task is not overly expensive because, when working with a Gaussian Process (GP), it can be done analytically. To compute the expected improvement, we define the improvement as,

$$I(x) = max(f^* - Y, 0). \tag{31}$$

Here, f^* represents the current best value, and $Y \sim \mathcal{N}(\mu, \sigma^2)$ is the predicted value of the objective function at point x based on the GP surrogate model. Introducing now the expectation,

$$\mathbb{E}\mathrm{I(x)} = \int_{-\infty}^{\infty} I(x)\phi(\epsilon)d\epsilon, \tag{32}$$

$$\mathbb{E}\mathrm{I(x)} = \int_{-\infty}^{(f^*-\mu)/\sigma} (f^* - \mu - \sigma\epsilon)\phi(\epsilon)d\epsilon,$$

$$\mathbb{E}\mathrm{I(x)} = (f^* - \mu)\Phi\left(\frac{f^* - \mu}{\sigma}\right) - \sigma\int_{-\infty}^{(f^*-\mu)/\sigma} \epsilon\phi(\epsilon)d\epsilon,$$

$$\mathbb{E}\mathrm{I(x)} = (f^* - \mu)\Phi\left(\frac{f^* - \mu}{\sigma}\right) + \frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{(f^*-\mu)/\sigma} (-\epsilon)e^{-\epsilon^2/2}d\epsilon,$$

$$\mathbb{E}\mathrm{I(x)} = (f^* - \mu)\Phi\left(\frac{f^* - \mu}{\sigma}\right) + \sigma\phi\left(\frac{f^* - \mu}{\sigma}\right), \tag{33}$$

where ϕ and Φ are the PDF and CDF of standard normal distribution, respectively. The probability density function, denoted as PDF, is a function that describes the relative likelihood of a continuous random variable taking on a specific value. For a continuous random variable, the PDF represents the probability of the variable falling within a particular range of values. In other words, it provides a probability distribution over the possible values of the random variable. The cumulative distribution function, denoted as CDF, is a function that gives the probability that a random variable takes on a value less than or equal to a specified value.

3.4 Gaussian Process Regression

Choosing the GP as a technique to generate a metamodel offers several advantages over other approaches. Here are some reasons why the Gaussian process is often preferred:

- Flexibility and Nonlinearity: Gaussian processes can model complex and nonlinear relationships between input variables and output responses. They are capable of capturing intricate paths, variations, and interactions in the data without imposing strict assumptions about the functional form of the underlying relationship.
- Uncertainty Quantification: GPs provide a principled way to quantify uncertainty in predictions. Along with estimating the mean response, GPs produce probabilistic predictions by computing the covariance or uncertainty associated with the predictions. This allows for a more comprehensive understanding of the model's confidence and enables robust decision-making, especially in scenarios with limited data or noisy observations.

- Bayesian Framework: GPs are inherently Bayesian models, which means they naturally incorporate prior knowledge and can be updated with new data. This is particularly valuable in the context of metamodeling, as it allows for sequential learning and refinement of the surrogate model as more observations become available. The Bayesian framework also facilitates the propagation of uncertainty from the surrogate model to downstream analyses, such as optimization or uncertainty-based decision-making.

- Interpolation and Extrapolation: GPs excel at interpolation, meaning they can accurately estimate responses for input configurations within the range of observed data. Additionally, GPs can provide reasonable predictions for extrapolation, meaning they can extend the surrogate model's predictions to regions of the input space where no observations are available. This makes GPs useful for exploring and analyzing the response surface beyond the sampled data points.

- Adaptive Sampling: GPs enable adaptive sampling, where new evaluations are strategically chosen based on the model's predictions and uncertainty estimates. Through active learning or acquisition functions, GPs guide the selection of informative points, effectively reducing the number of expensive evaluations needed to build an accurate metamodel. This makes GPs efficient in optimizing computationally expensive or time-consuming objective functions.

- Interpretable Covariance Structure: GPs utilize covariance functions, also known as kernels, to model the relationships between input variables. These kernels encode prior assumptions about the smoothness, periodicity, or other properties of the response surface. By selecting appropriate covariance functions, one can incorporate prior knowledge or domain expertise into

the metamodel, allowing for interpretability and insights into the underlying system.

While other techniques such as polynomial regression, neural networks, or radial basis functions also have their merits, Gaussian processes offer a unique combination of flexibility, uncertainty quantification, Bayesian framework, adaptability, and interpretability, making them a popular choice for metamodeling tasks.
Supervised learning can be categorized into two main types: regression and classification problems. In classification tasks, the outputs are discrete class labels that assign data points to specific categories or groups. On the other hand, regression focuses on predicting continuous quantities, which involve estimating numerical values rather than assigning class labels.
A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. It is completely specified by its mean function and covariance function [30]. We define mean function $m(x)$ and the covariance function $k(x, x')$ of a real process $f(x)$ as

$$m(x) = \mathbb{E}[f(x)], \tag{34}$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))], \tag{35}$$

and will write the Gaussian process as,

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')). \tag{36}$$

Usually the mean is set to be zero. The specification of the

covariance function implies a distribution over functions. In the case of noise-free observations the joint distribution of the training outputs f_T and the test outputs f_* according to the prior is,

$$\begin{bmatrix} f_T \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right). \tag{37}$$

If there are n training and n_* test points then $K(X, X_*)$ denotes the n x n_* matrix of the covariances evaluated at all pairs of training and test points, and similarly for the other entries $K(X,X)$, $K(X_*,X_*)$ and $K(X_*,X)$. To get the posterior distribution over functions we need to restrict this joint prior distribution to contain only those functions which agree with the observed data points. This operation corresponds to conditioning the joint Gaussian prior distributions on the observations to give,

$$\begin{aligned} f_*|X_*, X, f \sim \mathcal{N}(K(X_*,X)K(X,X)^{-1}f, \\ K(X_*,X_*) - K(X_*,X)K(X,X)^{-1}K(X,X_*)). \end{aligned} \tag{38}$$

In case of noisy observations, introducing the noise ϵ with variance σ_n^2, eq. 37 becomes,

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right), \tag{39}$$

with $y = f(x) + \epsilon$.

Deriving the conditional distribution corresponding to eq.

38 we arrive at the key predictive equations for Gaussian process regression,

$$f_*|X, y, X_* \sim \mathcal{N}\left(\overline{f}_*, \text{cov}(f_*)\right), \tag{40}$$

where,

$$\overline{f}_* \triangleq \mathbb{E}\left[f_*|X, y, X_*\right] = K(X_*, X)[K(X, X)+\sigma_n^2 I]^{-1} y, \tag{41}$$

$$\text{cov}(f_*) = K(X_*, X_*)-K(X_*, X)[K(X, X)+\sigma_n^2 I]^{-1} K(X, X_*), \tag{42}$$

that in compact notation becomes, for a single test point x_*,

$$\overline{f}_* = k_*^T (K + \sigma_n^2 I)^{-1} y, \tag{43}$$

$$\mathbb{V}[f_*] = k(x_*, x_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_*, \tag{44}$$

with $\mathbb{V}[f_*]$ the variance prediction for the posterior. Typically the covariance functions that are used will have some free parameters, the so-called hyperparameters. Optimizing them, for example estimating the maximum likelihood, corresponds to estimate the maximum likelihood of the observed data. Given a set of observed data points and their corresponding function values, the posterior distribution in GPs is obtained by combining the prior distribution (often assumed to be a Gaussian) with the likelihood function, which describes the probability of observing the data given the function values and noise level. The posterior distribution is proportional to the product of the prior and the

likelihood, as dictated by Bayes' theorem. Bayes' theorem is expressed as,

$$P(f_*|X, y, X_*) = \frac{P(X, y, X_*|f_*)P(f_*)}{P(X, y, X_*)}, \tag{45}$$

where $P(f_*|X, y, X_*)$ is the posterior probability, representing the updated probability of event f_* given events X, y, X_*, $P(X, y, X_*|f_*)$ is the likelihood, representing the probability of observing events X, y, X_* given event f_*, $P(f_*)$ is the prior probability, representing the initial probability of event f_* and $P(X, y, X_*)$ is the marginal likelihood or evidence, representing the probability of observing events X, y, X_*.

3.5 CoKriging

In general, high-fidelity models provide detailed and accurate representations of the system but are computationally expensive. On the other hand, low-fidelity models are computationally inexpensive but may lack the accuracy required for precise predictions. The MultiFidelity (MF) approach bridges this gap by leveraging both high- and low-fidelity models to obtain accurate results at a reduced computational cost.
The key idea of the coKriging method is to define the MF-surrogate at level l, denoted $Y^{(l)}$, as a correction of the MF-surrogate at level $l-1$. Prediction and variance become [31],

$$\bar{f}_*^{(l)}(x) = \rho_{(l-1)}\bar{f}_*^{(l-1)}(x) + k_{(l)}^T(x)\left(K_{(l)} + \sigma_{n_{(l)}}^2 I_{n_{(l)}}\right)^{-1} \cdot \cdot\Delta Y^{(l)}(\rho_{(l-1)}), \tag{46}$$

$$\mathbb{V}_{(l)}(x) = \rho_{(l-1)}^2\mathbb{V}_{(l-1)}(x) + k_{(l)}(x) + \sigma_{n_{(l)}}^2 - -k_{(l)}^T(x)\left(K_{(l)} + \sigma_{n_{(l)}}^2 I_{n_{(l)}}\right)^{-1} k_{(l)}(x). \tag{47}$$

The formulations for nested and non-nested infilling are largely analogous, with the primary distinction being the term $\Delta Y^{(l)}(\rho_{(l-1)})$ referred to as the vector of noisy observation residuals. For the nested approach,

$$\Delta Y^{(l)}(\rho_{(l-1)}) = \begin{cases} Y^{(l)} & l = 1 \\ Y^{(l)} - \rho_{(l-1)}Y^{(l-1)} & l \geq 1 \end{cases}, \tag{48}$$

while for the non-nested approach,

$$\Delta Y^{(l)}(\rho_{(l-1)}) = \begin{cases} Y^{(l)} & l = 1 \\ Y^{(l)} - \rho_{(l-1)}\bar{f}_*^{(l-1)} & l \geq 1 \end{cases}. \tag{49}$$

In the nested approach, this term depends on the observation from the preceding level. In contrast, the non-nested approach doesn't have this observation, so it relies on the prediction from the previous level.

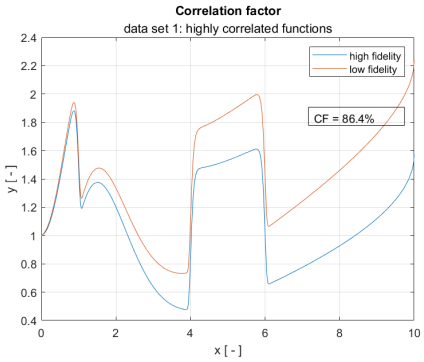
3.6 Nested vs. Non-nested Infilling Strategy

Multifidelity methods for constructing surrogate models in Bayesian optimization, which involve using evaluations at different levels of accuracy, represent a significant advancement in computational efficiency compared to traditional single fidelity strategies as demonstrated by Ariyarit [2] and Belakaria [5]. The idea of utilizing less accurate but faster data sets to gather information and train the surrogate model is a valuable resource for accelerating the optimization process. An example of multifidelity data is the use of FEM models with varying mesh refinement or contact settings, where the low-fidelity level corresponds to a coarser mesh or less stringent penetration constraints as proposed by Aruna [4]. However, when considering Gaussian Process regression emulators, it is not always the case that multifidelity cokriging outperforms single fidelity ordinary kriging. Lim [27] demonstrated that if the additional observations for the second process are located near the observations of the first process, but their number is limited and the correlation between the two processes is weak, cokriging does not improve prediction performance compared to kriging. This is especially true when the bivariate spatial processes possess an intrinsic coregionalization covariance structure. Similarly, the classic nested data sampling, despite satisfying the Markov assumption, does not necessarily outperform a non-nested structured data sampling. Sacher [31], for example, compared nested and non-nested formulations and concluded that the non-nested multi-fidelity approach is more efficient in terms of numerical cost and more robust against noisy evaluations of intermediate fidelity levels and low correlations between levels. The issue of correlation between data arises in both [27] and [31], and before delving into the project model, it is essential to determine the most suitable infilling strategy for multifidelity Bayesian

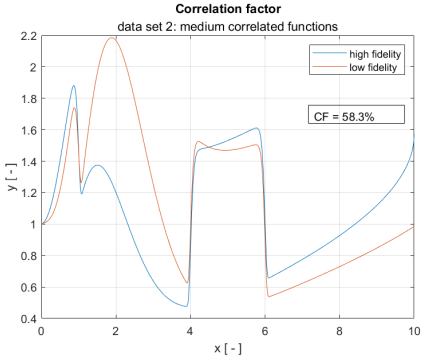
optimization in terms of convergence and computational costs [13]. This study begins with a preliminary 1-D analysis, where three different data sets are defined, each with a distinct level of correlation between high- and low-fidelity levels. These data sets are then tested using both nested and non-nested infilling strategies based on specific acquisition functions, and conclusions regarding the relationship between data correlation and infilling strategy are drawn. Once the optimal infilling strategy based on the available data is determined, the Bayesian optimization can be applied to the project model using the same settings as the 1-D analysis, including the acquisition function, exit strategy, and optimization algorithm.

3.6.1 Data Sets

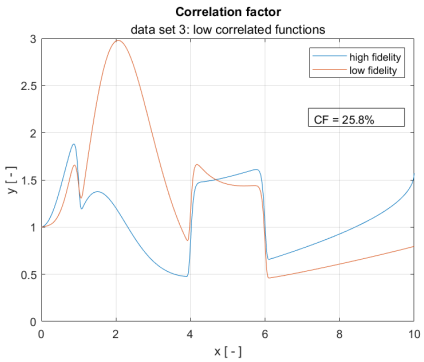
To assess the connection between the infilling strategy and data correlation, three distinct data sets are established. The initial data set comprises data exhibiting a strong correlation, with its DCF (Data Correlation Factor) approaching 90%. The second data set contains data displaying a moderate correlation, approximately 50%. Lastly, the third data set encompasses data with a DCF value of around 25%, indicating a low level of correlation. Each data set consists of (x, y) coordinates that correspond to points on 1-D functions. Furthermore, all three data sets possess two fidelity levels: high-fidelity and low-fidelity. The high-fidelity level accurately represents the true function to be minimized, employing noise-free sampling. On the other hand, the low-fidelity level differs across the data sets, generating varying correlation levels. This distinction allows for the creation of different correlation levels among the data sets (figure 3.1). In general, despite potential variations in the data's origin, all fidelity levels are established using the same input and output variables. The key distinction lies in the accuracy



(a)



(b)



(c)

Figure 3.1: Highly (a), medium (b) and low (c) correlated data sets

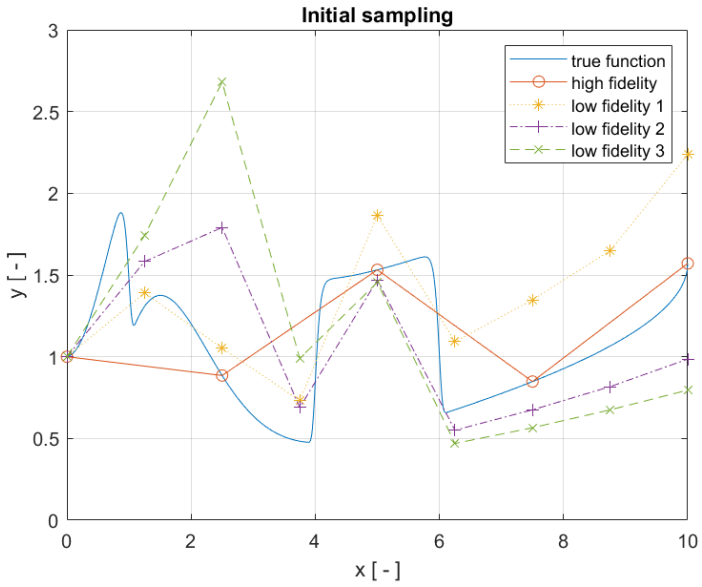
of the observed samples. For instance, if both data sets are simulations, the high-fidelity data might originate from a model with a finer mesh, whereas, if the data have different sources, the high-fidelity data could be experimental and the low-fidelity data simulations. In cases where low-fidelity data is expected to be more affordable to collect, a large number of observations can be easily amassed. Conversely, high-fidelity data is typically costly to obtain, necessitating a restriction on the number of observations. The initial data sets are structured nested, with the correlation factor based solely on the nested values present in both the high-fidelity and low-fidelity levels. Initially, the low-fidelity level comprises 9 points, evenly distributed, while the high-fidelity level consists of 5 points (figure 3.3).

It can be demonstrated that the minimum number of samples required to achieve a DCF close to the asymptotic value for this particular problem is 5. The low-fidelity function initially incorporates 9 equally spaced points to establish a nested relationship with the high-fidelity level (figure 3.2). The 1-D function is intentionally designed to be arbitrarily complex, featuring multiple local minima and steep gradients. This complexity serves the purpose of increasing the number of iterations necessary to achieve convergence and effectively exploit the impacts of various infilling strategies.

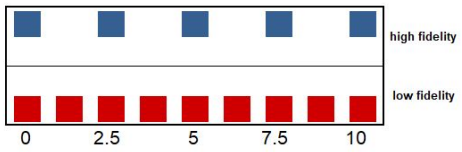
1. Highly correlated functions

   High-fidelity function:

$$Y = \arcsin\left(\frac{X}{10}\right) + X^{2-X} + \exp\left(5 - X\right)^{20} + \exp\left(-X\right)^{20}. \tag{50}$$



(a)



(b)

Figure 3.2: Initial sampling: high and low fidelity data (a) and visualization of the equidistant nested distribution (b)

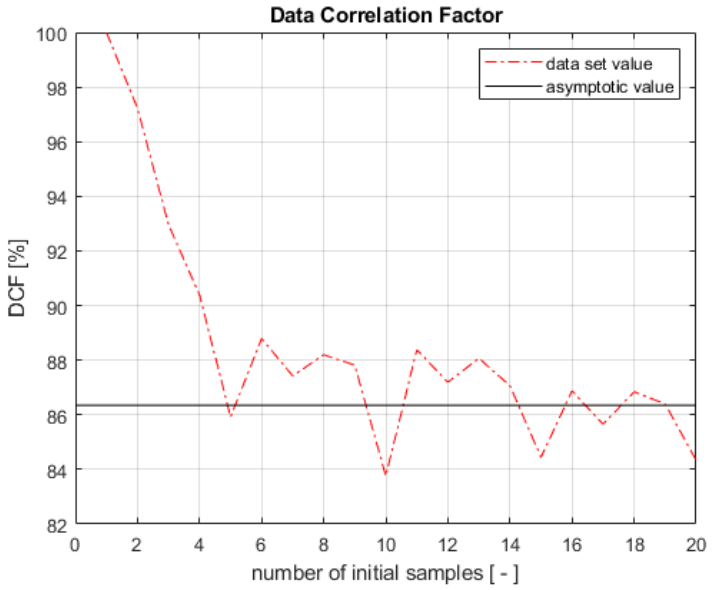


Figure 3.3: Convergence behavior of the Data Correlation Factor for the highly correlated functions

Low-fidelity function:

$$Y = \arcsin\left(\frac{X}{10}\right) + X^{2-X} + \exp\left(5 - X\right)^{20} + \exp\left(-X\right)^{20} + \frac{1}{15}X. \tag{51}$$

2. Medium correlated functions

   Low-fidelity function:

$$Y = \arcsin\left(\frac{X}{12}\right) + X^{3-X} + \exp\left(5 - X\right)^{20} + \exp\left(-X\right)^{18}. \tag{52}$$

3. Low correlated functions

   Low-fidelity function:

$$Y = \arcsin\left(\frac{X}{14}\right) + X^{\frac{7}{2}-X} + \exp\left(5 - X\right)^{20} + \exp\left(-X\right)^{16}. \tag{53}$$

3.6.2 Acquisition function and exit strategy

At each iteration of the process, the maximum Expected Improvement ($\mathbb{E}$I) is computed for both the low-fidelity ($\mathbb{E}\text{I}_L$) and high-fidelity ($\mathbb{E}\text{I}_H$) levels. The parameter X corresponding to the highest $\mathbb{E}$I value is selected, and new sam-

ples are added accordingly. If the highest $\mathbb{E}$I value is from the low-fidelity level ($\mathbb{E}\mathrm{I}_L > \mathbb{E}\mathrm{I}_H$), a new pair of points is added to the low-fidelity level by solving the known objective function at X. On the other hand, if the high-fidelity $\mathbb{E}$I value is higher than the low-fidelity $\mathbb{E}$I value ($\mathbb{E}\mathrm{I}_L < \mathbb{E}\mathrm{I}_H$), the additional sampling depends on the chosen infilling strategy. In the case of a nested approach, both the low-fidelity and high-fidelity levels are updated, whereas only the high-fidelity level is updated in the case of a non-nested strategy. The procedure is illustrated in figure 3.4. To account for the additional time required to solve the high-fidelity problem, the $\mathbb{E}$I values are weighted. In this study, a factor of 3 is used, representing the time ratio between the high-fidelity and low-fidelity models in a real-life laser welding simulation scenario. This means that the option is selected only if $\mathbb{E}\mathrm{I}_L$ is three times larger than $\mathbb{E}\mathrm{I}_H$. The iterations continue until the maximum $\mathbb{E}$I calculated for the high-fidelity level becomes smaller than 1e-10 or after a maximum of 40 iterations. Once the final surrogate model is obtained, a final optimization step is performed, this time minimizing the prediction to determine the minimum of the surrogate function.

3.6.3 Optimization Algorithm

The Adaptive Metamodel of Optimal Prognosis (AMOP) algorithm, available in Ansys Optislang, is employed as the optimizer to find the maximum Expected Improvement at each iteration and level. AMOP is an iterative metamodeling approach based on the Multi-Objective Prognosis (MOP) method proposed by Will [34]. It utilizes an adaptive refinement of data points and functions similarly to an optimizer, running a specified number of solver runs in multiple iterations. The convergence criterion for AMOP is to achieve a minimum value of the Coefficient of Prognosis for all selected responses. In comparison to gradient-

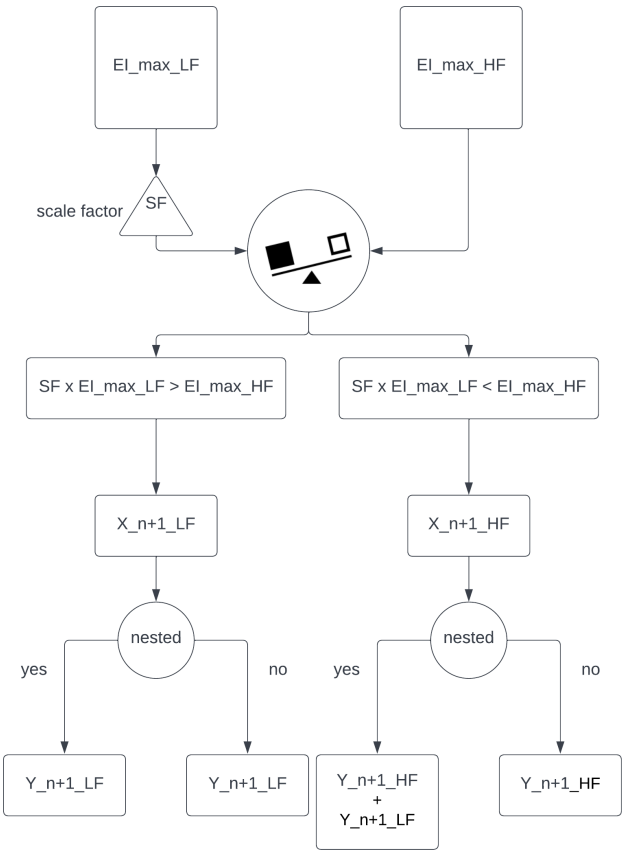
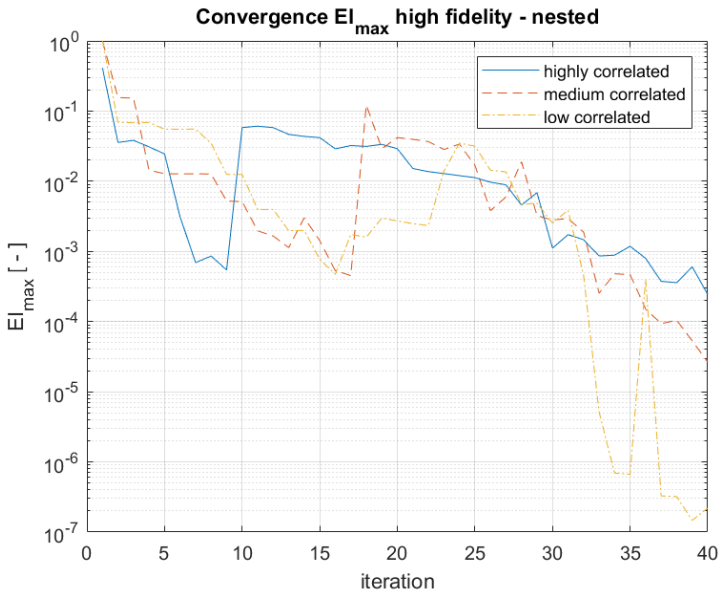


Figure 3.4: Schematization of the algorithm to select next sampling

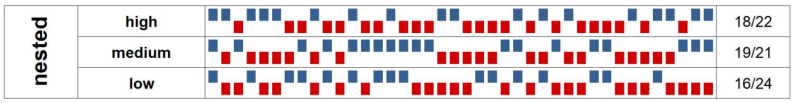
based methods such as Downhill Simplex, AMOP allows for a search throughout the entire domain, enabling exploration of the global minimum without being influenced by local minima. The sampling process is divided into two stages: a global search with 50 samples and a local refinement with 45 samples. The global search employs a plain Monte Carlo approach to ensure coverage of all areas within the domain without favoring specific locations. A total of 10 iterations are performed, with the first iteration dedicated to the global search, followed by 9 iterations of local refinements. Each local refinement stage allows a maximum of 5 samples, resulting in a total of 45 samples for refinement. To maintain a balanced distribution of samples between the local and global searches, the number of global samples is set to 50. This approach strikes a good balance between computational time and result quality for this specific case. Increasing the number of samples, both globally and locally, has shown no significant differences in results, while using fewer than 5 samples for refinement may be insufficient.

3.6.4 1-D Study

As expected, when using the nested strategy, the highly correlated data set exhibits the best convergence rate in the initial iterations. However, the behavior in the latter part of the optimization is not intuitive. All three cases experience a regression in convergence, with the highly correlated case showing a slow but steady convergence, the low correlated case demonstrating a faster but rough convergence, and the medium case falling in between (figure 3.5).
All cases exhibit a similar distribution between low- and high-fidelity infilling, with a slight preference for low-fidelity infilling (3.5). The ratio between high-fidelity sampling and total iterations is 45% for the first case, 47.5% for the second case, and 40% for the low correlated case. Consequently,



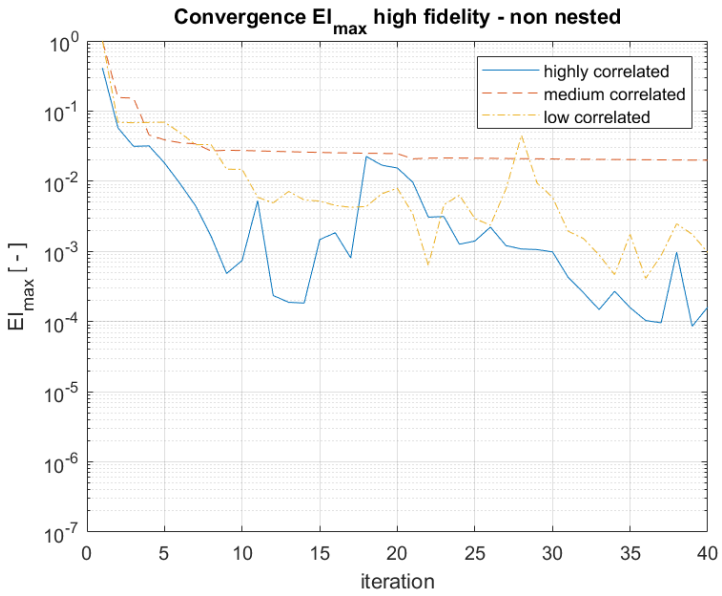
$\mathbb{E}$I convergence for the nested strategy



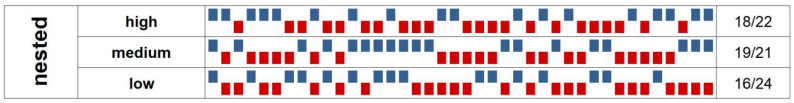
Infilling quantitative distribution for the nested strategy. Blue squares indicate high-fidelity+low-fidelity infilling and red one low-fidelity. In the last column total number of high-fidelity+low-fidelity samples and total number of low-fidelity samples are reported

Figure 3.5: Convergence and infilling sequence for the case nested

the low correlated case is the most cost-effective in terms of computation, as the number of iterations remains the same. Analyzing the results for the non-nested strategy, it is evident that the behavior is similar to the nested strategy for two data sets: both the low and highly correlated cases converge, while the medium correlated data set stagnates. This time, the highly correlated case demonstrates the best behavior in both the initial and latter parts of the optimization after the regression. The results are once again summarized in figure 3.6, similar to the way they were summarized for the nested case.
By considering +1 for an iteration where high-fidelity $\mathbb{E}$I dominates over low-fidelity $\mathbb{E}$I and -1 for the opposite, it is possible to plot the cumulative distribution of the acquisition function selection for both nested and non nested cases (figure 3.7). Although there are no significant differences among the three cases, in case of nested infilling, the highly correlated case tends to add more high-fidelity points in the initial iterations, which could partially explain the better convergence observed in this segment. In the latter part of the optimization, low-fidelity infilling dominates in all three cases (descending curve). Despite the apparent similarities in behavior, the distribution of $\mathbb{E}$I selection differs significantly in case of non-nested infilling: in this case, the number of high-fidelity samplings is consistently above 80% (82.5%, 90%, and 85% respectively), indicating a much higher computational cost.
Comparing now the two infilling strategies, nested and non-nested, for each case, the following results are observed: for the highly correlated data, the non-nested strategy performs well, even outperforming the nested approach. For the medium correlated data, the non-nested solution does not converge, and the nested approach is the only valid option. Similarly, for the low correlated data set, the performance of the two strategies is comparable, but the nested strategy exhibits a better convergence rate after a certain

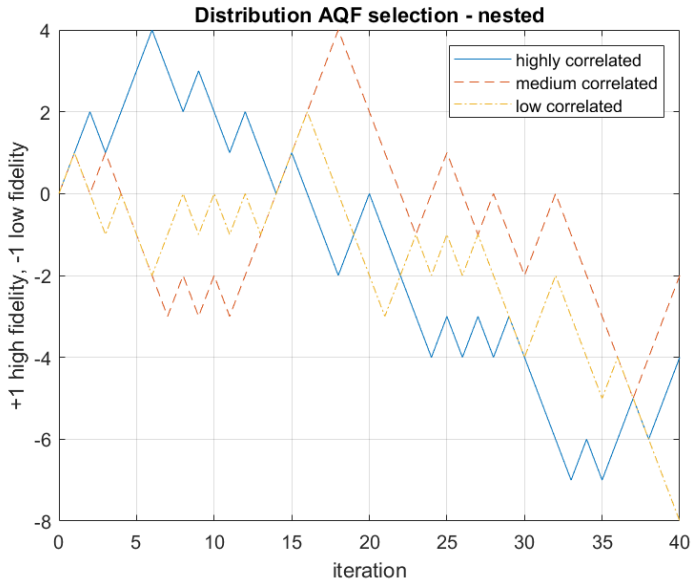


$\mathbb{E}$I convergence for the non-nested strategy

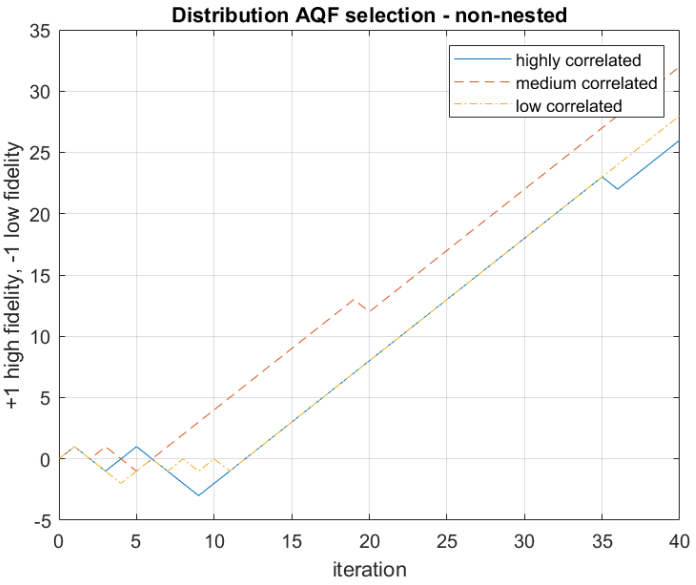


Infilling quantitative distribution for the non-nested strategy. Blue squares indicate high-fidelity infilling and red ones low-fidelity. In the last column total number of high-fidelity samples and total number of low fidelity samples are reported

Figure 3.6: Convergence and infilling sequence for the case non-nested



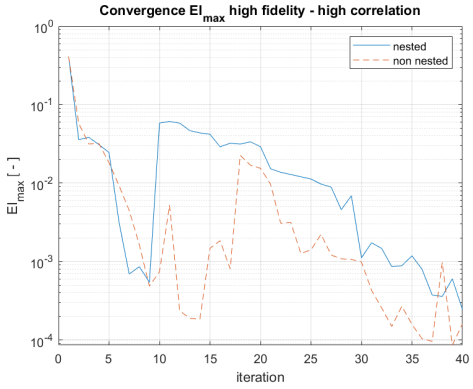
(a)



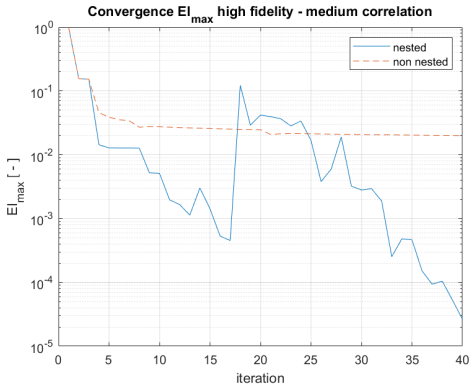
(b)

Figure 3.7: Infilling iterative distribution for the nested (a) and non-nested (b) strategy

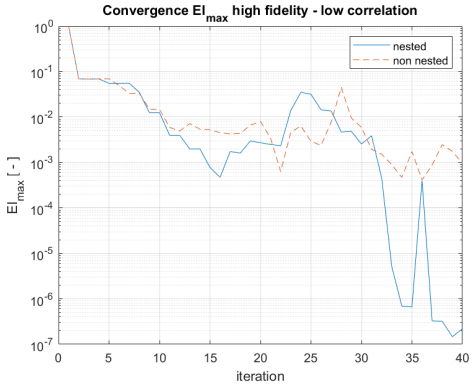
number of iterations (figure 3.8).
Although the performance for the low and highly correlated cases is similar, the cost of the two infilling strategies differs. While the non-nested strategy may seem tempting to save computational time (considering that when $\mathbb{E}$I_high-fidelity dominates over $\mathbb{E}$I_low-fidelity, only one simulation is needed instead of two), the total cost decidedly favors the nested approach. The non-nested approach could be considered for a "quick and dirty" optimization with few iterations. Assuming a unit cost of 1 hour for low-fidelity simulation and 3 hours for high-fidelity simulation based on the 3-D model, the cumulative cost is represented in figure 3.9. Finally, it is interesting to observe the distribution of samplings across the domain (figure 3.10). Both nested and non-nested strategies converge, which explains the majority of the points falling within the range of 4 (minimum at 3.88). The distribution is similar, but the nested strategy appears to be more sparse than the non-nested strategy.



(a)



(b)



(c)

Figure 3.8: $\mathbb{E}$I convergence for the highly correlated (a), medium (b) and low correlated (c) data set

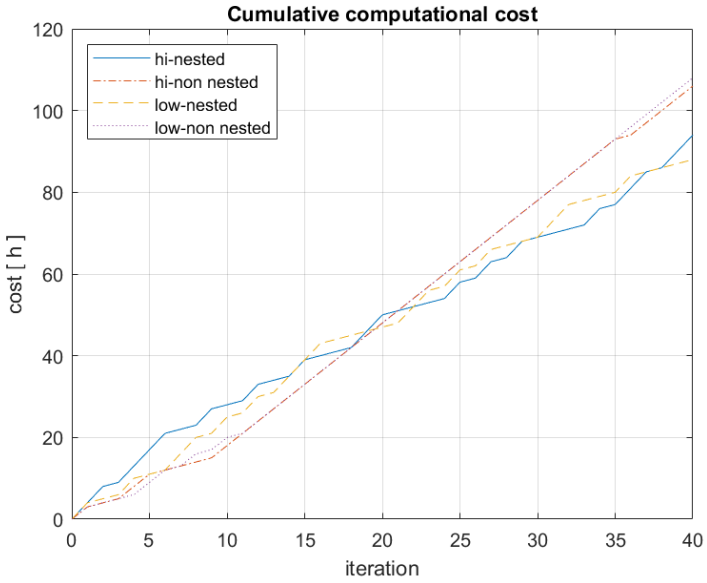
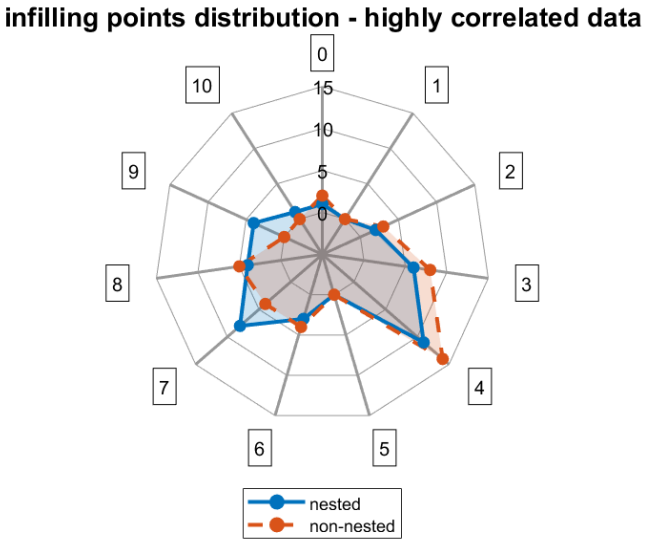
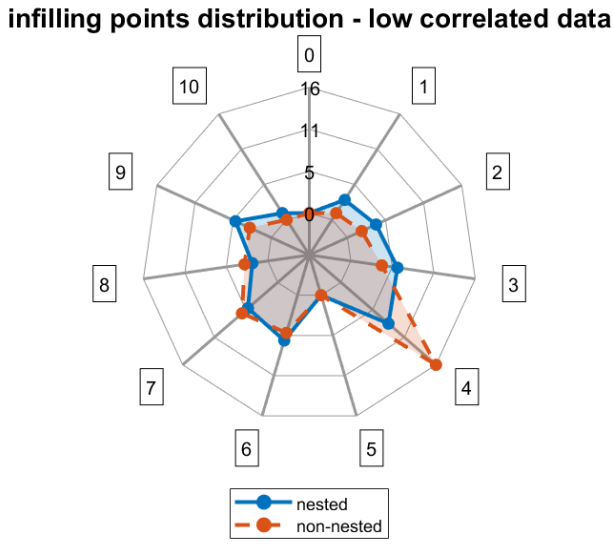


Figure 3.9: Cumulative costs based on infilling strategy and simulation time



(a)



(b)

Figure 3.10: Infilling point distribution for highly (a) and low (b) correlated data

Chapter 4

Optimization of Laser Welding Process

4.1 Workflow Overview

The optimization environment can be structured into four primary sections as underlined in figure 4.1:

1. Sampling: This involves the use of two specific Excel files, labeled as 'low fidelity' and 'high fidelity'. These files comprise tables filled with input and output values that are essential for the training of the surrogate model. The process is not static; it involves adding values in an iterative manner until either the point of convergence is reached or the maximum number of iterations is completed. The process incorporates two levels of fidelity, thereby making it a multifidelity approach.

2. Metamodeling – Training and Prediction: During this stage, the metamodel is trained with the data that has been collected. The training facilitates the generation of a prediction concerning the optimal value, which is determined based on a specific acquisition function. Here, the optimum is identified as the value that brings about the

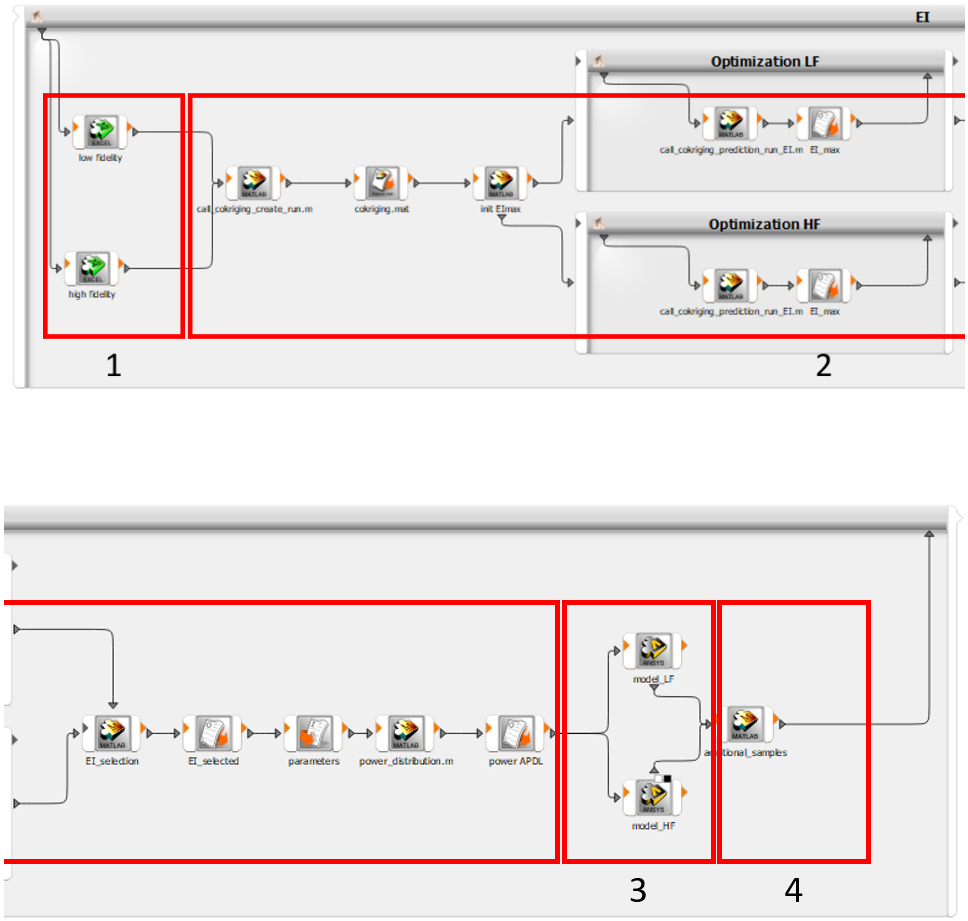


Figure 4.1: Full workflow of the optimization process in Optislang

greatest expected improvement.

3. FEM-Simulation: After determining the optimum, a new set of input data is created, and the corresponding output is evaluated using the FEM simulation model. There are two distinct models provided, each corresponding to one of the fidelity levels.

4. Bayesian Iterative Optimization: In this final phase, the newly obtained data set is incorporated into the existing training samples. This inclusion is a crucial part of the process, allowing for the continuous evolution and refinement of the optimization cycle.

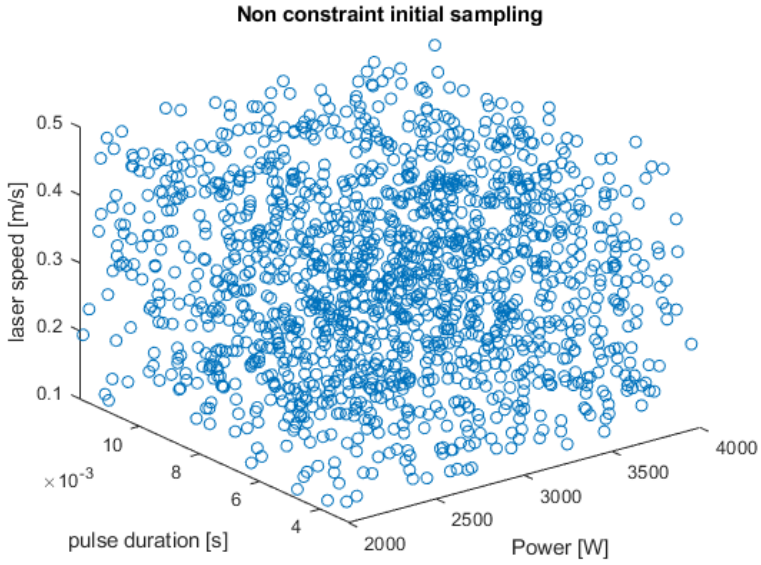
4.2 3-D Model Optimization

The project model involves a 3-D thermo-mechanical elasto-plastic FEM model for simulating a laser welding process as described in [12]. It comprises two fidelity levels, with the lower level featuring a coarser mesh for both thermal and mechanical aspects compared to the higher level. The objective function to be minimized is the residual deformation, while the laser power, pulse duration, frequency, and laser speed are the parameters to be optimized. Although surrogate models have been extensively used in manufacturing, including the laser welding field, such as the work by Heydari [18] who simulated the laser welding process of Ti6Al4V alloy steels and investigated the effects of welding speed, laser power, and pulse duration using a response surface based on a fixed series of DOEs using a Central Composite Design (CCD) matrix, and the study by Ai [1] who performed a complete Efficient Global Optimization (EGO) by simulating a fiber laser welding (FLW) process of dissimilar materials (low carbon and stainless steel) and

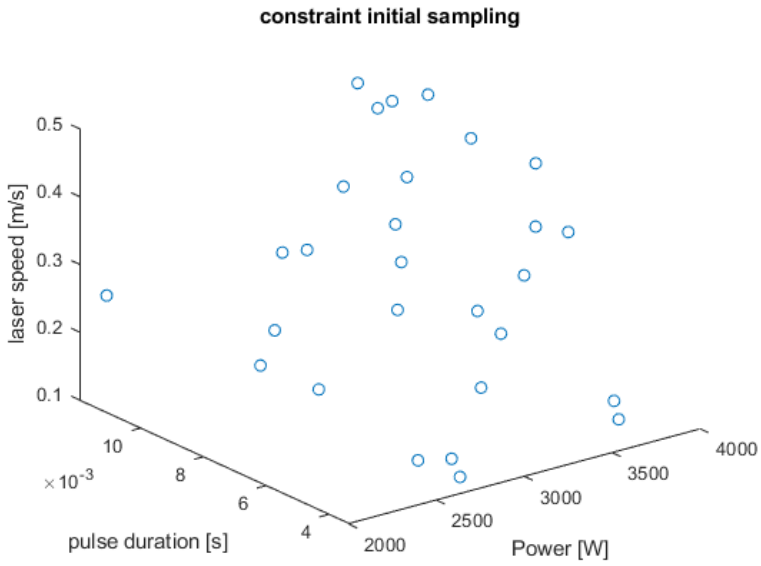
implemented optimization techniques combining a genetic algorithm optimized backpropagation neural network (GA-BPNN) and Particle Swarm Optimization (PSO) to maximize weld bead integrity and minimize the weld area, the novelty of this work lies in the application of a multifidelity approach based solely on simulations.
A common approach to define the initial DOEs involves using Latin hypercube sampling. This method ensures uniformly distributed samples across the entire domain. However, in this case, the inputs are not independent but linked through the equation:

$$\frac{Ppuf}{v} = constant \quad [\mathrm{J/m}]. \tag{54}$$

Here, P represents power [W], pu is pulse duration [s], f denotes pulse frequency [Hz], and v stands for laser speed [m/s]. This equation establishes that the energy introduced per unit length must remain constant to ensure solid bonding and prevent excessive or insufficient penetration. Consequently, the initial sampling is divided into two steps (figure 4.2): firstly, a Latin hypercube sampling across the entire domain for numerous combinations, including both feasible and non-feasible ones, and thus different values for the constraint equation, and secondly, filtering the candidates to extract a feasible subsystem by adding the mentioned constraint. The experimentally determined constant is 4500 J/m. To this value, a tolerance of $\pm$ 100 J/m is added, which allows for more feasible points by relaxing the constraint. When plotting parameters such as laser speed versus frequency, a positive proportionality becomes evident. Conversely, plotting pulse duration versus frequency reveals a negative proportionality (figure 4.3). There are two fidelity levels, with the higher level featuring a finer mesh. The initial DOEs consist of 28 points for the low-fidelity

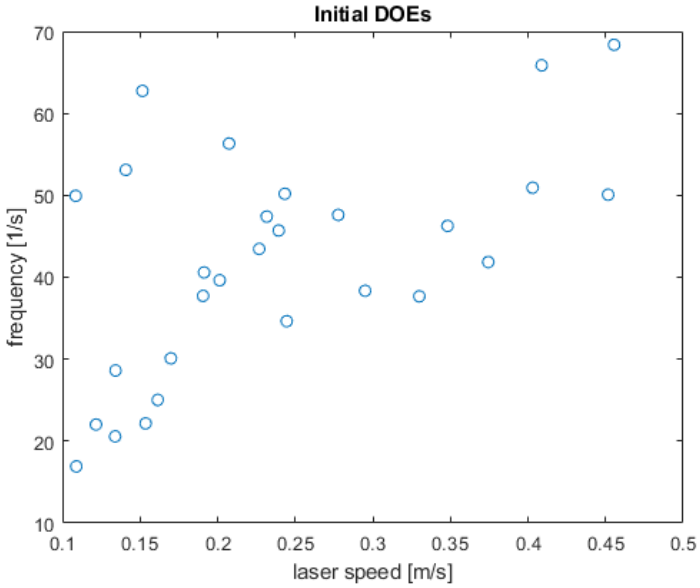


(a)

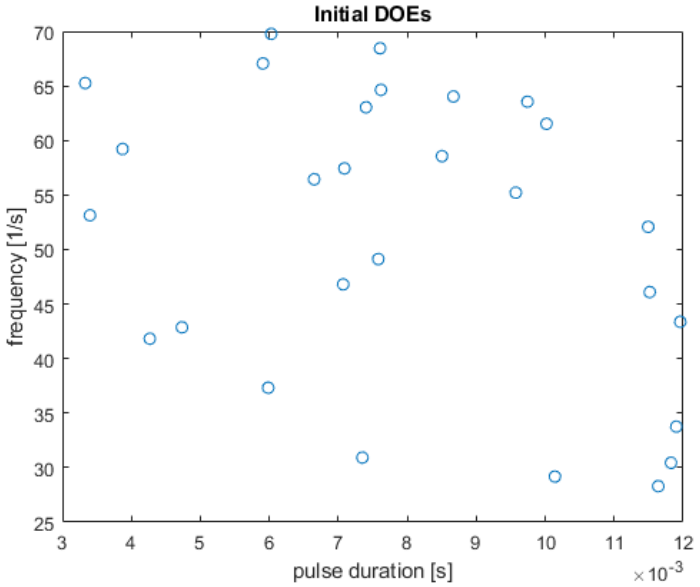


(b)

Figure 4.2: Latin hypercube sampling in the entire domain (a) and feasible configurations respecting the constraint (b)



(a)



(b)

Figure 4.3: Initial DOEs: proportional (a) and inversely proportional (b) correlation of the parameters

level and 6 nested points for the high-fidelity level, achieving an initial correlation coefficient of 0.6. The 6 points of the high-fidelity level are selected from the 28 low-fidelity points to satisfy the nested condition. These points encompass the upper and lower boundaries of the parameters, ensuring coverage across the entire domain. Among these points, 5 have at least one parameter with the highest or lowest value, while only one point is chosen with average values. Upper and lower boundaries are reported in the table 4.1:

	min	**max**
P [W]	2000	4000
pu [s]	0.003	0.012
f [Hz]	20	70
v [m/s]	0.1	0.4

Table 4.1: Upper and lower boundaries of the parameters based on experience and on the machine characteristics

The surrogate model is created using a proprietary tool based on Le Gratiet work [16]. It consists of two main subroutines: cokriging_create, responsible for metamodel creation, and cokriging_prediction, used for deformation prediction with the desired parameter configuration. The key properties are defined in table 4.2. For more information about the Gaussian Process employed to train the model, please refer to the original study [17] and the in-house implemented modifications [24] and [25].
Building upon the results of the 1-D analysis, a Bayesian optimization with nested infilling strategy is employed. The same AQF and optimizer from the 1-D analysis are implemented for the 3-D analysis. The model consists of two plates measuring 100 x 20 x 0.5 mm, which need to be joined together. In the transient thermal simulation,

(a)

(b)

Figure 4.4: Results of the laser welding simulation: initial (a) and optimized (b) status

the laser power source moves in the y-direction. Subsequently, in the mechanical simulation, the transient temperature field is imported, and the plates are fixed on one side in y- and z-direction with a remote displacement to allow shrinking in the x-direction. The other side of the plates remains free, and the maximum displacement is observed in the z-direction. With the current parameters, the maximum deformation, validated through experimental tests, amounts to 1.07 mm. However, with the optimized parameters, the residual deformation reduces to 0.39 mm, resulting in a 64% improvement as depicted in figure 4.4. The surrogate model provides a mean residual deformation of 0.44 mm for the same configuration, with a variance of 0.004 mm^2. As expected, the metamodel does not precisely match the simulation results but provides a good

approximation. More importantly, it efficiently helps identify the global minimum. To achieve even greater accuracy, a direct optimization of the simulation model around the metamodel's optimum could be conducted. However, for the present project, this step was not required. The actual and new parameters are highlighted in table 4.3. The new configuration prioritizes frequency and pulse duration over power, while the higher speed leads to a faster welding process.

nrestart	8	number of restart for hyperparameters optimization
kernel	matern52	kernel function
nugget	1.0e-7	noise factor
nuggetparam	fixed	the nugget term is fixed by the user
optimizer	fmincon	optimizer for hyperparameters optimization
nsamples hf	6	number of initial samples for the high fidelity level
nsamples lf	28	number of initial samples for the low fidelity level

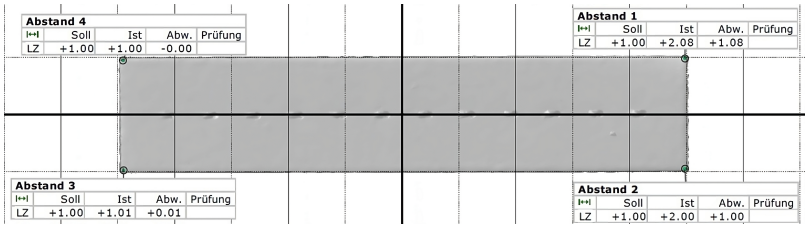
Table 4.2: Main properties of the Gaussian regression model

	actual	**new**
P [W]	3800	2412
pu [s]	9	11
f [Hz]	33.15	62.5
v [m/min]	15	21.6

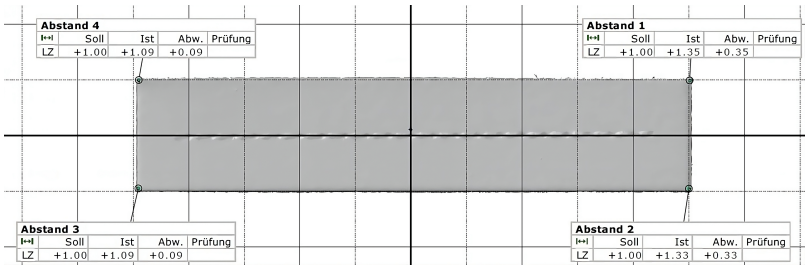
Table 4.3: Initial and optimized parameters

4.3 Experimental Validation

Although the 3-D model had previously undergone experimental validation, further testing was conducted using the new set of parameters. The boundary conditions, identical to those used in the optimization simulation, were duplicated. In this phase, two DC04 steel plates, each measuring 100x20x0.5 mm, were bonded, initially using the previously established settings and subsequently with the newly optimized parameter configuration (figure 4.5). The results were telling: with the default settings, the residual deformation measured 1.04 mm, but with the optimized parameters, it was significantly reduced to 0.43 mm. This outcome not only reaffirms the model's reliability but also highlights the enhancement in welding effectiveness.



(a)



(b)

Figure 4.5: 3-D scanned plates joint: initial (a) and optimized (b) welding path

Chapter 5

Conclusions

5.1 Time Cost Savings

In order to estimate the computational cost efficiency of the adopted method, a comparative analysis was conducted against an ideal approach that solely relies on simulations, thus not utilizing metamodelling or iterative optimum seeking. Moreover, it is assumed that only the high fidelity model is used, as direct use of the low fidelity model to obtain quantitative information on residual deformation would inevitably lead to incorrect results, despite its utility in finding the optimum in a multifidelity system. The computational advantage of this method has been estimated compared to a series of predefined simulations, rather than compared to an intermediate approach based on direct optimization of the FEM model. This is primarily to account for the worst-case scenario where neither optimization techniques nor surrogate models are employed. This choice is also motivated by the fact that applying deterministic optimization algorithms, such as the simplex method, does not guarantee finding the global optimum. The simplex method, like other deterministic methods based on gradient or Hessian matrix calculations, although suitable for problems with constraints, is a linear method capable of

quickly converging with convex functions. For more complex problems or in the case of "black box" scenarios, it is necessary to initially perform a global optimization to find the region of the optimum, followed by local refinement to accurately compute the sought maximum or minimum.
Therefore, a substantial series of simulations was envisioned, covering the entire allowed domain. Considering that only 3 out of the 4 variables are actually independent and that each simulation lasts on average 3 hours, and assuming that at least 5 configurations are desired for each input to achieve good accuracy, a total of 375 hours is required to complete the analysis. This computational cost is obtained by multiplying the time required for a single simulation, 3 hours, by the number of possible configurations, which is 125.
With the optimization model described in this text, convergence was achieved after 17 iterations, saving 75% in time. Considering that the low fidelity simulations have an average duration of one hour and that training the metamodel and searching for the optimum does not exceed 10 minutes per iteration, the total computational cost amounts to 3.5 days, instead of 15.6 days. Figure 5.1 illustrates the computational cost distribution for the two strategies. In addition, there would be no guarantee of identifying the optimum, but only the best configuration among those tested. To find the optimum, it would be necessary to resort to some form of interpolation, and the best way to do so would again be to use metamodels, even though in this case they are based on a high-density single-step approach to data.

5.2 Current method limitations and potential future developments

As much as the method presented in this work has allowed for obtaining useful information to optimize the laser weld-

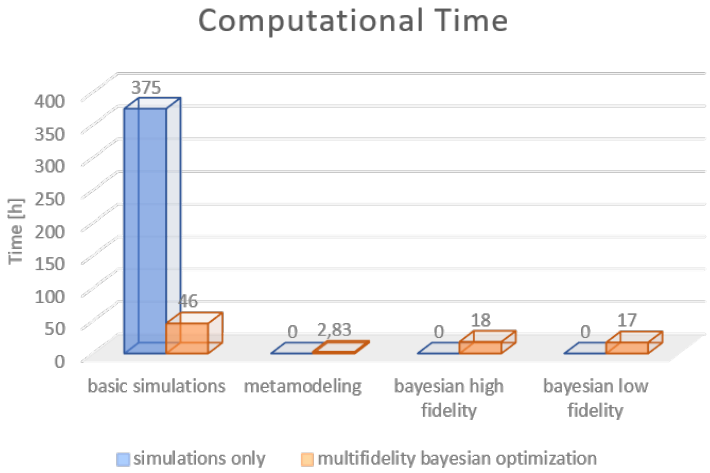


Figure 5.1: Comparison between "simulations only" approach and adopted multifidelity bayesian optimization

ing process within reasonable timeframes, it can certainly be further developed. While the accuracy of the FEM model is satisfactory on one hand, the computational efficiency does not yet seem suitable for conducting optimizations on a large scale characterized by hundreds of welding points. For this reason, the optimization focused solely on laser parameters and not on the geometric aspects related to the welding path.
A new FEM model is currently under study with the aim of drastically reducing computation times by exploiting the cyclic nature of the welding process. The method essentially involves simulating a single spot, as done previously, to highlight its residual deformation. Subsequently, a force is applied to the involved nodes to return them to their original position. Finally, utilizing the concept of submodeling, the obtained stresses are sequentially applied to all other welding spots. Once a method is available to work with large models, the implementation of the welding path variable in the optimization process can be pursued, allow-

ing for the analysis of non-linear paths as well.
Moreover, regarding the junction strength, in the experimental phase, the following observations have emerged: when applying the optimized input values, the two plates are indeed welded together, but the junction strength, i.e., the force required to separate them, has proven to be lower than the desired value, which typically corresponds to 80% of the material's tensile strength. This value is not linear, considering the time dependence of the process, and cannot be simply described by the constraint that the specific energy, expressed in [J/m], remains constant in all configurations. For example, with the same power and speed, doubling the frequency compared to a baseline configuration results in twice as many welding points, but each of them has a contact force less than half that of a single spot in the baseline configuration, leading to a lower total sum of all the points. Keeping the energy per spot constant, instead of specific energy, is not numerically feasible, as the system would inevitably tend to reduce the frequency to 0 in order to minimize deformations. Therefore, it is necessary to define a more specific correlation between the input parameters that takes into account the non-linearity of the bond force.

Appendix A

APDL Script

APDL (ANSYS Parametric Design Language) is a programming language specifically created for generating and modifying input files utilized by ANSYS simulation software. It serves the purpose of automating repetitive tasks, constructing intricate models, and conducting parametric investigations. APDL can be employed in conjunction with the ANSYS graphical user interface (GUI) or as an autonomous tool. The APDL script discussed in this article belongs to the former category and is intended for direct integration into Ansys Workbench. It functions as an additional command, serving as a boundary condition within transient thermal simulations, replacing the default heat source input. This script is divided into four sections: the initial part defines various constants that specify the dimensions of the domain, encompassing both spatial and temporal extents that encompass the mobile heat source, as well as the number of evaluation points. The combination of domain size and evaluation points establishes the resolution of the load: a smaller domain, with the number of points held constant, results in higher resolution. Similarly, resolution increases when the domain size remains consistent while the number of points increases. Nonetheless, it's important to note that the total number of evaluation points in Ansys Workbench is finite due to internal computational constraints. As a result, this method may not be suitable for exceedingly large models and domains. The required resolution naturally depends on the size and duration of the power source involved in the process. In instances where the power source has very small beam radii, such as in laser processes, a very high resolution is imperative.

This same requirement holds true for time resolution when the pulse duration is exceedingly brief. The heat source is depicted as a surface load, with only two in-plane dimensions specified, and no penetration by the heat source is considered. Achieving accurate results hinges on selecting appropriate numbers of points for both spatial and temporal dimensions. The distance between adjacent points should be finely tuned to accurately represent the physics of the problem and should be smaller than the beam size and pulse duration, respectively. Consequently, the mesh must be adequately refined. As an illustrative example, the constants corresponding to a single straight welding line moving in the y-direction are provided here below.

Script 1/4: Constants

```
lx = 0.02   m % extension of the domain in x-direction
ly = 0.1    m % extension of the domain in y-direction
tmax = 1    s % extension of the domain in time
npx = 241 % number of evaluation points in x-direction
npy = 1201 % number of evaluation points in y-direction
nt = 501 % number of evaluation points in time per welding line
k = 1 % number of welding lines
```

In the second part of the algorithm, the indices for the array named "lht," which contains all the information about the heat source, are created. The indexing has a 3-dimensional structure: two dimensions describe the spatial position of the source on the selected surface, and the third dimension represents the time step. In APDL, such an array is defined using the "TABLE" command. In this case, the array indices are real (non-integer) numbers, which must be defined when populating the table. The index numbers for the rows and columns are stored in the zeroth column and row of the "array elements," and they are initially assigned values close to zero. These index numbers must be in ascending order and are used solely for retrieving an array element. When retrieving an array element with a real index that does not exactly match a specified index, linear interpolation is performed among the nearest indices to obtain the corresponding array element values. It's important to note that this approach is valid only for plane surfaces. For 3-D surfaces, an additional dimension needs to be considered, and the "TAB4" command

is used instead. For each position defined by x-y coordinates and at each time step, a heat flux value expressed in W/m^2 will be assigned in the subsequent step. This implies that the heat source is defined for all points within the domain, irrespective of the power intensity, even if the heat flux at that particular position and time step is zero. This method may not be ideal for very large domains due to the limited resolution available. However, the advantage lies in the simplicity of having a single domain for the entire simulation. Using local domains that move with the heat source would allow for a much higher resolution in the area of interest while ignoring portions of the model not involved in the heating process. However, this approach would necessitate significantly higher mathematical complexity. Here below the indexing corresponding to the example provided upon is documented.

Script 2/4: Array Indices

```
*dim,lht,table,npx,npy,k*nt,x,y,time % size of the array
*do,ix,1,npx % x indices
  locx=(ix-1)/(npx-1)*lx
*do,it,1,k*nt
  lht(ix,0,it)=locx
*enddo
*enddo
*do,iy,1,npy % y indices
  locy=(iy-1)/(npy-1)*ly
*do,it,1,k*nt
  lht(0,iy,it)=locy
*enddo
*enddo
*do,it,1,k*nt % time indices
  time=(it-1)/(k*nt-1)*tmax
  lht(0,0,it)=time
*enddo
```

The formula to calculate the array size (AS) is simply the product of the sizes associated with the three dimensions: x, y, and t:

$$AS = np_x \cdot np_y \cdot (n_t \cdot k). \tag{A.1}$$

In this example, the total size for the array "lht" is 145009941. To calculate the resolution (RE) for each dimension, given the domain extension and the number of points, the following formulas are used:

$$RE_x = \frac{lx}{np_x - 1}\,[\mathrm{m}], \tag{A.2}$$

$$RE_y = \frac{ly}{np_y - 1}\,[\mathrm{m}], \tag{A.3}$$

$$RE_t = \frac{t_{max}}{(n_t - 1) \cdot k}\,[\mathrm{s}]. \tag{A.4}$$

In this case study, all three resolutions are equal to 8e-5 m for both x and y dimensions and 2e-3 s for the t dimension. These values are smaller than the smallest details they need to capture, such as the beam radius and pulse duration.
Once the array and indices are set up, the matrix needs to be populated with power intensity or heat flux (HF) values in $\mathrm{W/m^2}$ for all generated combinations, at every location and time step. This process is defined as:

$$lht(x, y, t) = HF. \tag{A.5}$$

The power intensity is not directly calculated or defined in Ansys Workbench; instead, it is generated using a Matlab script, as explained in the following section. The Matlab script generates a .txt output file containing all non-zero heat flux values, which is then called within the APDL script, as indicated here below:

Script 3/4: Heat Fluxes

```
/NOPR
/inp, powersource, txt, Z:...
/GOPR
```

In the end, the complete array, encompassing indices and heat flux values, is employed to specify the surface load on the designated surface. The last part of the script provides an explanation of the syntax of the APDL command "sf".

Script 4/4: Surface Load

```
sf,top_face,hflu,%lht%
allsel
```

Appendix B

Matlab Script

The Matlab script takes various input variables, such as laser power, pulse duration, frequency, and velocity. It then generates a text file in APDL format containing the power source data. It's possible to specify these input parameters directly in Matlab or, if using it as part of an optimization process, also input them from an additional .csv file referred to in the Matlab script. Both of these options are explained in detail here below. The optimization process generates a file named "parameters.csv."

Script 1/4: Input Parameters

```
% all units expressed in m, s and W
Pt = 3600; % total power
t1 = 0.009; % pulse duration
f = 33.15; % frequency
v = 0.25; % velocity
tt = 1/f; % cycle period
t2 = tt - t1; % off duration
% alternatively
% M0 = readmatrix('\\trtswiewfs01.euce.corp.bshg.com:..csv');
% Pt=M0(1);
% t1=M0(2);
% f=M0(4);
% v=M0(3);
```

Two other important input parameters are the standard deviation (σ), which characterizes the laser beam's shape, and the

overall process efficiency (η) as defined in [12]. These parameters are defined separately since they remain constant. Additionally, there are inputs related to the specific welding path. In this script, for instance, it has been customized to simulate 5 horizontal welding lines on a 100x100 mm steel plate. The laser moves in the y-direction while maintaining a constant x-direction. In this part of the script, there are three additional parameters needed:

Script 2/4: Specific Inputs

$p = [90 \times 10^{-3}, 70 \times 10^{-3}, 50 \times 10^{-3}, 30 \times 10^{-3}, 10 \times 10^{-3}, -10 \times 10^{-3}]$; % x positions
$dy0 = 0.007$; % distance from the border
$p_{\max} = \max(\text{size}(p) - 1)$; % number of parallel lines

pmax represents the actual number of welding lines, while p has one more value because it considers the laser returning to the initial position for the next line, even for the last welding line. The domain and resolution are defined in the next part of the script, and their values must match those specified in the APDL script.

Script 3/4: Domain and Resolution

$np = 601$; % number of space divisions
$nt = 501$; % number of time divisions per line
$lx = 0.1$; % x domain
$ly = 0.1$; % y domain

The remaining portion of the script is presented here below:

Script 4/4: Power Source Array

```
% initialization
it=[];
ix=[];
iy=[];
r=0;
```

```
fid = fopen('powersource.txt','wt');

for n=1:pmax    %for every line


tp=sqrt((p(n+1)-p(n))^2 + (0.1 - dy0 * 2)^2)/v;    % return to next starting point
tline=(0.1-dy0*2)/v+tp;    % welding line + return to next starting point
ttot=pmax*(0.1-dy0*2)/v;    % total welding lines
npuls=round(t1/ttot*nt*((0.1-dy0*2)/v)/((0.1-dy0*2)/v+tp)*pmax);    % time
steps per cycle with laser on
nf=round(tt/ttot*nt*((0.1-dy0*2)/v)/((0.1-dy0*2)/v+tp)*pmax);    % time
steps per cycle
i=1;

for it=1:nt    % for every time step per line
    time=(it-1)/(nt-1)*tline;
    x0=p(n);    % x position center heat source
    y0=dy0+v*time;    % y position center heat source
    if it==1
      r=0;
    end
    if r>npuls-1
      i=0;    % laser off
    end
    if r==nf
      i=1;    % laser on
      r=0;    % counter pulse number
    end
    r=r+1;
    for ix=1:np    % for every x position
      x=(ix-1)/(np-1)*lx;
      dx=x-x0;    % x position relative to x0
      for iy=1:np    % for every y position
        y=(iy-1)/(np-1)*ly;
        dy=y-y0;    % y position relative to y0

power=Pt/2/pi/(s)^2 * exp(-(dx^2 + dy^2)/2/(s)^2);    % power distribution
power=power*i*e;    % power * on/off * efficiency
        if y0>0.1-dy0    % ignore values beyond the limit
          power=0;
        end
        lht(ix,iy,it+(n-1)*nt)=power;    % assign power distribution to position
and time step
        if power>0    % ignore 0 values
        fprintf(fid,'lht(%6.0f,%6.0f,%6.0f)=%6.0f \n',ix,iy,it+(n-1)*nt,power)
% print the results
        end
```

```
        end
      end
    end
end
fclose(fid);
```

Upon initializing the indexes, a series of nested for-loops are used to determine the power intensity for each position and time step. The index r tracks the pulse duration, and the binary index i takes on a value of 0 when the laser is off and 1 when the laser is on. The power remains constant when the laser is on and is set to zero when the laser is off. Finally, the power is also scaled by the overall process efficiency. Using the *fprintf* command, the resulting array is formatted in APDL style and exported to a .txt file named "powersource" Only the non-zero values are included in the output.

Bibliography

[1] Yuewei Ai, Xinyu Shao, Ping Jiang, Peigen Li, Yang Liu, and Wei Liu. Welded joints integrity analysis and optimization for fiber laser welding of dissimilar materials. *Optics and Lasers in Engineering*, 86:62–74, 2016.

[2] Atthaphon Ariyarit and Masahiro Kanazaki. Multi-fidelity multi-objective efficient global optimization applied to airfoil design problems. *Applied Sciences*, 7:1318, 2017.

[3] Antoni Artinov, Marcel Bachmann, and Michael Rethmeier. Equivalent heat source approach in a 3d transient heat transfer simulation of full-penetration high power laser beam welding of thick metal plates. *International Journal of Heat and Mass Transfer*, 122:1003–1013, 2018.

[4] A. Aruna and Ranjan Ganguli. Multi-fidelity response surfaces for uncertainty quantification in beams using coarse and fine finite element discretizations. *International Journal for Computational Methods in Engineering Science and Mechanics*, 22:1–20, 2020.

[5] Syrine Belakaria, Aryan Deshwal, and Janardhan Doppa. Multi-fidelity multi-objective bayesian optimization: An output space entropy search approach. 2020.

[6] D. Bergström. *The Absorption of Laser Light by Rough Metal Surfaces*. PhD thesis, Luleå University of Technology, Sweden, 2008.

[7] M. Courtois, M. Carin, P. Le Masson, and et al. A new approach to compute multi-reflections of laser beam in a keyhole for heat transfer and fluid flow modelling in laser welding. *Journal of Physics D: Applied Physics*, 46:505305, 2013.

[8] M. Courtois, M. Carin, P. Le Masson, and et al. Guidelines in the experimental validation of a 3d heat and fluid flow model of keyhole laser welding. *Journal of Physics D: Applied Physics*, 49:155503, 2016.

[9] M. Dal and R. Fabbro. An overview of the state of the art in laser welding simulation. *Optics & Laser Technology*, 78:2–14, 2016.

[10] R. Fabbro. Melt pool and keyhole behaviour analysis for deep penetration laser welding. *Journal of Physics D: Applied Physics*, 43:445501, 2010.

[11] R. Fabbro, M. Dal, P. Peyre, and et al. Analysis and possible estimation of keyhole depths evolution, using laser operating parameters and material properties. *Journal of Laser Applications*, 30:032410, 2018.

[12] Piero Favaretti and Lucia Parussini. A mesoscale approach to simulate residual deformations in complex laser welding processes. *SIMULATION*, 003754972211070, 2022.

[13] Piero Favaretti and Lucia Parussini. Nested vs. non-nested sampling: Definition of an infilling strategy for multi-fidelity bayesian optimization based on data correlation. *SAE Technical Paper*, 2023-01-0084, 2023.

[14] Shaoning Geng, Ping Jiang, Xinyu Shao, Lingyu Guo, and Xuesong Gao. Heat transfer and fluid flow and their effects on the solidification microstructure in full-penetration laser welding of aluminum sheet. *Journal of Materials Science & Technology*, 46:50–63, 2020.

[15] J. Goldak, M. Asadi, and L. Karlsson. *Numerical Aspects of Modeling Welds*. ASME Press, New York, 2011.

[16] Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. Thesis, Université Paris-Diderot - Paris VII, Paris, France, 2013. Autres [stat.ML].

[17] Loic Le Gratiet and Claire Cannamela. Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57, 2012.

[18] Hamidreza Heydari and Mohammad Akbari. Investigating the effect of process parameters on the temperature field and mechanical properties in pulsed laser welding of ti6al4v alloy sheet using response surface methodology. *Infrared Physics & Technology*, 106:103267, 2020.

[19] A. Hozoorbakhsh, M. Hamdi, M. I. S. Ismail, and et al. Cfd modelling of weld pool formation and solidification in a laser micro-welding process. *International Communications in Heat and Mass Transfer*, 101:58–69, 2019.

[20] H. Huang, J. Wang, L. Q. Li, and et al. Prediction of laser welding induced deformation in thin sheets by efficient numerical modeling. *Journal of Materials Processing Technology*, 227:117–128, 2016.

[21] Frank P. Incropera and David P. DeWitt. *Fundamentals of Heat and Mass Transfer*. John Wiley & Sons, 2007.

[22] Bradáč J. Calibration of heat source model in numerical simulation of fusion welding. 2013.

[23] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.

[24] P. Z. Korondi, M. Marchi, L. Parussini, and C. Poloni. Multi-fidelity design optimisation strategy under uncertainty with limited computational budget. *Optimization and Engineering*, 2020.

[25] P. Z. Korondi, M. Marchi, L. Parussini, D. Quagliarella, and C. Poloni. Multi-objective design optimisation of an airfoil with geometrical uncertainties leveraging multi-fidelity gaussian process regression. In *UQOP: International Conference on Uncertainty Quantification & Optimisation*, 2020.

[26] P. Le Masson, M. Courtois, M. Carin, and et al. A complete model of keyhole and melt pool dynamics to analyze instabilities and collapse during laser welding. *Journal of Laser Applications*, 26:042001, 2014.

[27] Chae Lim and Wei-Ying Wu. Conditions on which co-kriging does not perform better than kriging. *Journal of Multivariate Analysis*, 192:105084, 2022.

[28] Jonas Mockus. On a bayesian method for seeking an extremum. *Automatika i vychislitelnaja tekhnika*, (3), 1972.

[29] T. Mościcki, J. Hoffman, and Z. Szymański. Modelling of plasma plume induced during laser welding. *Journal of Physics D: Applied Physics*, 39:685–692, 2006.

[30] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006. c 2006 Massachusetts Institute of Technology.

[31] Matthieu Sacher, Olivier Le Maître, Régis Duvigneau, Frédéric Hauville, Mathieu Durand, and Corentin Lothode. A non-nested infilling strategy for multi-fidelity based efficient global optimization. ResearchGate, 2020.

[32] N. Shanmugam, B. G. Naidu, K. Sankaranarayanasamy, and et al. A transient finite element simulation of the temperature and bead profiles of t-joint laser welds. *Materials Design*, 31:4528–4542, 2010.

[33] G. Tirand, C. Arvieu, E. Lacoste, and et al. Control of aluminium laser welding conditions with the help of numerical modelling. *Journal of Materials Processing Technology*, 213:337–348, 2013.

[34] Johannes Will and Thomas Most. Metamodell of optimized prognosis (mop) - an automatic approach for user friendly parameter optimization, 2009.

[35] H. Xu, X. Guo, Y. Lei, and et al. Welding deformation of ultra-thin 316 stainless steel plate using pulsed laser welding process. *Optics and Laser Technology*, 119:105583, 2019.

[36] R. Zhang, X. Tang, L. Xu, and et al. Study of molten pool dynamics and porosity formation mechanism in full penetration fiber laser welding of al-alloy. *International Journal of Heat and Mass Transfer*, 148:119089, 2019.

[37] Y. Zhang, S. Li, G. Chen, and et al. Experimental observation and simulation of keyhole dynamics during laser drilling. *Optics and Laser Technology*, 48:405–414, 2013.

[38] J. Zhou, H. L. Tsai, and P. C. Wang. Transport phenomena and keyhole dynamics during pulsed laser welding. *Journal of Heat Transfer*, 128(7):680–690, 2006.