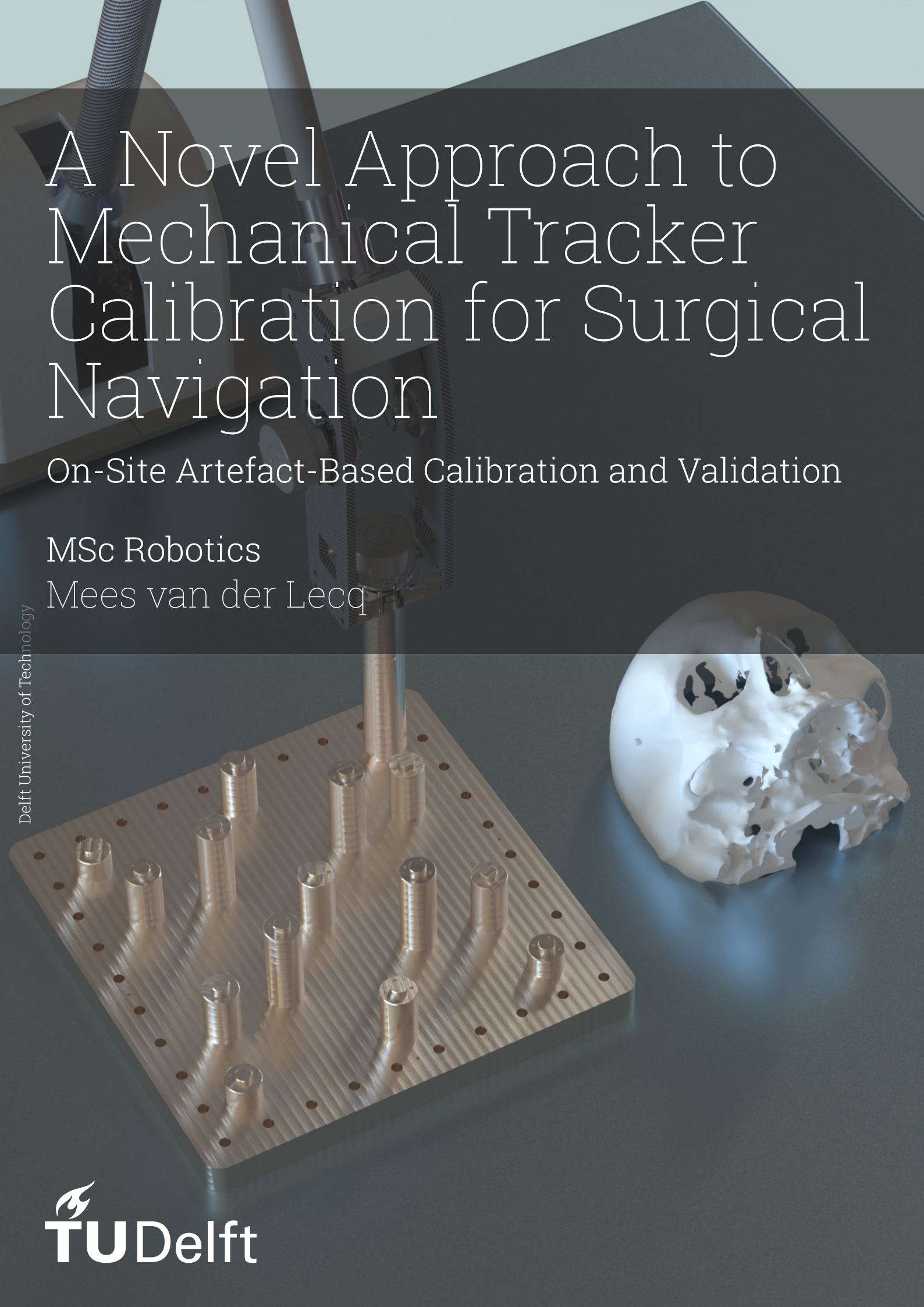
A Novel Approach to Mechanical Tracker Calibration for Surgical Navigation

On-Site Artefact-Based Calibration and Validation

MSc Robotics
Mees van der Lecq

Delft University of Technology



TU Delft

A Novel Approach to Mechanical Tracker Calibration for Surgical Navigation

On-Site Artefact-Based Calibration and Validation

by

Mees van der Lecq

Thesis committee:

Supervisor TUD:	Dr. Ir. Jens Kober
Supervisors Amsterdam UMC:	Drs. Naomi Rood
	Dr. Tom van Riet
	Dr. Ruud Schreurs
	Department of Oral and Maxillofacial Surgery
Faculty Department:	Faculty of Mechanical Engineering, Department of Cognitive Robotics
Student Number:	4887409

TU Delft

Acknowledgements

First and foremost, I would like to present my big thanks to drs. Naomi Rood for her dedication to my research, having to schedule her own systematic review and divide her time between ACTA and work in Haarlem. Her dedication over the duration of this project has been invaluable. I want to thank Ir. Willem Momma for his constant support. Though he graduated a year ago, he was still available between scheduled meetings, counseling me whenever I encountered problems in the research. We have known eachother since youth, but we have become real friends after this year of collaboration. In between working hours of saving lives, Dr. Tom van Riet was very involved during my research. He never failed to find time to assist outside scheduled appointments and could be contacted very directly. I may have been the first student to call rather than email, but direct contact and the friendly atmosphere made me very involved with the research group and excited for the project. I am thankful to Dr. Ruud Schreurs, whose frequent conference commitments did not prevent him from attending the meetings. His favorite time to support me appeared to be on sunny sundays, where we spent more than an hour working on the entire research structure. A special thanks to Dr. Ir. Jens Kober, who showed great flexibility in guiding a project with a little more on mechanical engineering, rather than his expertise in machine learning. Regardless of this small gap, he was always very involved in finding solutions to the dilemma's of this research. The 3ME FW team, especially Damian de Nijs and Reinier van Antwerpen who produced both the metal artefact and end-effector for me. You were very transparent and called me directly whenever there were questions or issues with the production. André Mendes Florindo of SamXL also deserves a special mention for having spent an entire day assisting me in co-certifying my calibration artefact with laser tracker hardware otherwise unavailable to me. I also thank the BEP team: Kris Klein Goldewijk, Rink van Hartingsveldt, Twan Grooff, and Wessel Wolterbeek Muller for developing the 6-DoF tracker arm that formed the foundation for this work.

To all those who made this adventure possible, thank you. Mees van der Lecq Delft, September 2025

Mees van der Lecq
Delft, September 2025

Abstract

Mechanical tracking systems offer potential advantages over optical and electromagnetic alternatives for surgical navigation, including freedom from line-of-sight constraints, immunity to metallic interference, and high-frequency position feedback. However, passive mechanical tracker arms lack the joint-locking capability required for traditional laser tracker calibration, and existing artefact-based methods fail to provide the complete six-degree-of-freedom constraints necessary for accurate kinematic parameter identification. This thesis presents a novel artefact-based calibration methodology specifically designed for passive mechanical tracking arms used in surgical navigation applications. The methodology employs a laser-certified steel calibration artefact featuring 15 asymmetric pillars that provide complete 6-DoF constraint through single contact points. Unlike traditional sphere-based methods that only constrain position, the asymmetric plug design mechanically prevents rotation, enabling simultaneous identification of all kinematic parameters without sequential joint locking. The artefact was precision-machined and certified using a Leica AT960 laser tracker to ±0.001 mm uncertainty, providing metrological traceability one order of magnitude better than the target accuracy. Applied to a custom-built six-degree-of-freedom Surgical Tracker Arm (SUTA), the calibration methodology achieved 1.01 mm RMS accuracy across 525 independent validation measurements, representing an 81.5% improvement from the 5.45 mm RMS baseline configuration. Dataset optimization analysis revealed that six measurement positions achieve near-optimal accuracy (1.09 mm RMS) while maintaining calibration efficiency, with marginal improvements beyond this point. The research identified mechanical interface compliance as the dominant error source, with clearance in the end-effector-pillar interface contributing approximately 0.5 mm to the total error. While the achieved accuracy does not yet match the 0.3-0.5 mm performance of state-of-the-art optical systems, the methodology successfully demonstrates that passive mechanical trackers can be calibrated to millimeter-level precision with a calibration artefact. The approach establishes a practical framework for calibrating passive tracking systems and identifies clear paths toward sub-millimeter accuracy through targeted hardware improvements, particularly in interface design and consistent seating mechanisms.

Contents

Nomenclature

Abbreviations

Abbreviation	Definition
BEP	Bachelor End Project
CFRP	Carbon Fiber Reinforced Polymer
CMM	Coordinate Measuring Machine
CMF	Craniomaxillofacial
CRC	Cyclic Redundancy Check
CT	Computed Tomography
DH	Denavit–Hartenberg
DoF	Degrees of Freedom
Hz	Hertz
ISO	International Organization for Standardization
MDH	Modified Denavit–Hartenberg
MPE	Maximum Permissible Error
RMS	Root Mean Square
RS-485	Recommended Standard 485 (serial communication protocol)
SUTA	Surgical Tracker Arm
TCP	Tool Center Point

1

Introduction

Image-guided surgical navigation has become an indispensable technology in modern surgical interventions. It enables the precise intraoperative localization of surgical instruments relative to imaging data and virtual surgical planning acquired prior to the operation. Robust tracking systems are essential for this technology, continuously monitoring patient and instrument pose throughout the surgical procedure. Modern surgical applications, particularly craniomaxillofacial (CMF) procedures, where target structures are frequently located within 1 mm of critical nerves and blood vessels, demand submillimeter accuracy for safe and effective surgical outcomes [1][2].

The most widely adopted navigation system is optical tracking, which tracks both instrument and patient using infrared cameras and reflective markers, representing the majority of the 1.2 billion US dollar surgical navigation market in 2022 [3]. The Brainlab Surgical Navigation System exemplifies optical tracking technology, shown in Figure 1.1. This setup illustrates the typical configuration required for optical tracking systems in surgical environments [4]. Despite the wide adoption of these technologies, current navigation systems suffer from significant limitations that create a substantial gap between the required precision and the achievable performance. Optical systems depend on continuous line-of-sight and provide limited update frequencies of 60-120 Hz with 5-7 ms latencies [5]. Electromagnetic tracking offers an alternative, overcoming line-of-sight constraints, but introduces systematic errors by metallic interference and thermal drift, while update rates are limited to 40 Hz [6][7]. These fundamental limitations prevent current surgical navigation technologies from meeting the precision and reliability requirements of demanding CMF procedures.

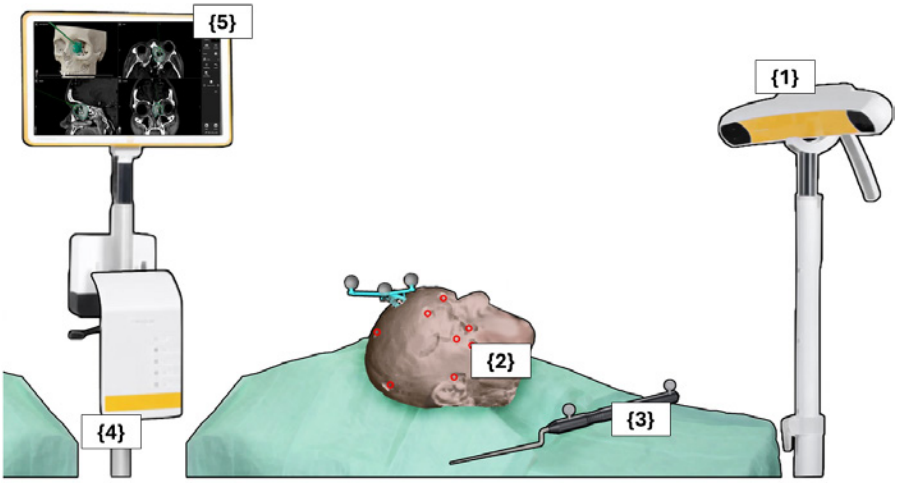


Figure 1.1: Brainlab optical tracking system: infrared cameras (1) track reflective markers on patient (2) and instruments (3), with positions registered to pre-operative imaging (5) on a portable monitor (4). Adapted from Schreurs et al. [4]

Mechanical tracking arms could address these core limitations by providing direct kinematic feedback through encoder measurements. Industrial articulated coordinate measuring machines (CMM) such as the Faro Quantum can achieve micrometer tracking accuracy across workspaces larger than the typical surgical workfield, with 600 Hz feedback and deterministic error characteristics [8]. Unlike optical systems, these mechanical trackers eliminate line-of-sight and electromagnetic interference dependencies, while providing predictable errors that can be systematically modeled and corrected. These characteristics make mechanical tracking particularly advantageous for surgical navigation applications where the operating field is often cluttered with personnel and equipment that obstruct optical line-of-sight, and where consistent high-frequency feedback enables smoother instrument guidance. Furthermore, the deterministic error characteristics and high update rates of mechanical systems provide the stable, predictable feedback necessary for potential future integration with surgical robotic systems, where control loop stability depends on consistent, high-frequency position updates. For surgical applications, tracking systems must monitor both the position and orientation of patient anatomy, as patient movement during procedures can alter the spatial relationship between preoperative imaging and actual anatomy [9]. Position and orientation tracking requires a robot arm that can measure six degrees of freedom (6-DoF) at it's end-effector. To achieve the highest possible accuracy, these arms require comprehensive calibration of all kinematic parameters.

Current active robot calibration relies on sequential parameter identification, where individual joints are actuated and locked to isolate specific kinematic parameters [10]. For active robotic systems, laser trackers represent the gold standard for this approach, providing highly accurate reference measurements during controlled joint movements [11]. Passive tracking arms however, are not equipped with motorized joints, removing the possibility for joint locking and controlled actuation. This constraint transforms traditional sequential calibration into a simultaneous parameter estimation challenge, where all kinematic parameters (link lengths, joint offsets and joint axis orientations) must be identified simultaneously without the ability to isolate individual parameters. Passive mechanical systems typically rely on calibration artefacts such as gauge blocks and ball bars that provide sufficient constraints for position-only tracking, but these methods cannot constrain the rotational parameters required for full 6-DoF tracking [12][13][14]. The custom-built Surgical Tracker Arm (SUTA), shown in Figure 1.2, was developed through collaboration with a Bachelor End Project team at TU Delft specifically for CMF surgical applications. As this passive mechanical system tracks both position and orientation (6-DoF) of the attached body, it requires a more comprehensive calibration approach than traditional position-only tracking systems.

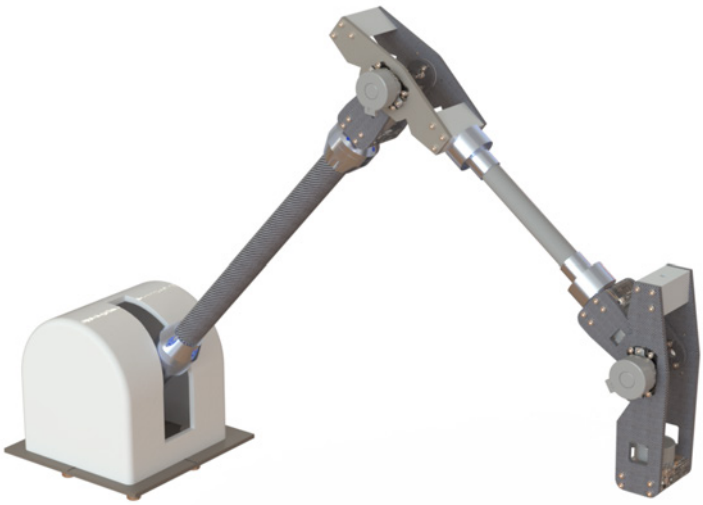


Figure 1.2: Custom-built mechanical Surgical Tracking Arm (SUTA) used as the platform for developing and validating the artefact-based calibration methodology. The 6-DoF passive kinematic chain enables comprehensive workspace coverage while maintaining low break-away forces for patient safety.

1.1. Research Objective

The objective of this thesis is to develop a calibration methodology specifically designed for passive mechanical tracking arms that require full 6-DoF accuracy while maintaining traceability to industrial standards.

1.2. Research Questions

To investigate whether this calibration approach can achieve the high accuracy needed for mechanical tracking in surgical navigation, the following research question is described:

How can a passive mechanical tracking arm be calibrated to achieve clinically acceptable accuracy using artefact-based methods?

This main question is addressed through three sub-questions:

RQ1: What artefact geometry enables efficient and fully constrained 6-DoF calibration of a passive tracker arm?

RQ2: What influence does calibration dataset size have on final tracking accuracy, and how can optimal data collection be determined?

RQ3: What is clinically acceptable accuracy and how can the calibrated system be validated against both industrial and surgical standards?

1.3. Contributions

This thesis makes four primary contributions toward enabling passive mechanical tracking for surgical navigation:

1. **6-DoF Artefact-Based Calibration Methodology:** Implementation of a laser-certified pillar-based calibration approach with asymmetric constraint geometry that, unlike traditional sphere-based methods, provides complete 6-DoF constraint through single contact points while maintaining metrological traceability.
2. **Systematic Dataset Optimization Framework:** Quantitative analysis shows how calibration is efficient for small dataset sizes while identifying trade-offs between efficiency and robustness. Optimal dataset configurations achieve accuracy comparable to larger data collection with significantly reduced measurement time.
3. **Hardware Enhancement of Surgical Tracker Arm (SUTA):** Refinement of the Bachelor End Project prototype through electrical grounding solutions and design of a specialized end-effector that enables 6-DoF calibration through asymmetric constraint interfaces.
4. **Validated Performance Achievement:** Demonstration that systematic calibration methodology achieves 1.01 mm RMS accuracy, transforming a manually configured tracker with 5.45 mm RMS into a precision tracking instrument with millimeter-grade accuracy.

1.4. Thesis Outline

This thesis is structured as follows:

- **Chapter 2** introduces the theoretical foundations of mechanical tracker calibration, including kinematic modeling frameworks, calibration optimization theory, and relevant ISO standards.
- **Chapter 3** presents the calibration methodology, including the custom tracker arm design, reference artefact development, experimental setup, and validation protocols.
- **Chapter 4** reports experimental results, including calibration accuracy, dataset size effects, and independent validation against industrial standards.
- **Chapter 5** discusses the findings, answers the research questions, analyzes limitations, and identifies opportunities for future development.

2

Related Work

This chapter establishes the theoretical foundation for mechanical tracker calibration. Section 2.1 presents the Modified Denavit-Hartenberg (MDH) kinematic framework used to model the tracker's joint transformations. Section 2.2 covers calibration optimization theory, including least squares formulation and robust loss functions for handling measurement outliers. Section 2.3 introduces ISO standards for systematic performance evaluation of articulated measurement systems. Section 2.4 analyzes existing calibration methods, evaluates their fundamental limitations for passive tracker systems, and identifies the critical gap in achieving complete 6-DoF constraint without specialized infrastructure.

2.1. Kinematic modeling

Mechanical tracking systems require precise kinematic models to transform joint angles into end-effector poses through forward kinematics. The accuracy of this transformation determines overall system performance, making kinematic modeling the foundation for achieving high accuracy.

2.1.1. Modified Denavit-Hartenberg Method

The MDH convention [15] provides systematic representation of kinematic chains through four parameters per joint: link length (a_i), link offset (d_i), joint angle (θ_i), and twist angle (α_i). The MDH convention locates coordinate frames at joint axes rather than link ends, improving numerical stability for calibration applications while maintaining computational efficiency for real-time robotic systems. The parameters used in this method will be referred to as DH parameters.

Figure 2.1 illustrates the DH parameters for a standard three-joint kinematic chain. At Joint 1, the z_1 axis is aligned with the joint's rotation axis. Moving to Joint 2, the parameters a_2, d_2, θ_2, and α_2 fully describe the spatial relationship between Joint 1 and Joint 2:

- a_2 (link length): distance from z_1 to z_2 measured along x_2,
- d_2 (link offset): displacement along z_1 to reach the intersection with x_2,
- θ_2 (joint angle): rotation about z_1 that aligns x_1 with x_2,
- α_2 (twist angle): rotation about x_2 that aligns z_1 with z_2.

Similarly, the parameter a_3 defines the link length between Joint 2 and Joint 3, measured along x_3. Together, these parameters describe the geometry of the chain as shown in the figure.

The homogeneous transformation matrix for joint i follows the MDH convention:

$$\mathbf{T}_i = \begin{bmatrix} \cos(\theta_i) & -\sin(\theta_i)\cos(\alpha_i) & \sin(\theta_i)\sin(\alpha_i) & a_i\cos(\theta_i) \\ \sin(\theta_i) & \cos(\theta_i)\cos(\alpha_i) & -\cos(\theta_i)\sin(\alpha_i) & a_i\sin(\theta_i) \\ 0 & \sin(\alpha_i) & \cos(\alpha_i) & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2.1}$$

This parametrization enables systematic representation of kinematic chains and is computationally efficient for real-time robotic applications.

2.1.2. Forward Kinematics

Forward kinematics computes end-effector pose from joint angles through sequential transformation multiplication. For a 6-DoF kinematic chain, the end-effector transformation becomes:

$$\mathbf{T}_{end} = \mathbf{T}_1 \cdot \mathbf{T}_2 \cdot \mathbf{T}_3 \cdot \mathbf{T}_4 \cdot \mathbf{T}_5 \cdot \mathbf{T}_6 \tag{2.2}$$

Position and orientation are then extracted from the final transformation matrix:

$$\mathbf{p} = \mathbf{T}_{end}[0{:}3, 3], \quad \theta_{end} = \text{atan2}(\mathbf{T}_{end}[1, 0], \mathbf{T}_{end}[0, 0]) \tag{2.3}$$

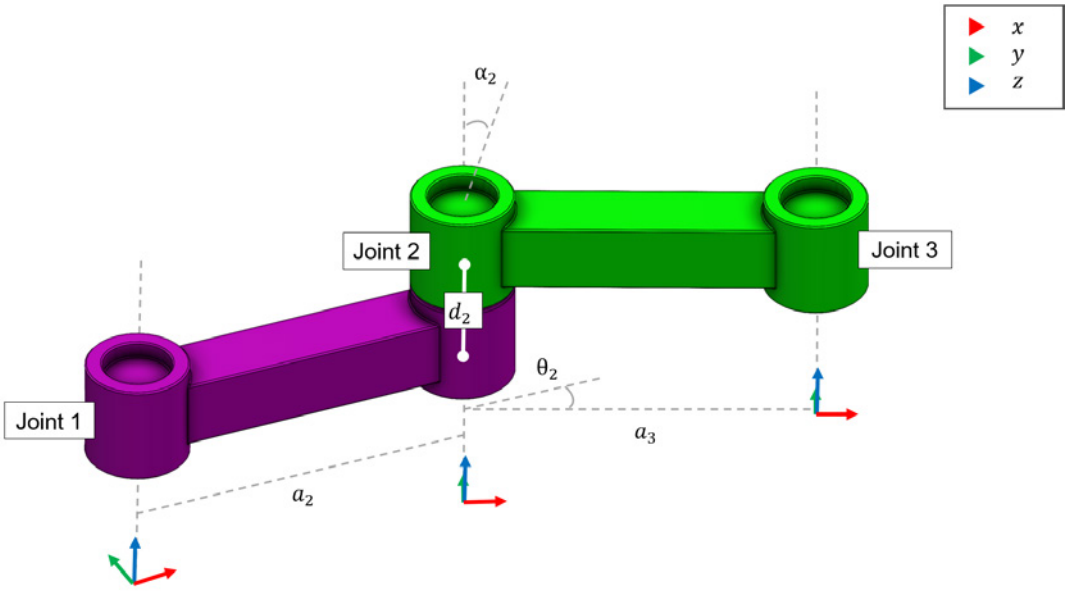


Figure 2.1: Modified Denavit-Hartenberg (MDH) parameter geometry for a three-joint kinematic chain. The four DH parameters are illustrated: a_2 and a_3 (link lengths), d_2 (joint offset), θ_2 (joint angle), and α_2 (twist angle). Coordinate frames follow the standard x-red, y-green, z-blue convention.

2.2. Calibration Optimization

Kinematic calibration requires solving a non-linear optimization problem to identify DH parameters from measurement data. This section presents the mathematical framework for parameter estimation.

2.2.1. Least Squares Optimization

The calibration objective minimizes the error between measured and reference distances:

$$\min \sum_{i,j} \|\mathbf{f}(\mathbf{q}_i, \mathbf{q}_j, \mathbf{p}) - d_{ij}^{GT}\|^2 \tag{2.4}$$

where $\mathbf{q}_i$ and $\mathbf{q}_j$ represent joint configurations at positions i and j, $\mathbf{p}$ contains DH parameters, and d_{ij}^{GT} are reference pairwise Euclidean distances.

The function $\mathbf{f}(\mathbf{q}_i, \mathbf{q}_j, \mathbf{p})$ computes the Euclidean distance between end-effector positions:

$$\mathbf{f}(\mathbf{q}_i, \mathbf{q}_j, \mathbf{p}) = \|\mathbf{x}(\mathbf{q}_i, \mathbf{p}) - \mathbf{x}(\mathbf{q}_j, \mathbf{p})\|_2 \tag{2.5}$$

where $\mathbf{x}(\mathbf{q}, \mathbf{p}) \in \mathbb{R}^3$ denotes the end-effector position computed via forward kinematics.

This distance-based formulation is independent from global coordinate frame registration between the measurement system and robot, and remains invariant to rigid body transformations. Multiple pairwise distance measurements are sufficient because they create an overdetermined system of equations, enabling robust least-squares solution of the DH parameters [16].

2.2.2. Robust Loss Function

Contact-based measurements may introduce outliers due to the mechanical and human influence on the measurements. The Huber loss function [17] provides robustness:

$$\rho(r) = \begin{cases} \frac{1}{2}r^2 & \text{for } |r| \leq \delta \\ \delta|r| - \frac{1}{2}\delta^2 & \text{for } |r| > \delta \end{cases} \tag{2.6}$$

Standard least-squares optimization applies quadratic penalties to all residuals, making the solution sensitive to outlier measurements. A Huber loss function maintains quadratic behavior for residuals within a certain threshold δ to preserve statistical efficiency. Beyond this threshold it transitions to linear growth for larger residuals. This bounds the influence of potential outliers without explicit identification. The threshold parameter δ determines the transition point between quadratic and linear regimes, typically set based on the expected measurement noise level.

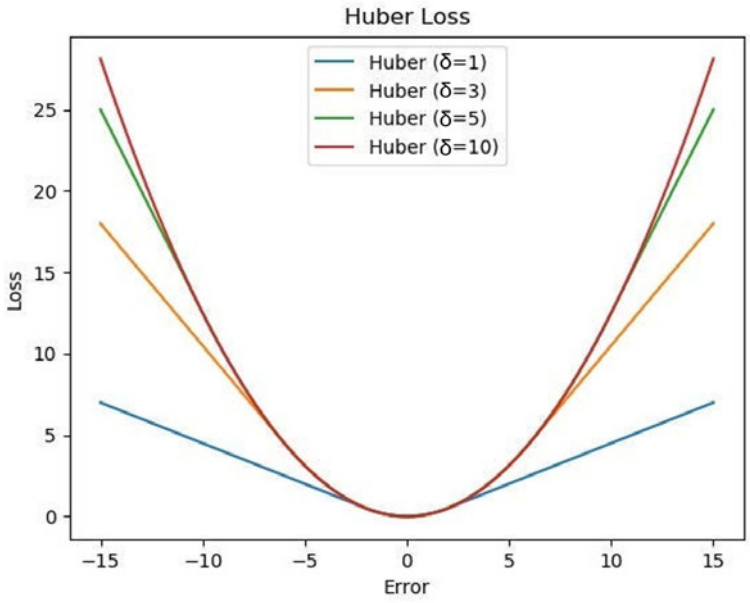


Figure 2.2: Huber loss function behavior for different values of threshold parameter δ, showing the transition from quadratic to linear loss regions.

2.3. Standards

This research adopts ISO 10360-12:2016 as an evaluation framework for the mechanical tracking system [18]. This standard, specifically designed for articulated arm coordinate measuring machines, provides systematic test protocols for multi-joint serial kinematic chains with encoder-based sensing. While the ISO standard specifies Maximum Permissible Error (MPE) as its primary metric, surgical navigation systems predominantly report Root Mean Square (RMS) error [19]. RMS provides a statistical measure that captures both magnitude and variability of errors, making it the established metric for comparing tracking system performance in surgical literature. Therefore, this research reports RMS error to maintain compatibility with surgical navigation literature while applying ISO 10360-12 test methodologies for systematic evaluation.

The ISO 10360-12 framework defines three primary tests for articulated arm evaluation. The length measurement error test evaluates dimensional accuracy using calibrated reference lengths across the workspace. Single-point repeatability assesses the system's ability to return to the same position through repeated measurements. Volumetric accuracy testing quantifies how joint configuration changes affect measurement accuracy by approaching the same point from different arm articulations. These tests characterize both local repeatability and global accuracy throughout the measurement volume. Additionally, ISO 9283 provides the standardized repeatability assessment protocol, requiring 30 consecutive measurements at a single position to statistically characterize short-term measurement variability [20]. This dual-standard approach enables both static repeatability characterization and dynamic workspace evaluation, essential for surgical tracking systems that must maintain accuracy throughout continuous movement.

To evaluate the performance of the SUTA system, a comparison will be made with the most widely adopted CMF surgical navigation systems, as well as an industrial reference. Table 2.1 presents the laboratory performance of these systems. Optical tracking systems are widely adopted in CMF surgery for their sub-millimeter accuracy and reliability [21]. Electromagnetic tracking achieves comparable laboratory accuracy but sees limited CMF adoption due to metallic instrument interference [6, 19]. YOMI represents the first FDA-cleared mechanical tracking system for dental surgery, demonstrating the clinical viability of mechanical approaches [22], however, there is little reporting information available for this system.

Table 2.1: Laboratory performance specifications of tracking technologies.

System	Technology	Accuracy	Metric
Surgical Navigation Systems			
Brainlab Curve/Kick	Optical	0.3 mm	RMS[a]
NDI Aurora	Electromagnetic	0.48 mm	RMS
YOMI	Mechanical	<1.00 mm	95% CI
Industrial Reference			
FARO Quantum Max S	Mechanical	0.024 mm	MPE

These accuracies are measured under controlled laboratory conditions with minimal environmental interference. In CMF surgery, performance accuracy degrades by a factor of approximately 1.5-3 times the laboratory values, with clinical accuracies ranging from 1.0-2.0 mm [23][24][25]. This degradation occurs due to multiple factors including increased distance between tracking cameras and the surgical field, environmental electromagnetic interference, patient movement, registration errors, and the presence of surgical instruments and staff that can obstruct line-of-sight in optical systems [26][27]. The optical and electromagnetic systems are widely used in clinical applications, which offers a benchmark for "clinically acceptable accuracy" in the range of 0.3-0.5 mm RMS for this research.

2.4. Calibration Artefact Design

Achieving surgical-grade accuracy through mechanical tracking requires precise calibration methods. This section analyzes existing calibration approaches and their fundamental limitations for passive tracker systems and identifying the gap that needs to be addressed for this research. The SUTA system is custom-built and its kinematics contain the following requirements for calibration.

2.4.1. Fundamental Calibration Requirements

The analysis of passive tracker calibration requirements reveals five essential criteria:

1. **Complete 6-DoF constraint**: Each measurement must constrain all six degrees of freedom to enable simultaneous parameter optimization
2. **Geometric determinism**: Contact geometry must provide unique, unambiguous positioning with single configuration
3. **Compact workspace**: Artefact must fit within surgical workspace constraints while providing geometric diversity

4. **Metrological traceability**: Reference coordinates must have precision at least one order of magnitude better than target accuracy
5. **Clinical accessibility**: Method must be portable and enable routine verification without specialized equipment or trained operators

2.4.2. Laser Tracker Calibration

Laser tracker systems represent the gold standard for kinematic calibration, providing sub-millimeter accuracy across large workspaces. However, their application to passive tracker arms faces a fundamental barrier. Laser tracker calibration of robotic systems uses sequential parameter identification where joints are actuated to specific positions and then locked. One joint remains free to rotate, while the other joints are locked, this isolates the moveable joint's kinematic parameters. By progressively progressing through the chain with each joint, the kinematic model can be accurately identified [28]. Passive tracker arms that are not motorized, do not allow for the joints to be locked so all kinematic parameters must be identified simultaneously from the measurement data. This transforms the single parameter problem into a complex multi-parameter calibration challenge, making laser-tracking incompatible for the SUTA system.

2.4.3. Calibration Artefacts

While laser trackers cannot be used directly for passive arms, various calibration artefacts have been developed to provide reference constraints for CMM's. These artefacts are described and compared in the following paragraphs.

Linear Artefacts

Ball-bar calibration systems consist of bars with precise lengths and spheres at each end that have known distances between measurement points. During calibration, the ball-bar is positioned between pairs of points inside the robot workspace, creating a network of distance constraints that can be used to identify kinematic parameters through optimization [14][16]. Gauge blocks similarly provide calibrated linear references that serve as a dimensional standard for length measurements in the workspace. While these linear measurement methods are well-established in metrology, they only provide distance constraints between points, failing to address the rotational parameters required for complete 6-DoF constraint.

Sphere-Based Artefacts

ISO 10360-2 sphere plates consist of precision-machined spheres mounted on stable granite or steel bases in calibrated patterns. The tracker end-effector contacts these spheres at various points to interpolate their centers which serve as the reference positions [29]. The spherical geometry allows contact from any approach angle, making the system versatile for different tracker configurations. However, sphere-based systems constrain only the three translational degrees of freedom at each contact, providing no direct orientation information. Additionally, the ambiguity of spherical contacts, where multiple contact points can satisfy the same constraint, fails the geometric determinism requirement.

Lattice Artefact

Zhao et al. developed a calibration artefact that uses precision-machined conical sockets arranged on three perpendicular planes within a compact volume. The artefact features multiple measurement points with each socket position certified to 9 µm uncertainty using laser tracker calibration [30]. The design enabled fast data collection (20 min rather than 2 hours with gauge block) through the high density of reference positions. However, this approach still relies on point-based contact features that provide primarily positional constraints, requiring additional procedures or indirect methods to determine the three rotational parameters necessary for complete 6-DoF calibration.

2.4.4. Comparison of Artefact Methods

Table 2.2 evaluates existing artefact-based calibration methods against the requirements from the beginning of the section. While each method offers certain advantages, none fully satisfy all criteria necessary for passive tracker calibration in clinical environments.

Table 2.2: Evaluation of calibration methods against passive tracker requirements. Checkmarks ✓ indicate the requirement is satisfied, crosses × indicate it is not met.

Method	6-DoF Constraint	Geometric Determinism	Compact Workspace	Metrological Traceability	Clinical Accessibility
Ball-bar/Gauge	×	✓	✓	✓	✓
Sphere Plates	×	×	×	✓	✓
Lattice (Zhao)	×	✓	✓	✓	✓

The analysis reveals a critical gap in existing calibration methodologies. Laser trackers provide complete 6-DoF constraint but require specialized infrastructure incompatible with routine clinical use. Artefact-based methods fail to constrain all rotational parameters necessary for the SUTA system. This fundamental limitation represents the primary challenge for passive tracker calibration.

3

Methodology

3.1. System Overview

This chapter presents the calibration methodology developed to achieve high-precision calibration of passive tracker arms without relying on laser tracker equipment; the inability to use joint-locking for calibration due to their lack of motor-controlled joints. It also overcomes the problems with traditional calibration artefacts; the insufficient constraining kinematic degrees of freedom. This chapter presents the calibration methodology developed that aims to achieve the highest possible accuracy for the SUTA system. The methodology integrates five components described in the following sections:

- **The custom 6-DoF TU Delft mechanical tracker arm** (Section 3.2.1) - the device requiring calibration, its specifications and uncalibrated kinematic model
- **Laser-certified steel calibration artefact** (Section 3.2.2) - the steel traceable metrological ground truth reference artefact, its design, manufacturing and certification process
- **System integration and coordinate frames** (Section 3.2.3) - mounting specifications, coordinate conventions and software implementation
- **Experimental procedures** (Section 3.3) - data acquisition protocols, calibration algorithms and independent validation procedures
- **Data processing and analysis** (Section 3.4) - performance metrics, dataset optimization analysis and repeatability tests

3.2. Experimental Setup

3.2.1. Surgical Tracker Arm (SUTA)

The SUTA system used in this research was initially developed as a prototype by a Bachelor End Project team from TU Delft's Mechanical Engineering department. The team designed a mechanical tracking arm to overcome line-of-sight limitations of optical systems while providing high-frequency feedback for surgical navigation applications. While the mechanical structure provided a functional foundation, the system exhibited position errors with a mean of 28 mm due to incorrect kinematic model configuration, making it unsuitable for precision measurement. This research contribution included the following enhancements: (1) reconfiguration of the kinematic model through approximation using a physical ground truth, reducing errors to approximately 5 mm, which is set as the baseline for the performance in this research; (2) development of a specialized end-effector with asymmetric interface geometry to enable 6-DoF constraint with the calibration artefact, shown in Figure 3.2; and (3) implementation of electrical grounding throughout all six links to prevent static discharge damage to the encoder transceivers. (4) disregarding measurement cache was necessary to obtain the real-time measurement data. Initially, the first 10 measurement were old cached encoder measurements, making the data acquisition a bad measurement representation. These hardware enhancements, detailed in Appendix B, were essential actions enabling the calibration methodology of this thesis.

The refined SUTA system, shown in Figure 3.1, consists of a 6-DoF passive kinematic chain designed for navigation in CMF surgery. Its workspace, a hemisphere with 100 mm radius, accommodates the motion range required for skull navigation while maintaining low break-away forces for patient safety. The system employs carbon fiber reinforced polymer (CFRP) links in an alternating Pitch-Roll (P-R-P-R-P-R) joint configuration, with spring-loaded gravity compensation and counterweights to minimize operational forces. Table 3.1 summarizes the key specifications of the system.

The SUTA employs high-resolution absolute encoders providing 22-bit angular resolution, enabling joint angle measurement with ±0.00004° precision. Data acquisition operates at 768 Hz with 2-second measurement averaging to reduce noise while maintaining operational efficiency. This averaging duration balances measurement stability against practical time constraints during the calibration procedure. Complete tracker arm design details, including the Bachelor End Project foundation, electrical grounding implementation, and end-effector development process, are documented in Appendix B.

Table 3.1: Key specifications of the Surgical Tracker Arm (SUTA) showing design parameters for 6-DoF passive patient tracking.

Parameter	**Value**
Degrees of Freedom	6
Link Material	CFRP
Encoder Type	22-bit ACURO AD35 Absolute
Data Acquisition Frequency	768 Hz
Workspace Radius	100 mm
Measurement Averaging	2 seconds

The kinematic model follows the MDH convention for systematic representation of the joint transformations. Table 3.2 presents the initial DH parameters before optimization. These initial parameters, derived from the mechanical design, require calibration optimization to account for manufacturing tolerances and assembly deviations that are present in the custom-built system. Furthermore, the joint encoders need to be calibrated to represent the their true angles, as they are assembled at uncontrolled initial rotations. Thus, the parameters optimized in this research are the link lengths d_i, a_i and joint angle offsets $\delta\theta_i$. At the end of the chain is the Tool Center Point, which indicates the measured point at the end-effector. d_7 indicates the length of the end-effector, which is also included in the calibration.

Table 3.2: Modified Denavit–Hartenberg parameters used in the implementation. Base at $z{=}0$ with $d_1 = 0.0458$. Longitudinal link lengths are placed in d_i; $a_i{=}0$ for all rows. Twist angles α_i in radians; encoder offsets $\delta\theta_i$ in degrees. The Tool Center Point (TCP, row 7) is a fixed translation that rotates with θ_6. The parameters included in the calibration are highlighted in bold.

Joint	θ_i *(rad)*	d_i *(m)*	a_i *(m)*	α_i *(rad)*	$\delta\theta_i$ *(deg)*
1	θ_1	**0.047**	0	$\pi/2$	**-1.6**
2	θ_2	**0**	0	$\pi/2$	**20.7**
3	θ_3	**0.571**	0	$-\pi/2$	**201.3**
4	θ_4	**0**	0	$\pi/2$	**117.1**
5	θ_5	**0.371**	0	$\pi/2$	**88.7**
6	θ_6	**0**	0	$\pi/2$	**-11.6**
7	- (TCP)	**0.253**	0	0	-

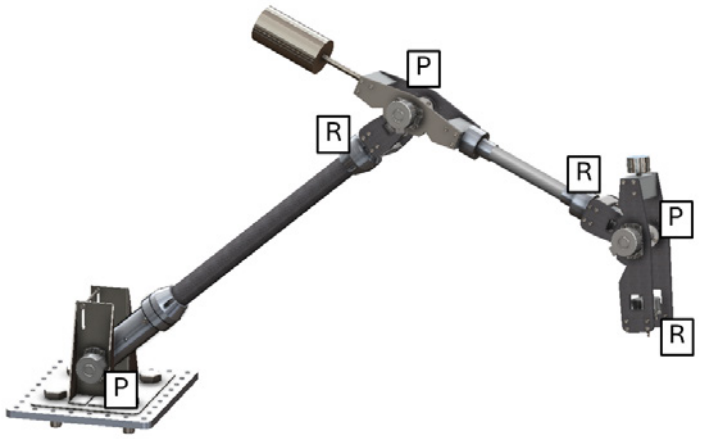


Figure 3.1: Surgical Tracker Arm (SUTA) with P-R-P-R-P-R joint configuration showing alternating pitch (P) and roll (R) joints connected by CFRP links, mounted on precision optical table for calibration experiments. The specialized end-effector developed for this research enables asymmetric coupling with calibration artefact pillars.

3.2.2. Artefact

The calibration artefact is a steel object that serves as the ground truth reference for all calibration and validation procedures in this research. Unlike traditional artefact calibration tools from related works that only constrain position, this design enables 6-DoF constraints through single orientation landmark and end-effector interface. Figure 3.2 illustrates this feature, it enables simultaneous position and orientation calibration for the SUTA kinematic configuration. The figure shows the end-effector socket in purple and the artefact pillar geometry in blue.

Geometric Design

The artefact employs an asymmetric pillar design to provide complete 6-DoF constraint at each measurement point. Unlike spherical or conical contacts that allow rotation about the contact normal, each pillar features a cylindrical base with an asymmetric plug that mechanically prevents rotation, ensuring unique positioning and orientation. This geometric constraint is essential for passive tracker calibration where all kinematic parameters must be identified simultaneously, without the ability to isolate individual joints, the optimization requires unambiguous pose constraints at each measurement.

The artefact consists of 15 pillars distributed across three height levels: short (S) 20 mm, medium (M) 40 mm, and tall (T) 60 mm, within a 180×180×60 mm workspace. This workspace corresponds closely to the hemispherical workspace defined in Section 3.2.1, moreover, it maximizes the milling machine workspace, offering the largest sampling volume (aiming for 100 mm radius) within the internal structure. The choice of 15 landmarks balances multiple practical constraints. With 13 kinematic parameters requiring identification, a minimum of 13 independent constraints is mathematically necessary. However, this bare minimum provides no robustness against measurement noise or outliers. The 15 landmarks generate 105 pairwise distance constraints per position, providing an overdetermination. This level of redundancy ensures robust least-squares optimization while remaining practical for manual data collection. Moreover, this layout leaves a margin for preventing collision of the SUTA body with the artefact. If a small pillar would be too close to a tall pillar, the final link before the end-effector could collide with the tall pillar, possibly damaging the SUTA or artefact structure. The height differences aim to simulate part of the hemispherical surface, while accounting for the production limitations of the milling machine, further described in the next paragraph. The O pillar serves as the homing position for each calibration and is designed to be the median height of all pillars. The reason for this was that initially, the algorithm would lock the measured O position to the ground truth O position whereafter the remaining measured points would be transformed as a solid body around the O position to find the best fit. This was later changed to a different approach, leaving the O position at the middle height rather

than the hemispherical surface. Each pillar is configured in a unique orientation to enable the widest range of end-effector orientation sampling. The symmetric layout of landmarks allows for the investigation of potentially problematic landmarks. For example, if T2 has a far greater error contribution than other landmarks, the artefact can be rotated to analyse whether this phenomenon is truly landmark dependent as the landmark would take on the position of one of the other landmarks.

Manufacturing constraints influenced the final geometry. While ideally the pillars would be oriented orthogonally to the local hemispherical workspace surface, the complexity of mounting the pillars to the base led to a practical design choice where all pillars are perpendicular to the ground plane. Unfortunately, this orientation forces the end-effector to remain perpendicular to the baseplate, but this limitation is acceptable because reaching pillars at different heights and radial positions still requires diverse joint configurations throughout the SUTA's kinematic chain. This is one of the reasons for placing the artefact at different positions outside of the original workspace. The varying joint angles needed to position the end-effector at each pillar location provide sufficient kinematic excitation for parameter identification, as the calibration objective focuses on identifying joint parameters rather than end-effector orientation accuracy. The 200×200×15 mm steel baseplate maximizes the DMG MORI MillTap 700 workspace, while the 10 mm offset between pillar structure (180×180 mm) and baseplate edge provides clearance for 6.5mm mounting holes for the mounting table explained in Section 3.2.3. The maximum pillar height of 60 mm represents the practical limit before milling vibrations compromise dimensional accuracy, as documented in Appendix C. Figure 3.3 shows the artefact geometry in isometric and top-down view, where the pillar orientations are chosen randomly and can be seen in the right image of the figure.

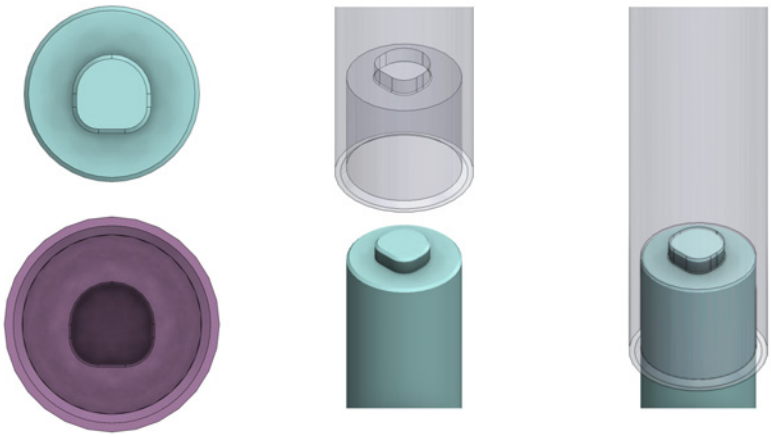


Figure 3.2: The end-effector and pillar interface. The dedicated asymmetric plug and hole design enables 6-DoF constraints of the entire SUTA kinematic configuration. Top left shows the pillar, bottom left the end-effector profile. The middle and right images illustrate the end-effector fitting over an artefact pillar.

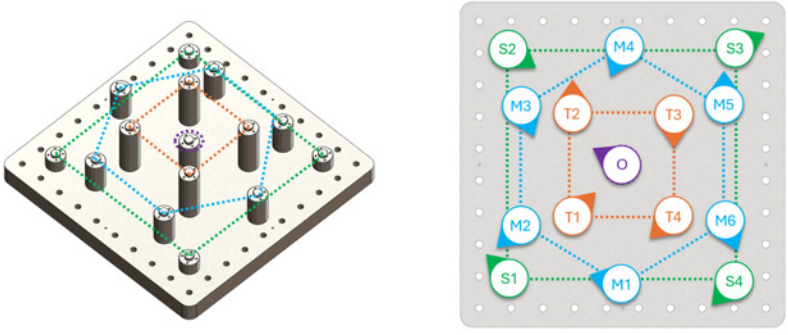


Figure 3.3: Calibration artefact design showing 15 precision-machined steel landmarks in 180x180x60 mm lattice. The right image is a top view of the artefact, it shows the landmark labels with the distinction in pillar heights and orientations.

Materials, Production and Laser-Tracker Certification

Steel construction (S235JR) was chosen to provide dimensional stability and wear resistance superior to 3D-printed alternatives while enabling precision machining of complex constraint features. Initial experiments with 3D-printed artefacts showed significant wear and dimensional drift after repeated use, making steel more suitable for long-term calibration accuracy. This material durability consideration is particularly relevant for the application of periodic tracking system maintenance. The artefact was precision-machined using a DMG MORI MillTap 700 and certified using a Leica AT960 laser tracker that provides ±0.001 mm measurement uncertainty [31]. The certification consists of a systematic seven-step protocol establishing a complete 6-DoF coordinate system of the artefact (pillar positions and orientations).

Extensive documentation of the artefact development is provided in two appendices: Appendix B describes the complete design evolution from initial 3D-printed prototypes through geometric optimization to the final steel design, while Appendix C covers material choice, manufacturing processes, laser tracker certification protocols, and detailed coordinate specifications.

3.2.3. System Integration

The complete system integrates the SUTA and artefact on a NEXUS Optical Table. While the table provides ±0.025 mm flatness specification, this is not critical for the calibration methodology since all measurements are pairwise distances between pillars at each artefact position, making them independent of the table's absolute flatness. Figure 3.4 shows the system configuration with global coordinate frame conventions. The tracker base (B) is positioned at the system origin (0, 0, 0), with the coordinate system following right-hand convention where the z axis is vertical marked in blue, positive x in green and positive y in red. The end-effector (E) attaches to the pillars of artefact (A), which is placed at several positions around the workspace, further explained in Section 3.3. The 25 mm separated M6 mounting holes on the NEXUS (N) table enable precise repositioning of the artefact for consistent sampling.

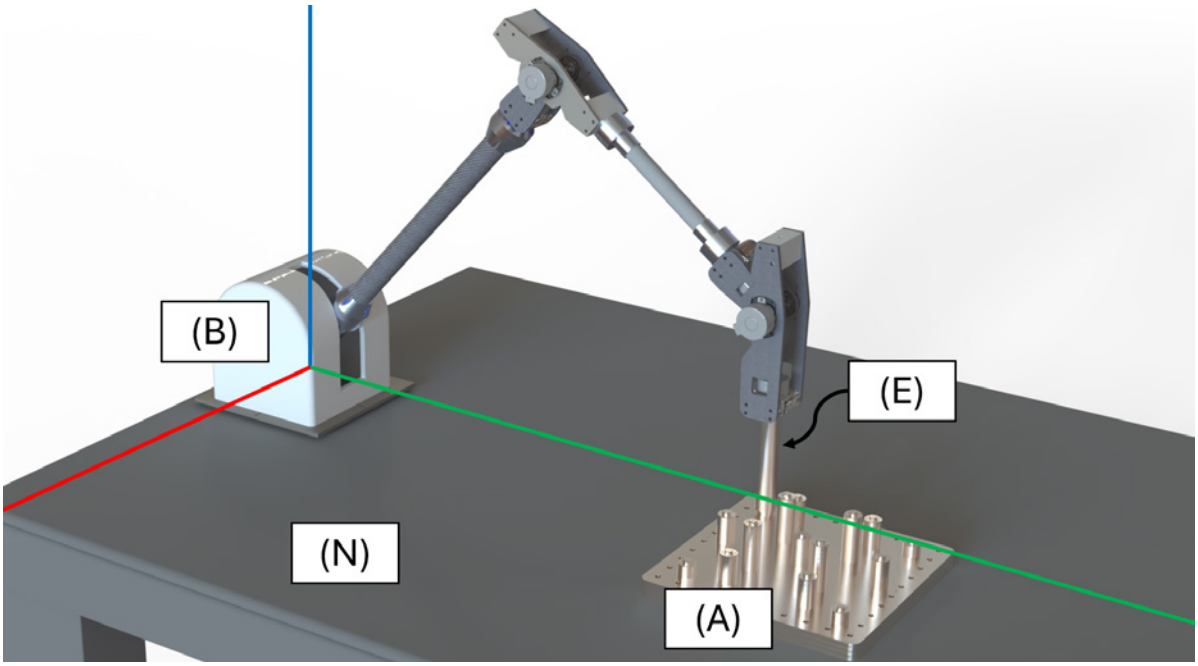


Figure 3.4: Calibration system setup with SUTA tracker arm (B: base, E: end-effector) and steel artefact (A) mounted on NEXUS optical table (N). Coordinate axes shown: z-axis (blue) vertical, x-axis (green), y-axis (red).

3.2.4. System Characterization

Prior to calibration optimization, the system repeatability was characterized through ISO 9283 protocols to establish baseline performance and identify mechanical limitations that might affect calibration accuracy. The repeatability assessment consisted of 30 consecutive measurements of a single landmark to isolate short-term measurement noise from systematic errors. This protocol quantifies the measurement uncertainty of the tracker system, establishing the theoretical lower bound for calibration accuracy expectations.

Additionally, an analysis was constructed to find a relation between a directional force on the end-effector's XY plane and the measurement consistency. The end-effector was pushed in multiple directions to identify an effect of mechanical compliance in the tracker-artefact interface. The analysis showed improved consistency (from 0.95 mm to 0.45 mm RMS) when consistent directional force was applied. This indicates that the backlash between end-effector and landmark can cause large errors if neglected. Thus, for each measurement, the end-effector was rotated counter-clockwise to lock the backlash, aiming for a consistent compensation. Full explanation of these experiments are described in Appendix E

3.2.5. Software Implementation

The SUTA calibration is implemented in Python3 using the Scipy library for least-squares optimization [32]. The software architecture employs a robust Huber loss function, described in Section 2.2.2 to handle measurement outliers while maintaining computational efficiency. Forward kinematics computations transform encoder readings to end-effector pose with respect to the system origin, enabling real-time position calculation during data acquisition. Data acquisition operates at 768 Hz with automatic averaging and logging functionality. The calibration algorithm interfaces directly with the tracker's encoder system through RS-485 communication protocols, ensuring reliable data transmission with 6-bit CRC error detection. Further algorithm implementations are described in Section 3.3.3 and complete software architecture of the SUTA are documented in Appendix A.

3.3. Experimental Design

3.3.1. SUTA Baseline

Initial position-based experiments revealed that the SUTA's virtual joint configuration differed strongly from the real-life configuration. To obtain a proper baseline, relatively close to the true kinematic configuration, a manual parametrization was performed. The end-effector was mounted to a fixed position on the XZ plane of the SUTA's coordinate system. This position was then manually measured and used as a single ground truth position. The configuration was then sketched in Solidworks, with the corresponding link lengths shown in Figure 3.5. The configuration was fully defined by the measured position, resulting in a set of joint angles. From these joint angles, the encoder offsets were computed by inverting the kinematic computations, which serves as the baseline of 5.45 mm RMS for the methodology in this research.

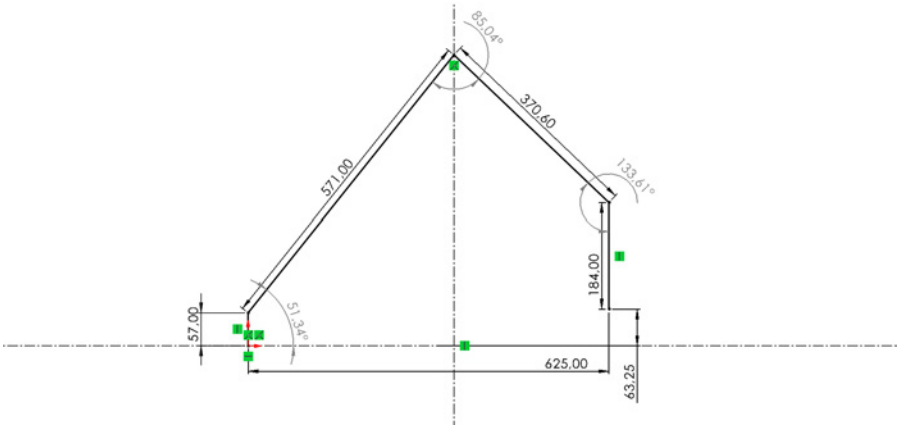


Figure 3.5: CAD sketch of the manually configured SUTA. From the link lengths and end-effector position, the joint angles were reverse computed to obtain a baseline of encoder offsets.

3.3.2. Data Acquisition Protocol

The calibration protocol follows an ISO 10360-12 inspired procedure [18] designed to capture comprehensive workspace sampling while maintaining measurement consistency. This standard defines testing protocols that obtain length measurements between two spheres at different configurations in the workspace. This research follows that same approach but uses the 15 internal landmarks of the artefact to obtain distance measurements. Each measurement session follows a systematic clockwise touch sequence: Origin (O) → Medium ring (M1-M6) → Large column (L1-L4) → Small row (S1-S4), shown in Figure 3.6. This ordering ensures consistent data structure during manual operation.

The artefact is positioned at five locations across the tracker workspace: central position A_O (512.5, 12.5, 0) mm shown in Figure 3.7 and four offset positions (C_A, C_B, C_C, C_D) shown in Figure 3.8. The original workspace was stated as a hemisphere with a radius of 100 mm as described in Table 3.1. Because full representation of this workspace is difficult to manufacture, the artefact was placed outside of this workspace to obtain an extensive set of unique kinematic configurations. Each landmark is measured three times per position, providing a robust set of 225 total measurements for calibration (15 landmarks × 3 repeats × 5 positions). This sampling density exceeds ISO 10360-12 requirements, which specify 25 points per sphere at two workspace positions, but is justified for a custom-built tracker arm that lacks the precision manufacturing and factory calibration of commercial systems. Moreover, this number enables the dataset size performance analysis performed in Section 4.3. Each measurement records all six encoders over a 2-second interval at 768 Hz, with mean values logged as the final measurement. This duration balances measurement noise reduction with practical operation speed. Each landmark contact includes counter-clockwise rotation locking before recording to ensure consistent rotational seating, compensating for the backlash between the end-effector and landmark plugs.

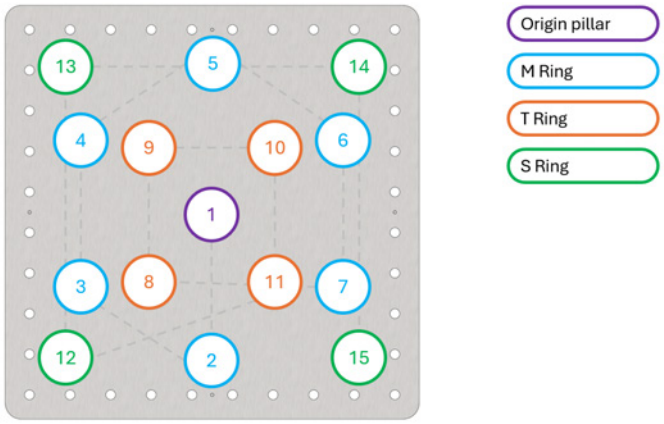


Figure 3.6: Systematic landmark measurement sequence: Origin → Medium ring (clockwise) → Tall column → Short column. Consistent ordering minimizes positioning errors and ensures repeatable data collection.

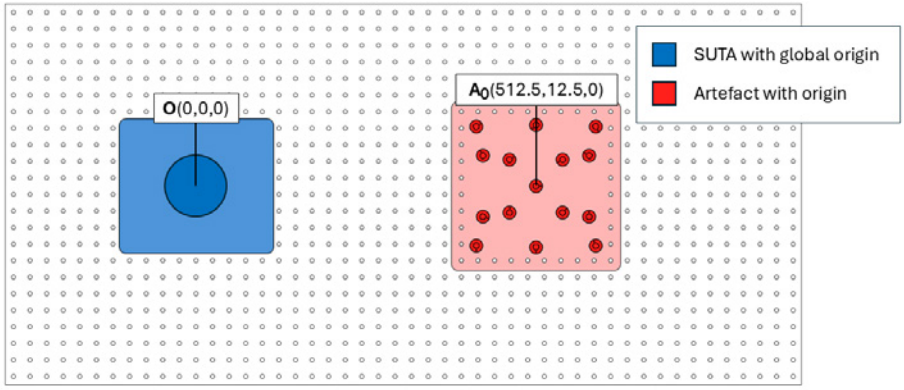


Figure 3.7: The SUTA base located at the global origin O and the artefact's standard position A_O located at (512.5, 12.5, 0). From this position, offset positions are used for calibration and independent validation.

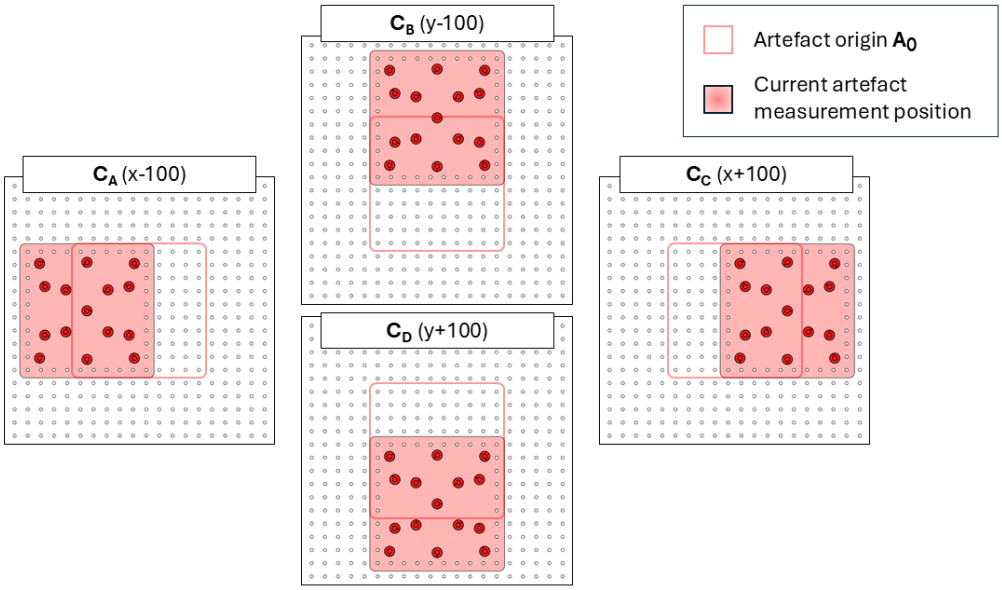


Figure 3.8: Four calibration artefact offset positions providing varied joint configurations for robust parameter identification. The hollow square represents the artefact standard position A_O and C_i represent the offset positions for calibration.

3.3.3. Calibration Algorithm

The calibration employs least-squares optimization combined with the Huber loss function ($\delta = 1.35$) as described in Section 2.2. The optimization targets link lengths (d_i) and joint angle offsets ($\delta\theta_i$) to account for manufacturing tolerances and joint angle deviations. The cost function minimizes pairwise distance errors between ground truth landmarks and measured positions:

$$\min \sum_{i,j} \rho \left(|p_{ij}^{measured} - p_{ij}^{truth}|_2 \right) \tag{3.1}$$

where p_{ij} represents the Euclidean distance between landmarks i and j. The Trust Region Reflective algorithm [32] solves this optimization problem with bounded constraints on the MDH parameters. Multiple (8) optimization attempts with randomized initial conditions ensure convergence to the global minimum rather than local minima.

3.3.4. Validation Procedure

Independent validation employs ISO 10360-12 compliance testing [18] with verification samples spanning the workspace using different sample positions than calibration. Four ground-plane artefact positions and one elevated configuration (110 mm) provide 225 validation measurements across varied joint configurations, shown in Figure 3.9. This approach ensures representative encoder combinations that may occur in service while providing independent assessment of calibration quality.

Performance metrics include RMS distance error and systematic bias analysis. Validation positions are computed by performing forward kinematics of recorded encoder values using optimized DH parameters, with errors calculated as pairwise distances between measured and ground truth landmarks

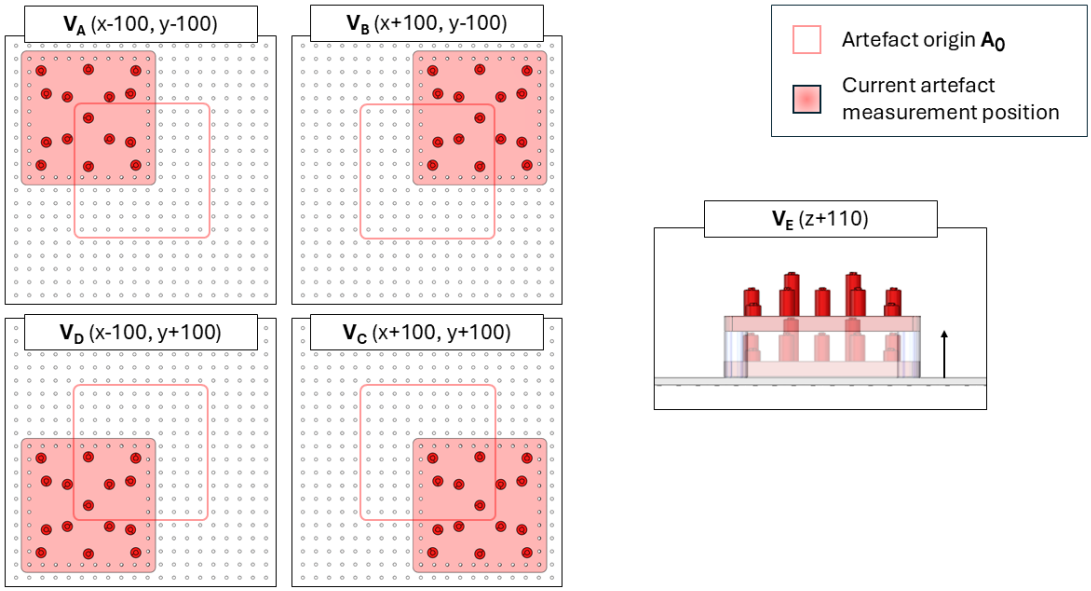


Figure 3.9: Independent validation artefact positions different from calibration locations. The hollow square represents the artefact standard position A_O and V_i represent the offset positions for calibration. Four ground-plane positions plus one elevated configuration (V_E) provide 225 validation measurements across varied joint configurations.

3.4. Data Processing

3.4.1. Performance Metrics

The primary performance metric employed throughout this research is Root Mean Square (RMS) error, calculated as:

$$\text{RMS} = \sqrt{\frac{1}{n}\sum_{i,j}\left(p_{ij}^{\text{measured}} - p_{ij}^{\text{truth}}\right)^2} \tag{3.2}$$

where p_{ij}^{measured} represents the Euclidean distance between landmarks i and j computed through forward kinematics, p_{ij}^{truth} is the laser-certified ground truth distance, and n is the total number of pairwise distance measurements. This distance-based formulation, as established in Section 2.2, is independent of global coordinate frames, eliminating the need for registration between tracker and artefact coordinate systems. Additional metrics include mean absolute error, standard deviation for precision, and maximum error to identify worst-case performance.

3.4.2. Dataset Size Optimization

To determine optimal calibration data requirements, a systematic analysis investigated the relationship between dataset size and calibration performance. The study performed 1000 total calibration trials: 100 independent calibrations for each dataset size from 1 to 10 measurement positions. The calibration algorithm was simplified by disabling the 8-fold configuration reset attempts described in Section 3.3.3.

The analysis procedure for each dataset size s was:

1. Randomly select s measurement positions from the complete calibration dataset
2. Perform full kinematic calibration using the sampled data
3. Validate calibrated parameters against an independent validation test set
4. Record the validation RMS

Each dataset size was tested 100 times with different random position selections to quantify both mean performance and calibration consistency (standard deviation). The standard deviation between calibration results was measured to test robustness. A standard deviation of 0 would mean that the calibration algorithm finds the same optimal solution each time.

3.4.3. Repeatability

After all validation measurements have been executed, each individual pillar position will have been visited 3 times. From this, the Cartesian repeatability is calculated. With the low number (3) of measurements for each position, this test does not represent the standard for repeatability tests but it could indicate the effect of the attempted compensation with rotational seating described in Section 3.2.4.

4

Results

This chapter presents the results of the calibration methodology applied to the SUTA system. The results demonstrate the effectiveness of the artefact-based calibration approach, achieving 81.5% RMS improvement from the baseline explained in Section 3.3.1. Section 4.1 describes the kinematic parameter optimization results. Section 4.2 presents the independent validation performance of the calibrated system. Section 4.3 explains how a comprehensive dataset size analysis is performed to determine optimal calibration requirements.

4.1. Calibration Results

The calibration algorithm successfully converged to an optimized set of DH parameters after a computation time of 30 seconds after roughly 1250 iterations. Table 4.1 shows the calibration deviations from the baseline parameters. The link length parameters d_i, a_i showed a mean deviation, calculated over the adjusted parameters, of 2.18 mm from initial values with the largest deviation of $\sim$10 mm (d_4). Joint angle offsets $\delta\theta_i$ showed a mean absolute deviation of 1.9° with an absolute range 0.6° to 4.1°. These deviations result in the final SUTA kinematic model described in Table 4.2.

Table 4.1: Deviations from initial Denavit Hartenberg parameters after calibration optimization, showing manufacturing tolerances and assembly corrections required for accurate kinematic modeling.

Joint	d_i **Deviation (mm)**	$\delta\theta_i$ **Deviation (deg)**
1	0.00	4.1
2	0.00	0.6
3	-0.13	-0.9
4	0.00	-1.7
5	-10.07	-2.4
6	0	1.7
7 (TCP)	2.9	-

Table 4.2: Final optimized Denavit Hartenberg parameters after artefact based calibration. These parameters represent the calibrated kinematic model of the SUTA system accounting for manufacturing tolerances and encoder angle deviations. The parameters included in the calibration are highlighted in bold.

Joint	θ_i *(rad)*	d_i *(m)*	a_i *(m)*	α_i *(rad)*	$\delta\theta_i$ *(deg)*
1	θ_1	**0.047**	0	$\frac{\pi}{2}$	**2.54**
2	θ_2	0	0	$-\frac{\pi}{2}$	**21.28**
3	θ_3	**0.571**	0	$\frac{\pi}{2}$	**200.44**
4	θ_4	0	0	$\frac{\pi}{2}$	**115.46**
5	θ_5	**0.361**	0	$\frac{\pi}{2}$	**86.23**
6	θ_6	0	0	0	**-9.88**
7	(TCP)	**0.254**	0	0	0

4.2. Validation Results

The calibrated system was validated using measurement data from five independent artefact positions outside of the calibration data. Table 4.3 presents the validation performance metrics comparing the baseline of manually-configured DH parameters against the optimized calibration results.

Table 4.3: A comparison of the SUTA independent validation results between the manual kinematic configuration and the artefact-based optimization calibration. It compares the performance RMS error, Mean error, Standard deviation and Max error for 525 pairwise distance measurements across five independent artefact positions.

System State	RMS (mm)	Mean (mm)	Std Dev (mm)	Max (mm)
Manual	5.45	4.02	3.68	21.73
Optimized	1.01	0.80	0.61	3.01
Improvement	**81.5%**	**80.1%**	**83.4%**	**86.2%**

The baseline system with manually-configured parameters achieved 5.45 mm RMS error across 525 pairwise distance measurements, with a standard deviation of 3.68 mm. After calibration optimization, the RMS error decreased to 1.01 mm with a standard deviation of 0.61 mm, representing 80.5% and 83.4% improvements respectively. The maximum error was reduced from 21.73 mm to 3.01 mm, while the mean error decreased from 4.02 mm to 0.80 mm.

Figure 4.1 shows the error distribution for both baseline and calibrated configurations. The reduced standard deviation in the calibrated system (0.61 mm vs 3.68 mm) indicates improved measurement consistency in addition to improved accuracy. The validation results show that the calibration optimization improved tracking accuracy from 5.45 mm to 1.01 mm RMS, reaching millimeter-level precision.

From the validation measurements, repeatability was computed according to Section 3.4.3. The rotational seating compensation for backlash resulted in a Cartesian repeatability of 0.29 mm RMS.

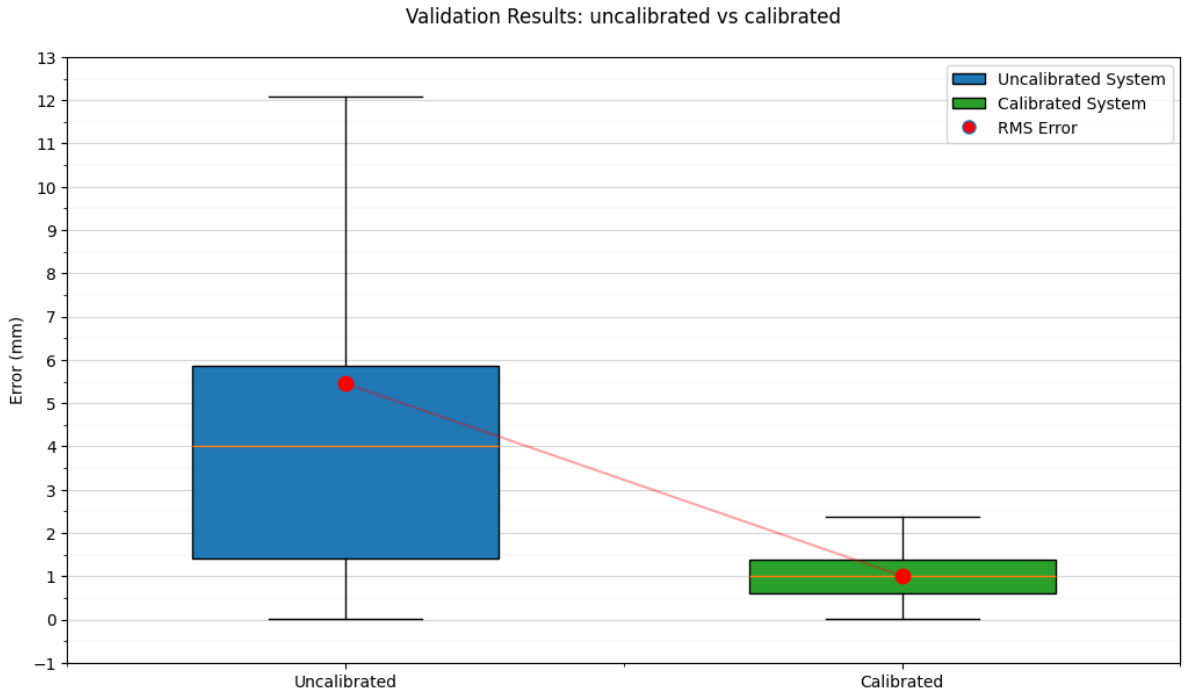


Figure 4.1: Validation error distributions for baseline (5.45 mm RMS) and calibrated (1.01 mm RMS) system configurations across 525 pairwise distance measurements.

4.3. Dataset Size Effects

The dataset size analysis investigated the relationship between calibration dataset size and tracking accuracy through 100 independent calibrations for each size from 1 to 10 measurement positions. Performance was evaluated using an independent validation dataset to compute RMS error for each calibration.

Table 4.4 presents the quantitative results, showing validation RMS decreased from 1.43 mm at size 1 to 1.03 mm at size 10, representing a 28.5% improvement. Equally important, the standard deviation decreased from 0.11 mm to 0.00 mm, indicating improved calibration consistency with larger datasets.

Table 4.4: Dataset size performance analysis showing continuous improvement in both calibration accuracy and reliability, with perfect consistency achieved at size 10.

Dataset Size	Mean RMS (mm)	Std Dev (mm)	Range (mm)	Improvement vs Size 1
1	1.43	0.11	1.27 – 1.60	Baseline
2	1.37	0.09	1.23 – 1.56	4.3%
3	1.35	0.09	1.21 – 1.53	6.0%
4	1.33	0.07	1.18 – 1.49	7.3%
5	1.20	0.07	1.06 – 1.43	16.5%
6	1.09	0.04	1.02 – 1.21	24.1%
7	1.06	0.03	1.01 – 1.16	26.0%
8	1.04	0.03	1.00 – 1.10	27.5%
9	1.03	0.02	1.01 – 1.07	28.2%
10	1.03	0.00	1.03 – 1.03	28.5%

Figure 4.2 illustrates the distribution of validation results across all dataset sizes. The figure shows two phases: rapid improvement from size 1 to 6, with the largest improvement from size 4 to 5 (1.33 mm to 1.20 mm). After size 6, the gains become marginal, where additional data provides less than 5% incremental improvement. The distributions show that with larger size, the calibration consistency increases, with a variability of 0.33 mm (1) to 0.00 mm (10) SD.

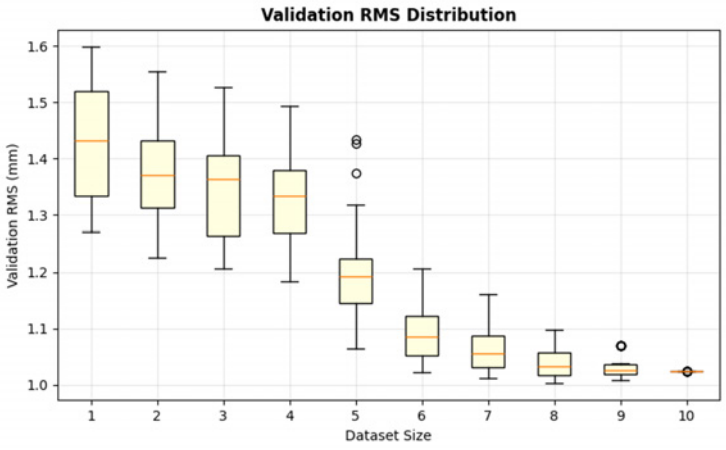


Figure 4.2: Distribution of validation RMS results across dataset sizes showing improvement in calibration consistency. Box plots demonstrate transition from highly variable results at small dataset sizes to perfect deterministic convergence at size 10 where all 100 calibrations yield identical results.

5

Discussion, Limitations and Future Work

5.1. Discussion

This research achieved 1.01 mm RMS accuracy for a passive mechanical tracking arm through artefact-based calibration, improving from a 5.45 mm RMS baseline. The baseline represents the SUTA system after correcting the initial incorrect kinematic model through manual configuration (Section 3.3.1). While the 81.5% improvement validates the calibration methodology, the final accuracy does not reach the 0.3-0.5 mm performance that state-of-the-art optical tracking systems achieve under laboratory conditions. The artefact-based calibration methodology addressed the fundamental challenge of calibrating passive tracker arms without sequential joint locking capability. The calibration algorithm consistently converged to optimal solutions across multiple trials at different starting configurations, indicating robustness of the methodology. In the following, the three research questions introduced in Section 1.2 are revisited and discussed in relation to the results of this thesis.

RQ1: What artefact geometry enables efficient and fully constrained six-degree-of-freedom calibration of a passive tracker arm? Each pillar consists of a cylindrical base equipped with an asymmetric plug at the top. This asymmetric plug prevents rotation of the SUTA end-effector, enabling recording of a fixed position and orientation at each measurement. This solved the fundamental limitation of traditional sphere-based methods that only constrain position. The 15 pillar arrangement across three height levels (20, 40, and 60 mm) within a 180x180x60 mm volume enabled systematic workspace sampling. This generated 105 pairwise distance constraints per measurement cycle, providing redundancy for a robust least-squares optimization. A practical compromise had to be made for the artefact's geometric design between ideal workspace coverage and manufacturing limitations. Ideally, pillars would be oriented perpendicular to a hemispherical surface, matching the SUTA's natural workspace. Manufacturing limitations however, required all pillars perpendicular to the baseplate, constraining the end-effector to be vertical for each measurement. An attempt to address this was made by repositioning the artefact at multiple locations around the workspace to obtain 75 unique kinematic configurations.

RQ2: What influence does calibration dataset size have on final tracking accuracy, and how can optimal data collection be determined? The dataset size analysis revealed a significant accuracy improvement up to 6 measurement positions, followed by marginal accuracy gains with the following 4 dataset sizes. The performance improved from 1.43 mm RMS at size 1, to 1.09 mm RMS at size 6, representing 24% improvement (Section 4.3. Adding more positions up to 10 yielded only 5.5% further improvement with 67% more measurement effort. Calibration consistency showed continuous improvement with dataset size. Standard deviation decreased from 0.11 mm at size 1 to 0.04 mm at size 6. Calibration reached perfect convergence at size 10 where all 100 independent calibrations obtained the same parameter results. The practical implication is that 6 measurement positions offer optimal balance between accuracy (1.09 mm) and efficiency for routine calibration, while applications requiring absolute repeatability should use the full 10-position protocol.

RQ3: What is clinically acceptable accuracy and how can the calibrated system be validated against both industrial and surgical standards? The SUTA's performance must be contextualized against existing surgical navigation technologies. According to the laboratory performance specifications presented in Table 2.1, the Brainlab Curve/Kick optical tracking system achieves 0.3 mm RMS, the NDI Aurora electromagnetic system reaches 0.48 mm RMS, and the YOMI mechanical system reports <1.00 mm at 95% in clinical experiments. The optical and electromagnetic systems are widely used in clinical applications, setting the benchmark for "clinically acceptable accuracy" of 0.3-0.5 mm RMS as explained in Section 2.3. The FARO Quantum Max industrial CMM achieves 0.024 mm MPE, where the RMS would be even lower but not determined from MPE alone, demonstrating that professional mechanical systems can achieve accuracy two orders of magnitude better than the SUTA. While these industrial CMM's are superior in accuracy, their cost would make research of this kind impossible. Clinical reality shows degraded performance compared to laboratory specifications, typically by a factor of 1.5-3 as discussed in Section 2.3. This indicates that the calibrated SUTA system, with 1.0 mm RMS accuracy under laboratory conditions, already approaches the error range that is achievable in clinical navigation. Based on this, the SUTA system would likely show performance in the 2-4 mm range under clinical conditions. The 768 Hz update rate of the SUTA system exceeds both optical systems (60-120 Hz) and electromagnetic tracking (40 Hz) by an order of magnitude. This higher frequency feedback could theoretically benefit applications requiring real-time instrument guidance or robotic control, where low latency and smooth trajectory tracking are critical. However, without dynamic tracking validation or trajectory following tests, this advantage remains theoretical. Moreover, clinical navigation systems must simultaneously render 3D anatomical models, display the axial CT views and perform registration updates. These tasks are computationally intensive and probably influence the theoretical advantages of high-frequency position updates. Mechanical tracking eliminates line-of-sight requirements that can be problematic when surgical staff and equipment obstruct the camera visibility. It also avoids metallic instrument interference that affects electromagnetic tracking. However, introducing a robotic arm into the operating room causes a physical obstruction directly connected to the patient, which could negatively influence the surgical workflow. With dual-robot operation however, this obstruction could be modeled with robotic collision avoidance methods.

Taken together, the results show that the artefact-based calibration methodology addressed the core challenge of constraining all degrees of freedom, identified an efficient calibration protocol, and demonstrated that the accuracy of a passive mechanical tracker can be significantly improved. However, there are clear limitations that must be addressed before clinical implementation can be considered. The mechanical interface and the lack of dynamic validation remain the primary barriers, and only after solving these can the methodology progress towards clinical translation.

5.2. Limitations

Mechanical interface compliance dominates the error budget, where the clearance contribution could be up to 0.5 mm of the total error. This clearance resulted from iterative manual machining during manufacturing, where achieving the precise fit for the asymmetric plug design was very challenging and required manual trial and error. Despite implementing counter-clockwise rotation for consistent seating, the interface showed 0.45 mm repeatability error over the separate repeatability experiment, described in Section 3.2.4 and extensively explained in Appendix E. The force dependency tests showed that consistent directional force improved repeatability by 50%, directly correlating with the clearance magnitude. While an attempt was made to compensate for this backlash by rotational seating during the validation measurements, the validation data showed an overall cartesian repeatability of 0.29 mm RMS (Section 4.2), still a possible 30% of the total RMS error. The design of the plug was difficult to manufacture and the end-effector was iteratively adjusted until it fit over each pillar. This single issue probably prevented achieving sub-millimeter accuracy for the SUTA system and a new interface design could show great improvements to the final performance.

The validation scope was limited to static point-to-point measurements. Surgical applications use continuous trajectory following with varying velocities and accelerations. Path following accuracy, latency and behavior during direction changes were not characterized. Without these tests, the practical benefit of high-frequency mechanical tracking over 60-120 Hz optical systems cannot be determined properly.

The methodology was validated only on the SUTA system. This prototype was originally built by a

bachelor student group and contained 3D-printed components rather than professional-grade parts and manufacturing. Testing on different passive tracker architectures would establish whether the artefact-based approach generalizes with other systems than this specific tracker arm. Moreover, this research was conducted entirely under controlled laboratory conditions. Clinical environments introduce temperature fluctuations, mechanical vibrations, and electromagnetic interference. The manual data collection process introduced human factors including variable contact forces and possible operator fatigue over 225 measurements per session.

5.3. Future Work

While this research does not offer the immediate solution to state-of-the-art tracking for surgical navigation, it shows promising future perspectives. The calibration methodology could offer accessible tracker arm calibration to maintain the mechanical tracking advantages. When tracking has been perfected, steps can be taken towards autonomous robotic surgery, where two robots work together to operate on the patient. Both of these robots could be working from the same base, with a direct transformation from the patient's anatomy to the autonomous robot's equipped instrument. This eliminates the need for continuous line-of-sight while working at very high (>500 Hz) sample rates.

Current mechanical trackers often do not publish their calibration methodologies, such as the YOMI or FARO system. For the benefit of future research, it is important to offer as much transparency as possible. With a custom-built tracking arm and a newly developed calibration method, many issues and advantages come to light which would otherwise remain hidden. By identifying research limitations and providing recommendations for future work, development of these innovations will be far more efficient. The limitations of this research have provided insights in what could be done to improve, and what steps need to be taken to take this research to a higher level. Thus, the following future recommendations are described:

The immediate priority is eliminating the interface clearance that dominates the error budget. Replacing the current plug with a kinematic coupling that does not contain backlash could remove seating variability and reduce the error budget by 30% to 50%. This is expected to bring the system into the 0.5 mm to 0.8 mm range under laboratory conditions. A solution could be where the seating of the end-effector is guided and then gradually locked in place with a sloped plug and slot. The plugs and socket from this research contained chamfers of 1 mm which did not offer an easy connection. Proper research into mechanical interface connections could offer a proper solution to solve this problem.

The second priority is dynamic validation to test whether the 768 Hz update rate provides an advantage in practice. Point-to-point measurements do not reflect surgical use, where instruments follow continuous trajectories with varying speed. Path following experiments and latency tests during rapid direction changes will show whether mechanical tracking is suitable for robotic guidance. With confirmation of this advantage, a second robot could be implemented to discover the latency of an autonomous system. If this computation shows great computational efficiency, together with a high tracker-to-instrument accuracy, the world of surgical robotics would see great changes. With autonomous surgical robot systems, human error and fatigue could become problems of the past.

The third priority is to establish generalisability and robustness. The calibration method was only tested on the SUTA system in a laboratory. Validation on other tracker architectures and professional-grade hardware would show whether the approach extends beyond this prototype. Long-term stability studies are also needed to quantify calibration drift and determine recalibration intervals, which would provide a grade of necessity for the calibration methodology of this research. Cadaver experiments could then provide the first realistic clinical evaluation by directly comparing mechanical tracking in laboratory conditions to actual clinical appliance.

This work shows that mechanical tracking can be calibrated to 1.0 mm accuracy. With improvements to the interface, validation under dynamic conditions, and broader generalisation, sub-millimeter performance and clinical translation are realistic next steps. Achieving this level of accuracy would place the system within the range of approximately 2-3 mm error that has been reported as clinically relevant for midface registration.

References

1. Reinhart, E., Friedl, W. & Hülsmann, M. Navigation in craniomaxillofacial surgery. *Journal of Oral and Maxillofacial Surgery* **66,** 1446–1454 (2008).
2. Scolozzi, P., Martinez, A., *et al.* Surgical Navigation: A Systematic Review of Indications, Treatments, and Outcomes in Oral and Maxillofacial Surgery. *Journal of Oral and Maxillofacial Surgery* **75,** 1987–2001 (2017).
3. Wani, G. & Faizullabhoy, M. *Surgical Navigation Systems Market - by technology, by application, by end use forecast, 2024 – 2032* tech. rep. (Oct. 2024). https://www.gminsights.com/industry-analysis/surgical-navigation-systems-market.
4. Schreurs, R. *et al.* Registration-free workflow for electromagnetic and optical navigation in orbital and craniofacial surgery. *Scientific Reports* **11.** https://doi.org/10.1038/s41598-021-97706-5 (Sept. 2021).
5. Ferreira, A. *et al. Latency in surgical navigation systems: implications for robotic integration* in *IEEE International Conference on Robotics and Automation (ICRA)* (2020), 9336–9342.
6. Gonzalez, D., Dey, R. & Joskowicz, L. Electromagnetic tracking accuracy in the operating room: Effects of metallic instruments. *Computer Assisted Surgery* **27,** 30–38 (2022).
7. King, F. *et al.* Performance assessment of the NDI Aurora electromagnetic tracking system. *Medical Physics* **39,** 6754–6764 (2012).
8. Metrology, F. V. *FAR Quantum Max 2M Portable CMM | Fox Valley Metrology* May 2025. https://www.foxvalleymetrology.com/sales/metrology-systems/portable-metrology-systems/portable-arm-coordinate-measuring-machines/faro-quantum-max-2-0m-portable-arm-coordinate-measuring-machine/.
9. Furuse, M., Ikeda, N., Kawabata, S., *et al.* Influence of surgical position and registration methods on clinical accuracy of navigation systems in brain tumor surgery. *Scientific Reports* **13,** 2644 (2023).
10. Siciliano, B. & Khatib, O. *Springer Handbook of Robotics* 2nd. ISBN: 978-3-319-32550-7 (Springer International Publishing, 2016).
11. Kiełczewski, K. & Gierlak, P. Programming of Industrial Robots Using a Laser Tracker. *Sensors* **22,** 6464 (2022).
12. Wikipedia. *Gauge block* Accessed January 2025. 2025. https://en.wikipedia.org/wiki/Gauge_block.
13. Technologies, F. *Measuring a Gauge Block or Gauge Ring Using the FaroArm and CAM2* FARO Knowledge Base Technical Documentation. 2023. https://knowledge.faro.com/Software/FARO_CAM2/CAM2/Measuring_a_Gauge_Block_or_Gauge_Ring_Using_the_FaroArm_and_CAM2.
14. Nubiola, A. & Bonev, I. A. Absolute robot calibration with a single telescoping ballbar. *Precision Engineering* **38,** 472–480 (2014).
15. Craig, J. J. *Introduction to Robotics: Mechanics and Control* 3rd. ISBN: 0-13-123629-6 (Pearson Prentice Hall, Upper Saddle River, NJ, 2005).
16. Hollerbach, J. M. & Wampler, C. W. The calibration index and taxonomy for robot kinematic calibration methods. *The International Journal of Robotics Research* **15,** 573–591 (1996).
17. Huber, P. J. Robust estimation of a location parameter. *The Annals of Mathematical Statistics* **35,** 73–101 (1964).
18. *ISO 10360-12:2016 Geometrical product specifications (GPS) – Acceptance and reverification tests for coordinate measuring systems – Part 12: Articulated arm coordinate measuring machines* International Organization for Standardization, 2016.
19. Franz, A. *et al.* Electromagnetic tracking in medicine—a review of technology, validation, and applications. *IEEE Transactions on Medical Imaging* **33,** 1702–1725 (2014).
20. *ISO 9283:1998 Manipulating industrial robots – Performance criteria and related test methods* International Organization for Standardization, 1998.

21. Mezger, U., Jendrewski, C. & Bartels, M. Navigation in surgery. *Langenbeck's Archives of Surgery* **398,** 501–514 (2013).
22. Block, M. S. & Emery, R. W. Implant accuracy using a robotically guided dental system: A prospective clinical study. *Journal of Prosthetic Dentistry* **126,** 435–442 (2021).
23. Yu, H. *et al.* An excellent navigation system and experience in craniomaxillofacial navigation surgery: a double-center study. *Scientific Reports* **6,** 28242 (2016).
24. Sozzi, D. *et al.* 15-Year Experience in Maxillofacial Surgical Navigation with Tracked Instruments. *Journal of Personalized Medicine* **13,** 10 (2024).
25. Widmann, G., Stoffner, R. & Bale, R. Errors and error management in image-guided craniomaxillofacial surgery. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* **107,** 701–715 (2009).
26. Conrad, E. U., Dai, E., Sippola, E., McKellop, H. & Sangeorzan, B. J. Accuracy and precision of a surgical navigation system: Effect of camera and patient tracker position and number of active markers. *The Open Medical Informatics Journal* **11,** 37–44 (2017).
27. Yaniv, Z., Wilson, E., Lindisch, D. & Cleary, K. Electromagnetic tracking in the clinical environment. *Medical Physics* **36,** 876–892 (2009).
28. Nubiola, A. & Bonev, I. A. Absolute calibration of an ABB IRB 1600 robot using a laser tracker. *Robotics and Computer-Integrated Manufacturing* **29,** 236–245. `https://doi.org/10.1016/j.rcim.2012.06.004` (July 2012).
29. *ISO 10360-2:2019 Geometrical Product Specifications (GPS)–Acceptance and Reverification Tests for Coordinate Measuring Machines–Part 2: CMMs for Measuring Linear Dimensions* tech. rep. (International Organization for Standardization, Geneva, Switzerland, 2019).
30. Zhao, H. *et al.* 3D artifact for calibrating kinematic parameters of articulated arm coordinate measuring machines. *Measurement Science and Technology* **29** (May 2018).
31. *Leica Absolute Tracker AT960 Data Sheet* Document 771288. Hexagon Manufacturing Intelligence (2016).
32. Virtanen, P., Gommers, R., Oliphant, T. E., *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17,** 261–272 (2020).

Appendices Overview

The following appendices document both the developmental process and final implementation of the calibration methodology. Appendices A and B present the iterative design evolution that informed the final approach, while Appendices C through F detail the validated methodology and its implementation. This developmental documentation is included to provide insight into design decisions and to assist future researchers in avoiding unsuccessful approaches.

A

Calibration and Validation Protocol for Future Research

To enable replication of the experiments performed in this research, this appendix described the full calibration and validation methodology.

A.1. Mounting of Components

All components are mounted on the NEXUS Optical Table, which employs 6 mm tapped mounting holes to secure the compontents.

1. Mount the SUTA to the table using 6 mm M6 hex screws and two washers. The first set of washers fits in the slot of the SUTA base, and the second set is larger to secure the base to the table.
2. Place the Artefact at 512.5 mm in the positive X direction and 12.5 mm in the positive Y direction from the base (the mounting holes are 25 mm apart and can be counted).
3. Plug in the power cable of the electronic unit to a 220 V wall socket.
4. Connect a laptop to the Arduino of the electronic unit via a micro-usb cable.

A.2. Data Acquisition

1. Begin by checking if the encoders are working properly by running *Rawdata.py*. If no values appear, check the encoder connections, this is often the issue due to handling and mounting of the SUTA.
2. To begin a measurement sequence, create a .json file with a unique name inside the calibration or validaiton folder (depending on which acquisition you are performing), then run *Measurement.py*.
3. The instructions in the script will ask the user to attach the end-effector to position O, then press enter.
4. The script will now run the user through the measurement sequence from Figure 3.6, holding position for 2 seconds until prompted to move to the next position.
5. After finishing the measurement sequence, copy the path of the .json file and paste it into the terminal.
6. The measurement set is now saved and the process can be repeated 3 times for each position (or just 1 if you want to follow the optimal 6 measurement sets of Section 4.3.
7. Move the artefact to the next position, described in Chapter 3 and perform the acquisition sequence again.
8. The number of datasets can be chosen by the operator, make sure to separate the calibration and validation sets to ensure a reliable validation.

A.3. Calibration and Validation

1. Ensure that the calibration data files are saved in the calibration folder and the validation data files in the validation folder.
2. To perform calibration, run the *Calibration.py* script and wait until calibration is finished.
3. The script will have saved the calibrated DH parameters in *calibrated_dh_parameters.json"*. A report will be shown with the calibration results.
4. To perform validation, run the *validation.py* script. The validation results will be shown with plots and printed in the terminal.

This concludes the calibration and validation protocols for future research.

B

Tracker Arm Design and Validation

This appendix provides complete details of the tracker arm construction, building upon the BEP research foundation and providing structural validation results. An overview of the CAD model and workspace is shown in Figure B.1; the yaw–pitch–roll convention is illustrated in Figure B.2.

B.1. BEP Research Foundation

The tracker arm design is based on the Bachelor End Project (BEP) from BSc Mechanical Engineering at TU Delft, developed by Klein Goldewijk, R. van Hartingsveldt, T. Grooff and W. Wolterbeek Muller. Due to its relatively short development timeline, the tracker arm required refinement within an ISO-grade calibration framework (see Figure B.1 for the clamped setup and workspace hemisphere). The end-effector axis convention used throughout this work follows Figure B.2.

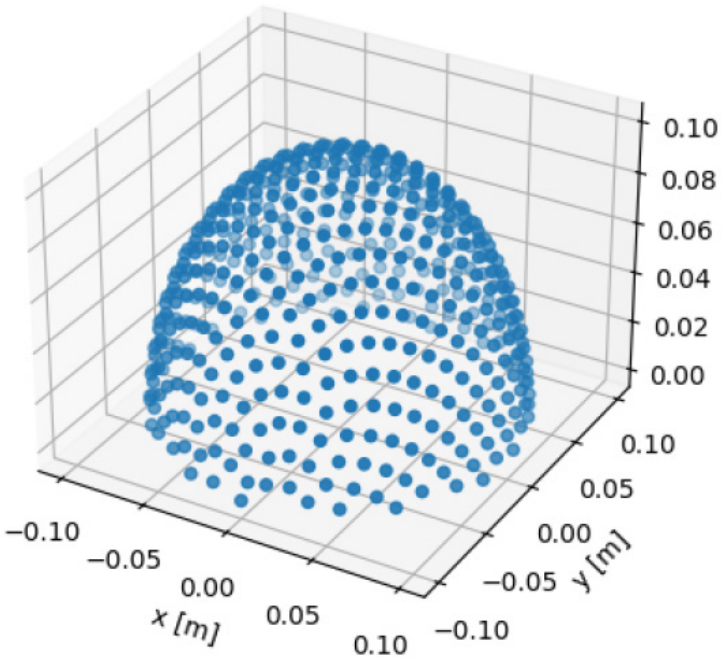


Figure B.1: Overall CAD model of the tracker arm clamped to the theatre rail. The translucent hemisphere visualizes the required 100 mm surgical workspace established by the BEP research.

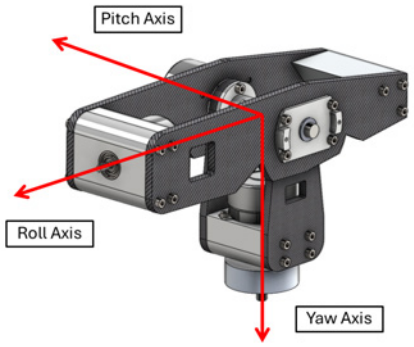


Figure B.2: Yaw-pitch-roll axis convention of the tracker-arm end-effector established by the BEP research.

B.2. System Grounding Requirements

The tracker arm is equipped with 6 wires running from each joint encoder to the main unit. Without proper grounding, static buildup in the links would discharge through encoder wires, disabling transceivers in the main unit. After several transceiver replacements, all links were grounded to enable safe static electricity discharge. This adjustment eliminated electrical problems and ensured reliable encoder operation (configuration as in Figure B.1).

B.3. Kinematic Architecture

The system employs a six-joint serial revolute configuration in a P-R-P-R-R-P roll-pitch-yaw sequence maintaining tool-normal orientation throughout the workspace. The link dimensions and joint specifications provide the foundation for all calibration work; the axis convention in Figure B.2 is used for all references.

C

Artefact Design Evolution

This appendix documents the iterative development process from initial prototypes to the final steel artefact design. Sections B.1 and B.2 describe preliminary experiments that identified critical design requirements but do not represent the final methodology's performance. Section B.3 presents the final artefact design used for all results reported in Chapter 4.

C.1. Preliminary Experiment A: Position-based Calibration on 3D-printed Artefact

Note: This preliminary experiment with 3D-printed materials identified fundamental limitations that informed the final design requirements. Results from this experiment are not included in the final methodology validation.

C.1.1. Objective

From the Zhao et al. study [30], the first calibration artefact design emerged. This artefact was designed for use with the original tracker arm end-effector (Figure C.1).

C.1.2. Artefact Design

The artefact was composed of 15 cylinders at 3 different heights (Z) and 15 different XY positions spanning the required workspace on a flat mounting surface. These points formed a labeled landmark map used as ground truth for initial robot kinematic model calibration. The original spherical end-effector and the first-generation artefact are shown in Figure C.1 and Figure C.2, respectively.

Figure C.1: Shows a front- and isometric CAD view of the original spherical end-effector developed by the BEP research.

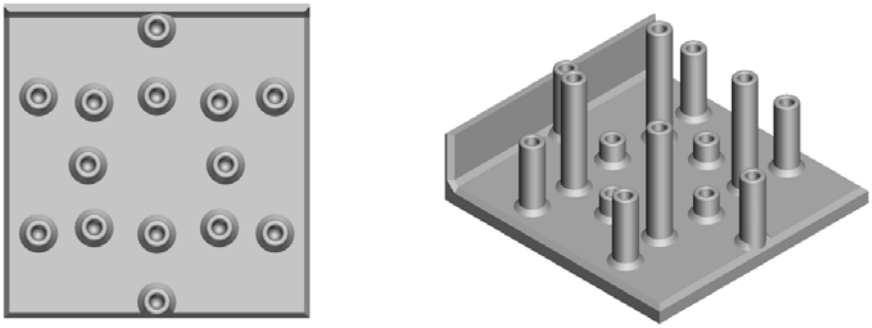


Figure C.2: Top- and Isometric CAD view of first-generation 3D-printed cylindrical calibration artefact with 15 landmarks at three heights, used in Preliminary Experiment A.

C.1.3. Experimental Procedure

With this artefact, the tracker arm end-effector position could be calibrated using pairwise point distances and simple least-squares optimization. Each landmark was visited consecutively, repeated 3 times. Raw encoder data was logged for all 6 joints, and RMS error was calculated for each paired distance (setup as in Figure C.2).

C.1.4. Identified Limitations

This preliminary experiment revealed three critical shortcomings that necessitated redesign:

1. The artefact only calibrated the first 5 joints, neglecting the 6th rotational joint and losing 6-DoF functionality
2. 3D-printed construction allowed bending under tracker arm forces (see Figure C.10)
3. Dimensional variations in 3D-printed holes created uncertainty exceeding target accuracy

These limitations led to the development of an orientation-aware design in Preliminary Experiment B.

C.2. Preliminary Experiment B: Orientation-based Calibration on 3D-printed Artefact

Note: This experiment established the need for 6-DoF constraint but revealed material limitations of 3D-printed construction. The asymmetric constraint principle developed here was carried forward to the final steel design.

C.2.1. Enhanced End-effector Design

Version 2 was designed to accommodate end-effector orientation through asymmetric plug design. The new end-effector could attach with single orientation constraint, enabling 6-DoF measurement capability (Figure C.3 and Figure C.4). The full setup on the NEXUS table is shown in Figure C.5.

Figure C.3: Front- and Isometric CAD view of the v2 end-effector, enabling measurement of the end-effector orientation.

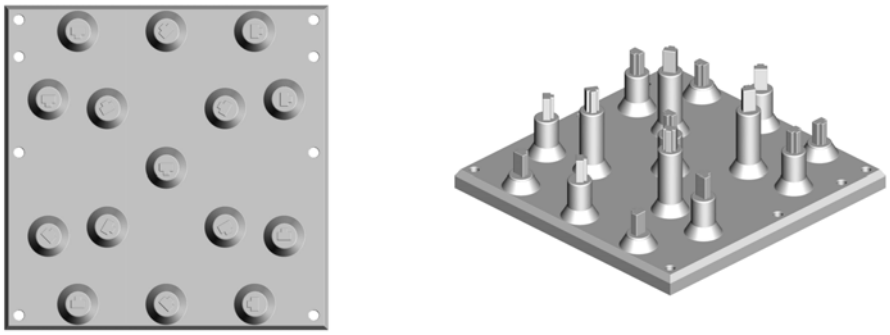


Figure C.4: Top- and Isometric CAD view of keyed pillar plug and matching single-orientation end-effector.

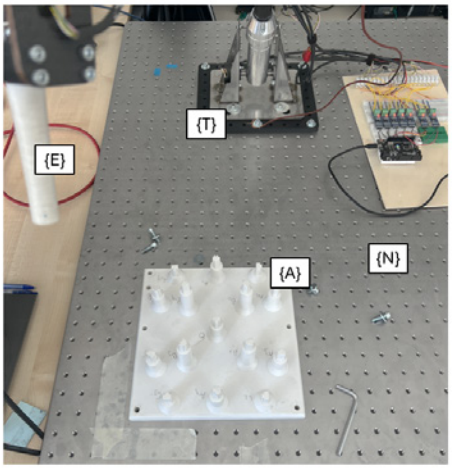


Figure C.5: Second-generation artefact A accommodating the tracker arm T end-effector E orientation via asymmetric plugs, mounted on the NEXUS table N.

C.2.2. Configuration Resolution

Initial experiments revealed that virtual joint configuration differed completely from real-life configuration (Figure C.6). This was resolved by the steps listed below, aided by the sketches and comparisons shown in Figure C.7 and Figure C.8, and finalized with the end-effector frame correction demonstrated in Figure C.9.

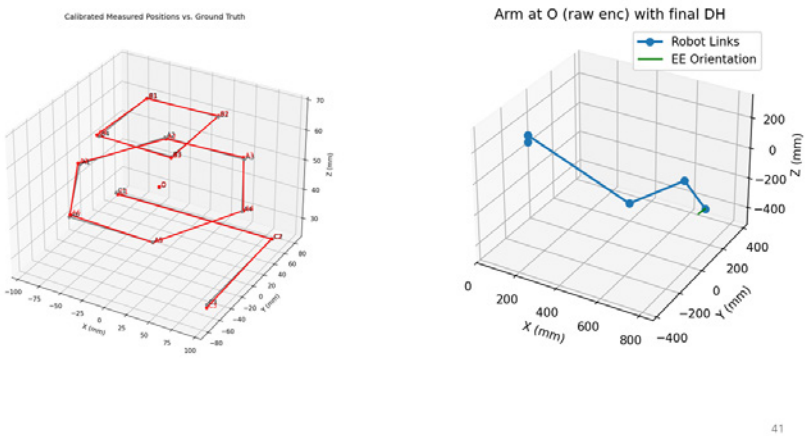


Figure C.6: Incorrect virtual joint configuration revealed after initial orientation-aware calibration.

1. Mounting artefact at known global position relative to tracker arm base
2. Using SolidWorks sketch to mimic kinematic configuration (Figure C.7)
3. Reverse-computing required joint angles and signs
4. Correcting end-effector frame axis computation (Figure C.9)

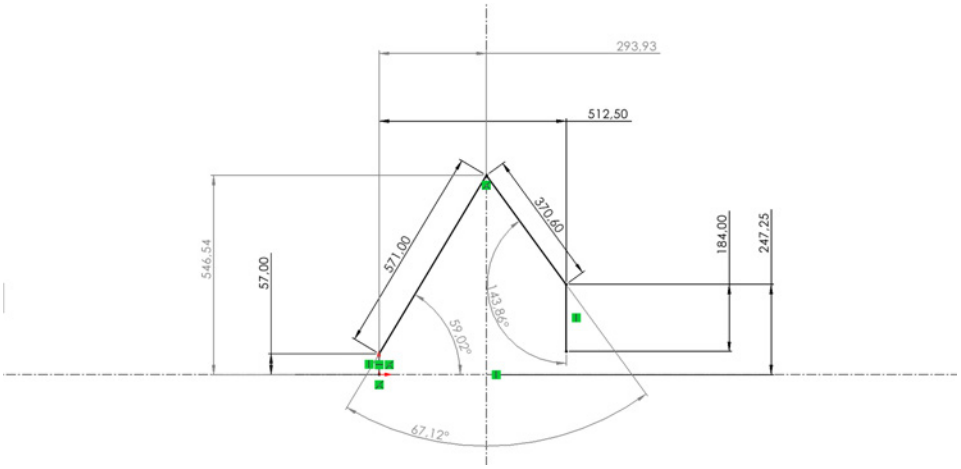


Figure C.7: Rough configuration sketch based on real world position approximation.

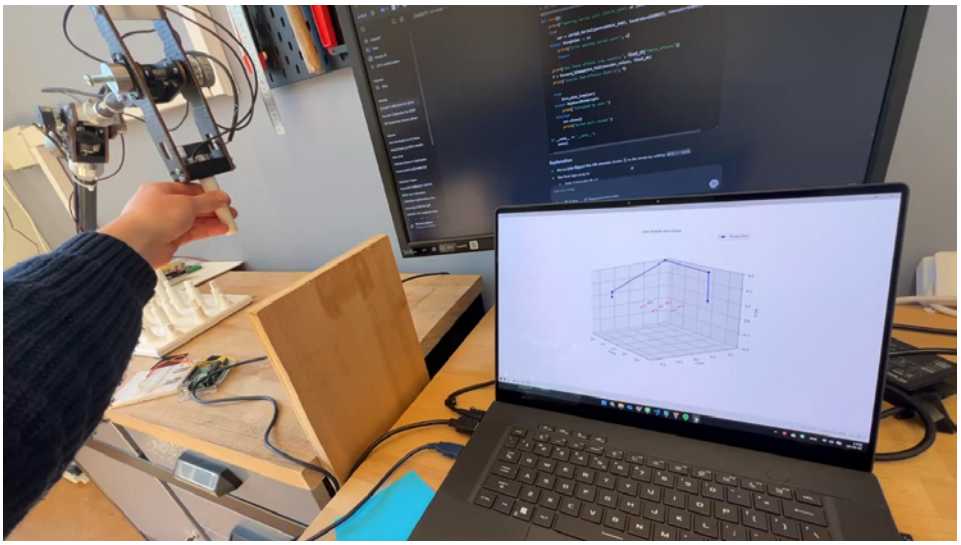


Figure C.8: Visualization of virtual- vs real tracker arm configuration.

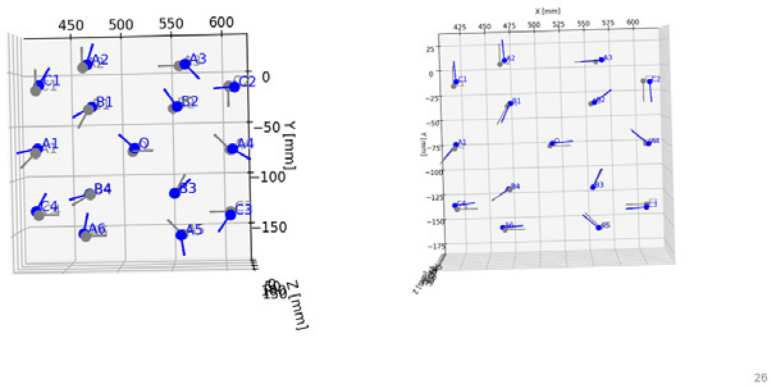


Figure C.9: Before-and-after orientation vectors after fixing the end-effector frame axis.

C.2.3. Material Limitations

The 3D-printed artefact deteriorated quickly with significant backlash between end-effector and landmark plugs. Heavy tracker arm components and manual experiments caused landmarks to bend or break (see Figure C.10), with measured deformations exceeding 2 mm under normal operating forces. This magnitude of deformation was incompatible with the target millimeter-level accuracy, necessitating transition to more robust materials.

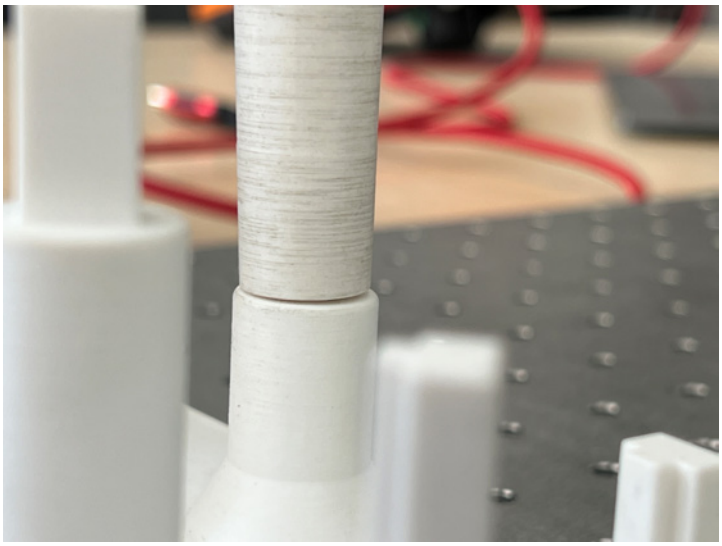


Figure C.10: Wear and bending damage observed on 3D-printed pillars after repeated use. Visible deformation exceeded 2mm, incompatible with millimeter accuracy requirements.

C.3. Final Design: Steel Artefact Implementation

This section describes the final artefact design used for all calibration and validation results presented in Chapter 4. This design incorporates lessons learned from preliminary experiments while achieving the required dimensional stability and wear resistance.

C.3.1. Design Requirements

Based on limitations identified in preliminary 3D-printed experiments, the final steel design requirements were established:

- Dimensional stability under repeated loading (<0.01 mm deformation)
- Wear resistance for long-term use (>1000 measurement cycles)
- Precision machining capability for tight tolerances (±0.05 mm)
- 6-DoF constraint through asymmetric plug geometry (single unique orientation)

The resulting design is shown in CAD form (Figure C.11) and as manufactured (Figure C.12).

C.3.2. Material Selection

S235JR steel was selected for the final artefact based on:

- Superior dimensional stability compared to polymers (Young's modulus 200 GPa vs 2-4 GPa)
- Machinability for complex constraint features using standard CNC equipment
- Resistance to deformation under tracker arm forces (<0.01 mm under 10N load)
- Cost-effectiveness for research applications (€300 material cost)

C.3.3. Manufacturing Process

The final artefact was manufactured using:

- 15 mm S235JR steel baseplate (200×200×15 mm)
- Cylinder welding and positioning to ±0.1 mm tolerance
- DMG MORI MillTap 700 precision machining
- Asymmetric plug feature creation with 0.05 mm clearance
- Quality control and dimensional verification using coordinate measuring machine

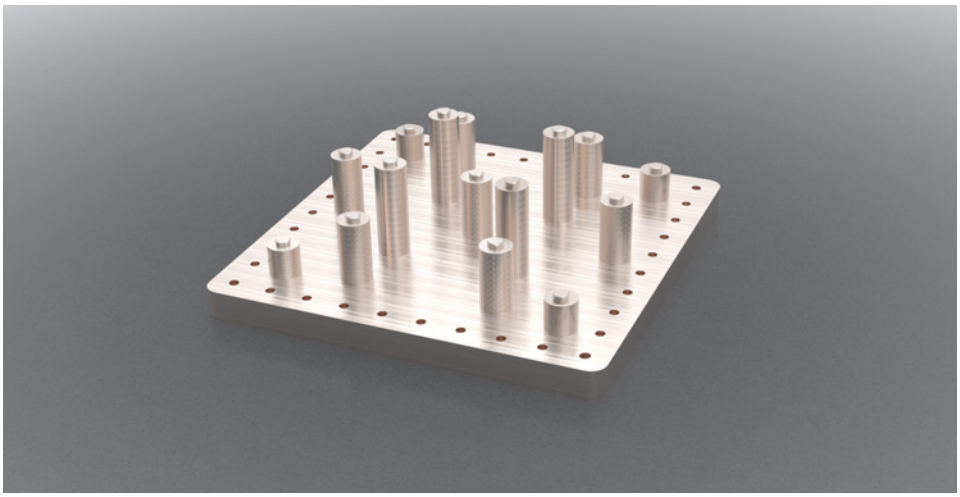


Figure C.11: Final steel artefact CAD design showing 15 precision landmarks used for all results in Chapter 4.

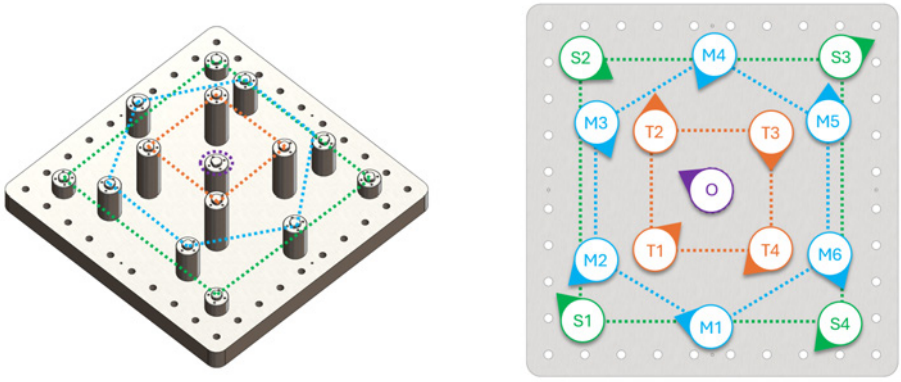


Figure C.12: Manufactured steel artefact showing precision machined features. This artefact achieved the 1.01 mm RMS accuracy reported in the main results.

D

Laser-Tracker Certification Protocol

This appendix details the certification procedure for the final steel artefact, establishing the ground truth reference with ±0.001 mm uncertainty for all calibration and validation results presented in Chapter 4.

D.1. Certification Overview

The artefact geometry serves as the length standard for every calibration and validation routine. Absolute correctness matters less than knowing its deviations precisely. The plate was fully certified with a Leica AT960 laser tracker in collaboration with André Mendes Florindo at SamXL metrology laboratory (setup in Figure D.1).

D.2. Equipment and Setup

All measurements used the AT960's T-probe with certified accuracy of ±0.001 mm:

- 6 mm ruby tip for surface touches
- 3 mm needle tip for small reference holes
- Siegmund 22 Professional Extreme Welding Table as reference surface (flatness ±0.02 mm)

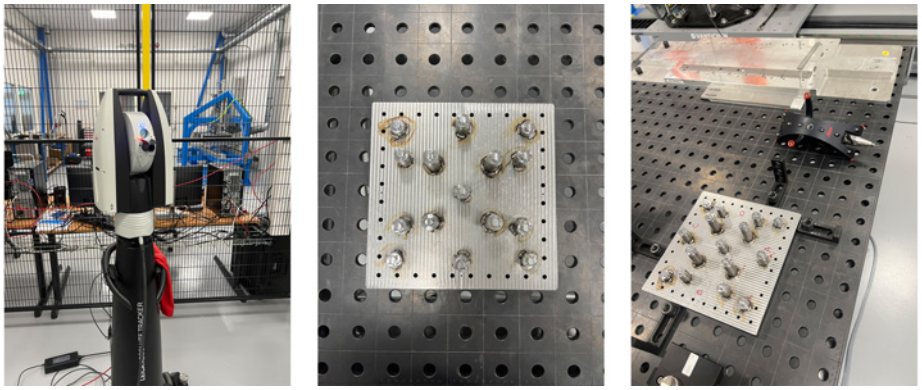


Figure D.1: The Leica AT960 laser tracker [left], the steel calibration artefact [middle] and the mounted artefact with the T-probe [right].

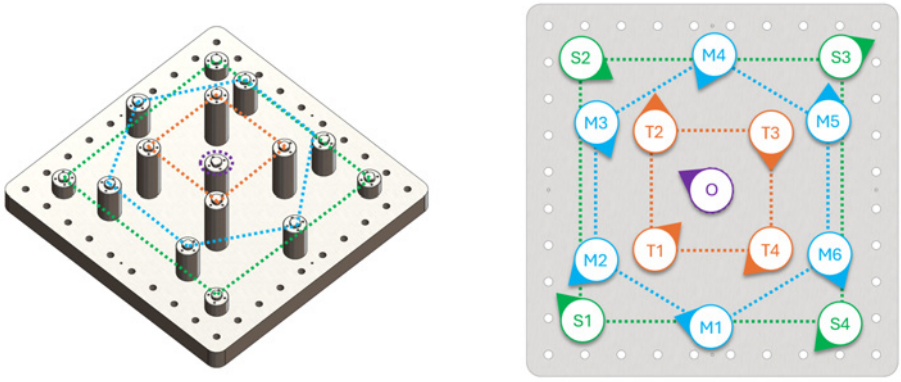


Figure D.2: Top- and isometric view of the final steel artefact design showing 15 landmarks at three height levels.

D.3. Five-Step Certification Protocol

D.3.1. Step 1: Ground Plane Establishment

Ten points were captured over the 250×250 mm space covered by the artefact. A best-fit plane defines the XY-plane for the entire certification session. Subsequent coordinates were automatically projected onto this plane by tracker software (reference surface shown in Figure D.3).

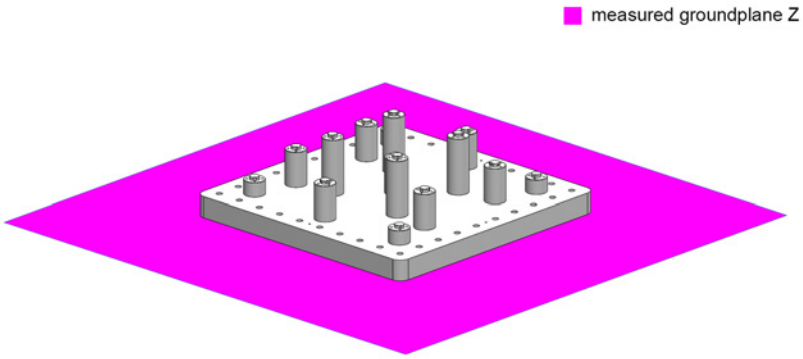


Figure D.3: Siegmund 22 welding table reference surface measured using 6 mm ruby tip.

D.3.2. Step 2: Pillar Heights

The flat tops of all fifteen pillars were touched in a star pattern (five points each) to yield Z-coordinates with high statistical confidence. Mean standard deviation across all pillars was 0.003 mm (Figure D.4).

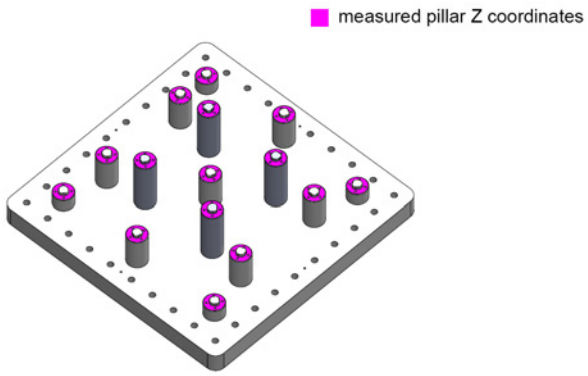


Figure D.4: Surfaces used for five-point sampling of the Z coordinates.

D.3.3. Step 3: XY Centers via Reference Holes

Switching to 3 mm probe tip, four 3 mm calibration holes around each pillar were touched. The intersection of diagonal lines through opposite holes defines the XY-center of each pillar with ±0.01 mm uncertainty (Figure D.5).

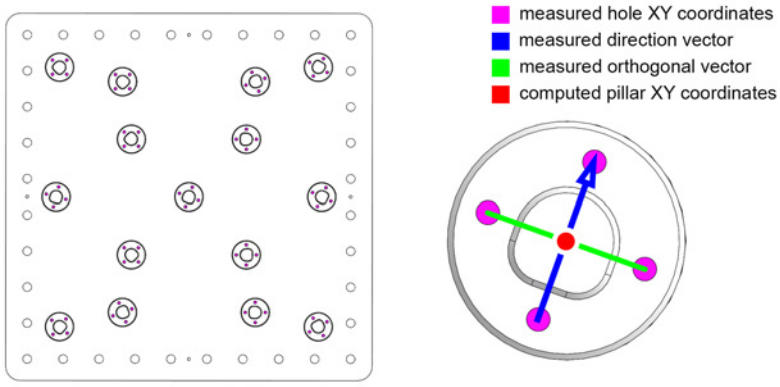


Figure D.5: XY-center determination using the four reference holes.

D.3.4. Step 4: Pillar Orientation

Front- and back-faces of cup flats were touched with 6 mm tip. The vector between these points gives the in-plane orientation of each pillar with angular uncertainty of ±0.1° (procedure area shown in Figure D.5).

D.3.5. Step 5: Side-Mount and Spacer Verification

The milled flank was probed in grid fashion to certify ”side-on” configuration for elevated mounting, and four aluminum spacers were measured end-to-end to validate 110 mm elevation height (Figure D.6 and Figure D.7).

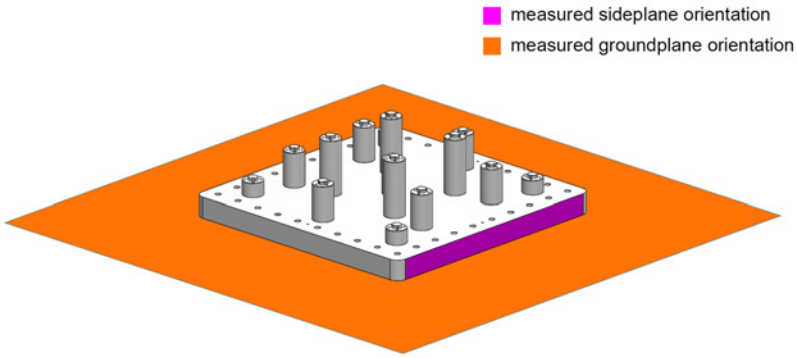


Figure D.6: Side-plane sampling for angular offset determination.

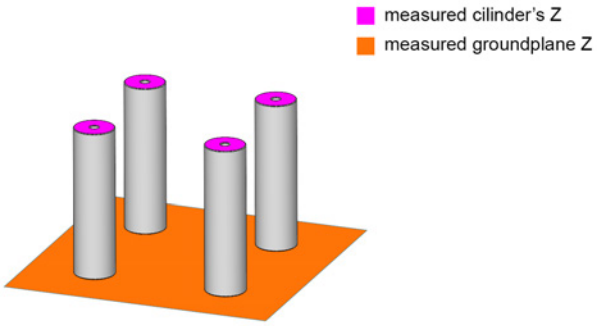


Figure D.7: Spacer-height verification using 6 mm tip for elevated mounting configuration.

D.4. Certification Results

The certification achieved 0.19 mm RMS positional accuracy across all 15 landmarks when excluding the intentionally damaged T1 landmark. This provides ground truth coordinates with uncertainty one order of magnitude better than the target tracker accuracy of 1 mm, satisfying the metrological requirement that reference standards should be 10× more accurate than the system being calibrated.

E

Backlash Compensation Methodology

This appendix documents the comprehensive repeatability analysis and backlash compensation development that forms part of the final calibration methodology, following ISO 9283 guidelines.

E.1. Rotational Backlash Protocol

Due to the 0.05 mm clearance between conical stylus and pillar cup (necessary for reliable seating), each touch was seated with counter-clockwise twist before recording. This procedure produced consistent rotational bias that was included in the calibration model (Figure E.1).

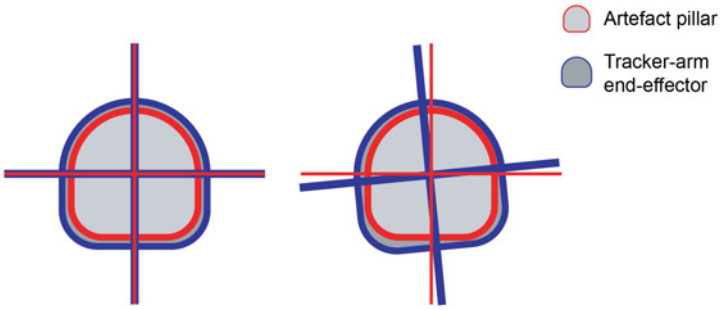


Figure E.1: Counter-clockwise twist procedure for consistent rotational offset compensation. This technique reduced measurement variability by 53%.

Figure E.2: Front- and isometric CAD view of the final steel end-effector with debossed interface profile used for all results in Chapter 4.

E.2. Systematic Repeatability Testing

E.2.1. Test A: Encoder Sanity Check

The tracker arm was attached to landmark O and left stationary for 2 minutes to verify encoder stability. Results showed ±0.00004° to ±0.00013° precision with no signs of encoder noise, confirming that measurement uncertainty originates from mechanical rather than electronic sources.

E.2.2. Test B: Regular Repeatability (ISO 9283)

Thirty measurements of landmark O showed repeatability of 0.094° RMS across all six joints. Individual joint repeatabilities ranged from 0.020° (Joint 3) to 0.14° (Joint 1), with mean of 0.085° ± 0.041°. This translates to 0.95 mm RMS Cartesian repeatability without force control.

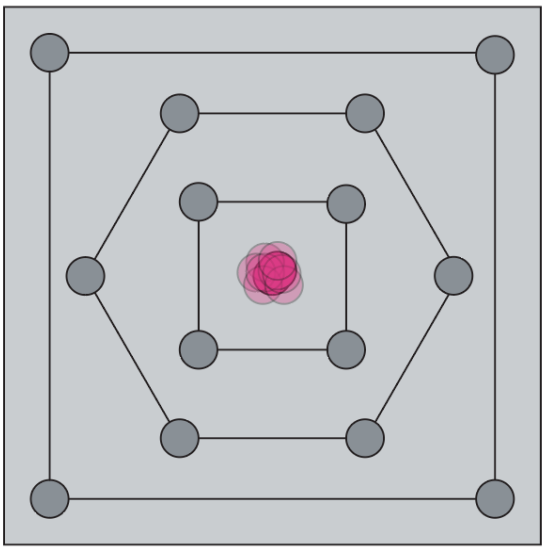


Figure E.3: Repeatability test setup showing landmark O used for 30 consecutive measurements per ISO 9283.

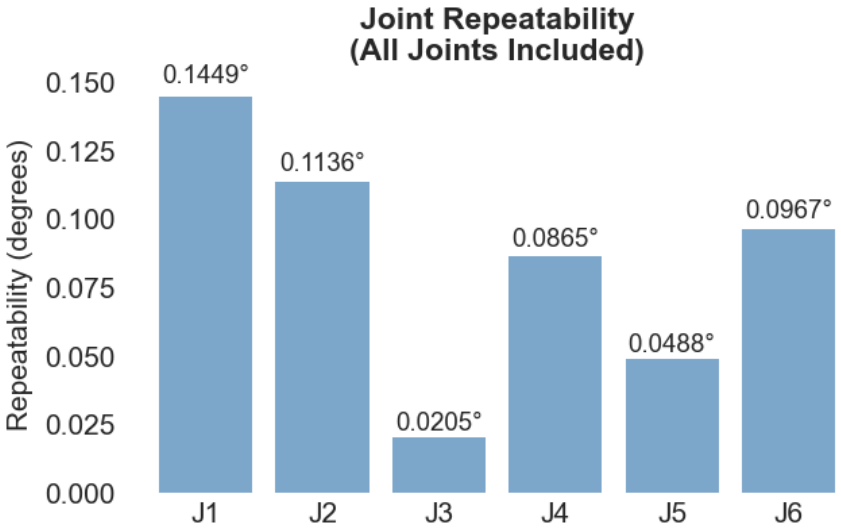


Figure E.4: Individual joint repeatability performance showing Joint 1 as the primary source of system uncertainty (0.145°) and Joint 3 achieving highest precision (0.020°).

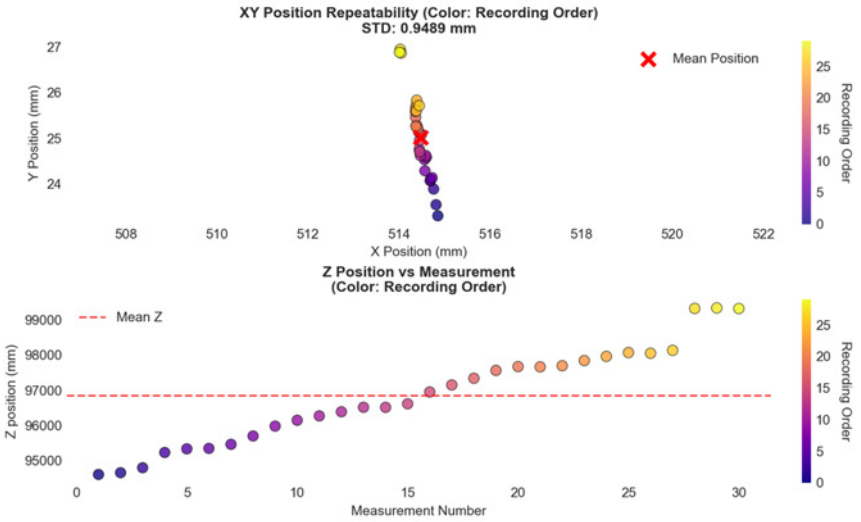


Figure E.5: Cartesian repeatability performance without force control: 0.95 mm RMS.

E.2.3. Test C: Directional Force Analysis

End-effector was pushed in 5 directions (North, South, East, West, Down) with approximately 5N force. Peak-to-peak angular deviations ranged from 0.19° (Joint 1, East-West) to 2.9° (Joint 6, North-South), showing direct correlation between joint rotation plane and highest deviation direction.

E.2.4. Test D: Single-Directional Force Repeatability

Consistent south-direction force application (approximately 5N) resulted in 0.067° repeatability RMS, a 30% improvement over Test B. Cartesian repeatability improved to 0.45 mm RMS (53% reduction), demonstrating that mechanical compliance dominates the error budget (Figure E.6).

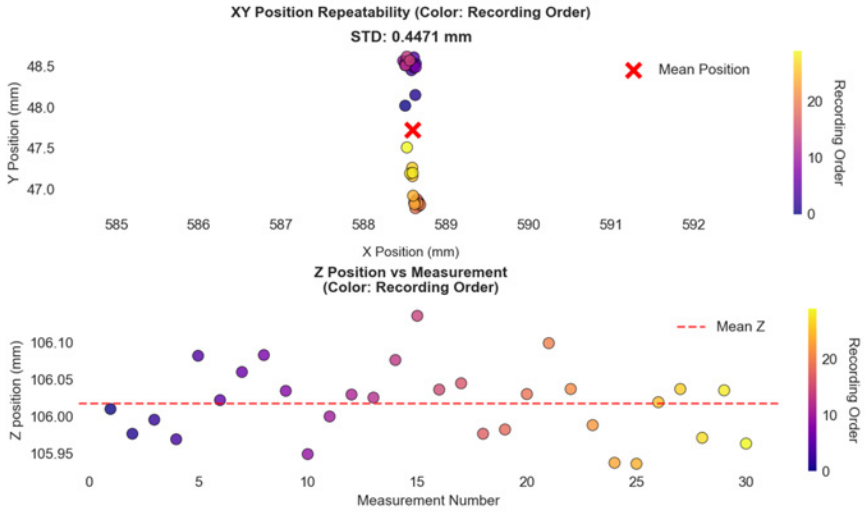


Figure E.6: Cartesian repeatability with consistent directional force: 0.45 mm RMS (53% improvement).

E.3. Compensation Strategy

Based on these findings, the calibration protocol incorporates:

1. Counter-clockwise rotation before each measurement for consistent seating
2. 2-second averaging to reduce measurement noise

This compensation strategy is integral to achieving the 1.01 mm RMS accuracy reported in Chapter 4.

F

Multi-Metric Outlier Detection Framework

This appendix provides the mathematical framework and implementation details for the multi-metric outlier detection methodology that forms part of the final calibration approach.

F.1. Framework Development Rationale

Due to the mechanical interface variability (0.05 mm clearance with ±0.45 mm repeatability), single-metric outlier detection proved too aggressive, removing valid measurements that reflected interface variability rather than true systematic errors. A multi-metric approach was developed to identify genuinely problematic measurements.

F.2. Four-Metric Statistical Framework

F.2.1. Error Contribution Analysis

Each pillar's contribution to total calibration error was calculated as:

$$\text{Contribution}_i = \frac{\sum_j |e_{ij}|^2}{\sum_{i,j} |e_{ij}|^2} \times 100\%$$

With 15 pillars, each should contribute approximately 6.7% in well-distributed datasets. Contributions exceeding 10% (1.5× expected) indicated potential problems requiring further investigation.

F.2.2. Cook's Distance Analysis

Cook's Distance measures each pillar's influence on calibration parameters:

$$D_i = \frac{(\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\hat{\boldsymbol{\beta}}_{(i)} - \hat{\boldsymbol{\beta}})}{p \cdot MSE}$$

where $\hat{\boldsymbol{\beta}}_{(i)}$ represents parameters estimated without pillar i. Values above 0.2–0.267 (4/n threshold) indicate high leverage points disproportionately influencing the model.

F.2.3. Leave-One-Out Analysis

Direct quantification of each measurement's impact on overall accuracy:

$$\text{LOO Impact} = \frac{RMS_{full} - RMS_{(-i)}}{RMS_{full}} \times 100\%$$

Improvements exceeding 5% upon exclusion indicate measurements that significantly degrade calibration quality.

F.2.4. Systematic Bias Magnitude

Consistent measurement errors were identified using:

$$\text{Bias}_i = \sqrt{\frac{1}{n_i}\sum_{j=1}^{n_i}(\mathbf{p}_{measured,ij} - \mathbf{p}_{truth,i})^2}$$

Bias exceeding 1.0 mm for millimeter-accuracy systems indicated systematic mechanical problems requiring investigation.

F.3. Implementation Results

Analysis of the calibration dataset identified landmark S2 as problematic across all four metrics:

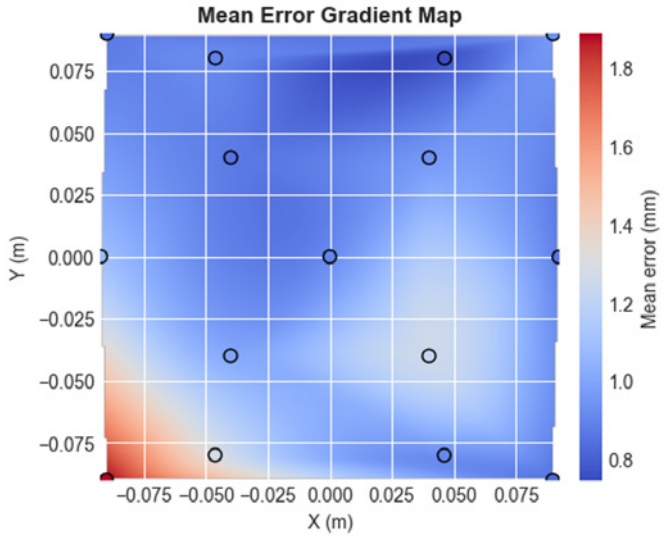


Figure F.1: Spatial error heat-map showing dominant contribution of landmark S2 to calibration error.

- Error Contribution: 13.2% (nearly 2× expected 6.7%)
- Cook's Distance: 0.22 (>5× higher than next highest)
- Leave-One-Out: 6.1% improvement upon removal
- Systematic Bias: 2.0 mm (54% higher than next largest)

S2 exclusion during preliminary analysis suggested a 26.6% RMS improvement. However, the final methodology retains all landmarks to maintain calibration robustness, accepting slightly higher error for improved generalization.

G

Experimental Results Summary

This appendix presents the complete experimental progression, including preliminary experiments that informed methodology development. The final validated system performance of 1.01 mm RMS (81.5% improvement from baseline) is presented in Chapter 4 of the main text.

G.1. Developmental Progression

The experimental development proceeded through several stages, each addressing specific limitations identified in previous iterations:

Figure G.1: Incremental improvement analysis showing performance changes with each refinement. Note that experiments 1-2 represent preliminary investigations, not final methodology.

Table G.1: Complete experimental progression including preliminary experiments (1-2 with 3D-printed artefacts) and final methodology validation (3C-5 with steel artefact). The final system achieves 1.01 mm RMS as reported in Chapter 4.

Experiment	Artefact	Key Finding	RMS (mm)	Status
1: Initial test	3D-printed	Need for 6-DoF constraint	N/A	Preliminary
2: 6-DoF test	3D-printed	Material inadequacy	N/A	Preliminary
3A: Certification	Steel	Ground truth established	0.194	Validation
3B: Repeatability	Steel	Backlash compensation needed	0.447	Final method
3C: Baseline	Steel	Manual configuration baseline	5.45	Baseline
4: Calibrated	Steel	Optimization successful	1.01	**Final result**