HARMONI at ELT: an evolvable software architecture for the instrument pointing model

Gonzalo José Carracedo Carballal\textsuperscript{a}, Javier Piqueras López\textsuperscript{a}, and Fraser Clarke\textsuperscript{b}

\textsuperscript{a}Centro de Astrobiología (CSIC/INTA), 28850 Torrejón de Ardoz, Madrid, Spain
\textsuperscript{b}Dept. of Astrophysics, University of Oxford, Keble Road, Oxford, OX1 3RH, U.K.

ABSTRACT

HARMONI is the first light visible and near-IR integral field spectrograph for the ELT. To achieve its optimal image quality, an accurate measurement of the telescope’s pointing error is necessary. These measurements are affected both by systematic and random error contributions. To characterise the impact of the latter (as well as the performance of any necessary corrective model), simulations of the pointing error measurement process are required. We introduce \textit{harmoni-pm}: a Python-based prototype which, departing from a geometric optics modelisation of the instrument, attempts to reproduce the main drivers of the instrumental pointing error. \textit{harmoni-pm} features a software architecture that is resilient to instrument model refinements and enables performance analyses of corrective models based on simulated calibrations. Results showed that the relay optics are the main drivers of the instrumental pointing error (order 100 \(\mu\)m). Additionally, simulated calibrations of corrective models based on truncated Zernike expansions can compensate for systematic pointing errors up to a residual of order 1 \(\mu\)m.

Keywords: HARMONI, pointing model, corrective model, calibration, pointing simulators, harmoni-pm

1. INTRODUCTION

The Extremely Large Telescope (ELT) is a 39-m class adaptive telescope led by the European Southern Observatory (ESO), currently under construction in Cerro Armazones (Chile). By the time of its completion (circa 2026) it will be the largest VIS-IR telescope ever built. As other contemporary telescopes, ELT will be able to perform diffraction-limited observations by adaptive optics (AO) techniques. However, while current 8-m class telescopes can achieve angular resolutions of up to \(\sim 50\) mas, ELT expands this limit up to \(\sim 10\) mas. These unprecedented figures will vastly advance our knowledge in many branches of Astrophysics, from the study of the earliest galaxies to exoplanets.\textsuperscript{1}

HARMONI (Fig. 1) is a slicer-based integral field spectrograph (IFU) designed to operate in the 0.47 \(\mu\)m – 2.45 \(\mu\)m range (3500 < \(R\) < 18000) in a broad variety of scientific programs.\textsuperscript{2} Its different spectral set-ups will enable observations in the NIR bands (including H, J, K and portions thereof). It will enable both AO (including laser-tomography adaptive optics –LTAO–, single-conjugate adaptive optics –SCAO–, and high-contrast adaptive optics –HCAO–) and non-AO observations, with fields of view (FoV) of 9\" \times 6\”, 4\” \times 3\”, 2\” \times 1.5\” and 0.8\” \times 0.6\” depending on the spaxel scale. In its highest resolution configuration, it will be able to resolve objects at a pixel scale of 4 mas/px in a 0.8\” \times 0.6\” FoV with a plate scale of 3.3mm/\".

Further author information:
G.J.C.C.: E-mail: gcarracedo@cab.inta-csic.es
J.P.L.: E-mail: piqueraslj@cab.inta-csic.es
F.C: fraser.clarke@physics.ox.ac.uk

\textsuperscript{1} Proc. of SPIE Vol. 12189 121892G-1
In order to achieve this resolution level, both HARMONI and ELT must co-operate in closed loop mode, correcting its pointing continuously (order ~ 1 Hz). This is done by measuring the position of natural guide stars (NGS) outside the science field (in the so-called technical field) using a guiding probe in the form of a mobile mirror. The mobile mirror is installed at the end of a 200 mm Pick-Off Arm (POA) that can be placed anywhere in the technical field with a $\theta - \phi$ positioning stage. This positioning stage consists of two precision rotors with rotary encoders inserted in the Low-Order Wavefront Sensing subsystem (LOWFS, see Fig. 2). They provide the POA with the two necessary degrees of freedom that ensure complete coverage of the technical field, namely primary axis ($\theta$) and secondary axis ($\phi$).

LOWFS belongs to a higher-level system called Natural Guide Star Sensing System (NGSS). Its primary role is to track the position of natural stars both for pointing and wavefront sensing.
Prior to its arrival to the NGSS, light entering the instrument (which includes science, laser stars and natural guide stars) has traversed multiple stages (Fig. 3). In a first stage, the prefocal station (PFS) installed in a Nasmyth platform routes the beam of light coming from the telescope to the different focal planes in which instruments are installed. Dichroic filters installed in the Calibration and Relay System (CARS) separate natural star light from laser star light. Finally, the optics of the Focal Plane Relay Subsystem (FPRS) re-focuses the natural star light from the telescope’s focal plane to the NGSS in the so-called relayed focal plane. This plane holds a 1:1 magnification ratio with respect to the telescope’s focal plane, and it is the plane upon which the POA senses guide stars.

1.1 Calibration

In order to exploit the diffraction-limit resolution of \(\sim 10 \text{ mas}\), HARMONI’s detector must Nyquist-sample the field of view. The current pixel scale (4 mas/px, 3.3mm”) ensures this requirement. Nonetheless, this also implies that the instrument must know the WCS coordinates of every spaxel with certain accuracy (R-HRM-153). In practice, this is not only determined by the current pointing of the instrument (determined by the reference pixel of the guide star), but also by the instrumental signature due to opto-mechanical aberrations in the optical path. This signature manifests itself as a certain distribution of systematic pointing errors along the relayed focal plane. A non-comprehensive list of contributions to this error could be the relay optics in the FPRS, wobblings of the derotator’s bearings or mechanical tolerances of the structural components.

This error is calibrated by the calibration module, in which the Geometrical Calibration Unit (GCU) is located. The GCU consists of a mask with a well-known pattern of barely resolved point sources that can be inserted across the telescope beam over the telescope focal plane. As the GCU mask pattern is known, the POA can be used to measure the apparent location of these point sources in the relayed focal plane and calculate the displacement with respect to their expected locations.

The result of the calibration is a corrective model that, as part of the instrument’s pointing model, compensates mechanical and optical displacements of the system up to the positioning accuracy of the POA (\(\sim 10 \mu\text{m}\)). This model is then used to provide an absolute geometrical reference for the guiding probe over the whole technical field.

During operation, pointing error measurements will be fed to certain calibration software to obtain a corrective model. In order to optimise the number of free parameters of the model, the software must take into account not only the error measurements but also the nature of the processes causing them. Additionally, the corrective model may have less degrees of freedom than points in the GCU mask used for calibration. This means that the software may also provide hints to the control logic on how many and which calibration points are preferred.
1.2 Motivation for this work

The motivation of this work is the need for a characterisation of the performance of a corrective model under different sets of calibration points. This will be achieved by a software that will produce simulated measurements of the pointing error in different focal plane locations. The simulator must take into account all effects that may affect the pointing, and therefore an opto-mechanical model of the instrument is required. The simulated measurements will then be used to fit a corrective model whose performance will be measured in terms of model complexity versus model residual.

For the present work, we have focused on the main potential contributions to the pointing error. Both the simulator and the instrument model are designed to ensure their extensibility in the form of additional error contributions.

In Section 2 we detail the opto-mechanical model instrument as well as the form of its corrective model. In Section 3 we introduce the simulator architecture of harmoni-pm. In Section 4 we present the results produced by the simulator. Finally, in Section 5 we enumerate the main conclusions of this work.

2. OPTO-MECHANICAL MODEL

The opto-mechanical model of the instrument must attain the following objectives:

1. Enable the characterisation of the accuracy of the measurement process of pointing errors by the POA.
2. Provide simulated measurements of the pointing error over the whole technical field, given the previous characterisation of the POA accuracy and
3. Evaluate different calibration strategies.

2.1 Working hypotheses

In this first version of the model, we agreed on the following set of assumptions. Many of them are based on technical documents of the instrument describing the analysis of the WCS budgets in relation to top-level requirements of the instrument:

- The behavior of light is well described by geometric optics.
- Focal surfaces are flat (focal planes). Although this is not exactly true, we do not expect the pointing to be affected by the field curvature.
- Beams of light are described only by the location of their focal points. The description of the radiation field inside the telescope is reduced to 2D flux density fields (W/m²). Diffraction effects are modelled as convolutions with the PSF in the flux fields.
- Pointing errors are observed in the focal planes and Z-axis errors (focus) are compensated by the wavefront sensor (WFS). This eliminates the necessity of tracing individual rays and reduces the description of light beams to 2D (R² → R²) transforms between focal points. These transforms map the x₁ = (x₁, y₁) coordinates of the point in the input focal plane to the x₂ = (x₂, y₂) coordinates of the output focal plane.
- Caustic-less regime. This implies that the reciprocal of each transform exists and is unique.

Hence the description of the behavior of light is reduced to providing the set of T : R² → R² coordinate transforms such that x₂ = T[x₁]. This definition of the transforms –along with their invertibility– leads to some convenient properties. In particular, the total error transform can be split into individual transforms applied sequentially, each representing a conceptually different error contribution.
2.2 Parametrisation

The model accepts a set of parameters that may be either fixed values or random variables, being the latter the most numerous. For each random variable, not only the nominal value of the variable but also the specific probability distribution (along with other non-central parameters like its variance) must be provided. Additionally, parameters of the model are classified according to the time at which they are sampled:

- Fixed. The parameter describes a quantity that is known with absolute precision (e.g. bit resolution of an encoder).
- Manufacturing. The variable describes a tolerance that is sampled upon manufacturing.
- Calibration. The variable describes a quantity that remains constant during calibration (e.g. misalignment of the GCU after deployment).
- Positioning. The variable describes a quantity that remains constant after the arm is moved (e.g. true primary axis angle).

2.3 GCU mask

The fundamental reference for geometric calibration is the GCU mask (Fig. 4). The GCU mask consists of a circular plate with a pattern of barely resolved point sources whose geometry is assumed to be known beforehand with certain tolerance. When deployed, the GCU mask is inserted in the telescope’s focal plane –right before the FPRS– and enables the characterisation of the opto-mechanical aberrations of the instrument.

![Geometrical Calibration Unit (GCU)](image)

Figure 4. Detail of the GCU and the technical field mask (also known as GCU mask).

In the current iteration, the GCU mask is simply modelled as a square grid of homogeneously-illuminated circular points. The parameters defining the GCU mask model are described in Tab. 1.

| Symbol | Description                                      | Sampling time |
|--------|--------------------------------------------------|---------------|
| $h_p$  | Point separation between consecutive centres     | Fixed         |
| $D_p$  | Point diameter                                   | Fixed         |
| $F_p$  | Point’s spectral flux density (in dimensions of $MT^{-2}$) | Fixed         |
| $x_0$  | X coordinate of the central point                | Fixed         |
| $y_0$  | Y coordinate of the central point                | Fixed         |
| $D_M$  | Mask diameter                                    | Fixed         |

Table 1. Parameters describing the GCU mask.
This model can be used to compute the locations of the mask dots, which are given by:

\[
G = \{(x_0 + ih_p, y_0 + jh_p)\mid i, j \in \mathbb{Z}, \|x_0 + ih_p, y_0 + jh_p\| < D_M \} \subset \mathbb{R}^2
\] (1)

Future iterations of the model may require a more realistic description of the GCU, with contributions like manufacturing tolerances acting individually on each mask dot.

### 2.4 Transform pipelines

The entry point of light of the instrument is always the telescope’s focal plane, which may come either from the sky or the dot pattern of the GCU mask. Depending on whether we are characterising the behavior of the POA as a measurement device (objective 1) or simulating measurements of the pointing error (objective 2), we will use a different sequence of composed coordinate transforms (henceforth transform pipelines):

- The POA pipeline \( (T_{POA}) \), which translates coordinates in the telescope’s focal plane to coordinates in the POA detector. It will be used to simulate the results of the pointing error measurement process, and

- The pointing error pipeline \( (T_\varepsilon) \), which translates coordinates in the telescope’s focal plane to coordinates in the relayed focal plane as measured by the POA.

These pipelines, at the same time, can be further decomposed in two sets of transforms: a shared set (namely the common path) and a pipeline-specific set (Fig. 5).

![Figure 5. Block diagram of the transform pipelines. The POA pipeline (circled in purple dashed line) considers the light detected in the POA’s focal plane (potentially the WFS detector plane). The pointing error pipeline (circled in green dashed line) considers light arriving at the relayed focal plane plus measurement uncertainties. IRW refers to the Instrument Rotator and Cable Wrap.](image)

Both pipelines differ not only in the destination plane, but also in the interpretation of the coordinate change. While the POA pipeline models the physical behavior of light beams as they travel through the instrument, the pointing error pipeline also models the measurement uncertainty of the true location of bright objects in the relayed focal plane. This extra contribution is necessary as the measurement of the central location of a star (or a GCU mask dot) is performed by a fitting algorithm that can be affected by the resolution of the detector and the shape of the measured object.
2.4.1 Common path

In both cases, the optical path is shared up to the relayed focal plane in the Natural Guide Star Sensors System (NGSS). As previously mentioned, each pipeline can be broken down into a common transform $T_c$ and a pipeline-specific transform as:

$$
T_{\text{POA}} = T_{\text{POA}}^s(\theta, \phi) \circ T_c
$$

$$
T_c = T_c^s \circ T_c
$$

In the common path, the light from the telescope’s focal plane is relayed to the NGSS by the optics of the Focal Plane Relay Subsystem (FPRS) with an ideal magnification ratio of 1:1 (Fig. 6). The parameters defining the total pointing error of this path are summarised in Tab. 2.

| Symbol  | Description                                      | Sampling time  |
|---------|--------------------------------------------------|----------------|
| $\Delta x_G$ | GCU X axis misalignment (only if the GCU is inserted) | Manufacturing |
| $\Delta y_G$ | GCU Y axis misalignment (only if the GCU is inserted) | Manufacturing |
| $\Delta \omega_{IRW}$ | Angular error of the instrument’s derotator | Calibration |
| $\Delta x_N$ | X-axis misalignment of the NGSS structure | Manufacturing |
| $\Delta y_N$ | Y-axis misalignment of the NGSS structure | Manufacturing |

Table 2: Random error contributions in $T_c$

![Figure 6. Modelisation of the common path.](image)

The common transform $T_c$ can be therefore expanded as:

$$
T_c = T(\Delta x_N) \circ R(-\Delta \omega_{IRW}) \circ F \circ T(\Delta x_G)
$$

where $T$ is a translation transform by a given offset and $R$ is a counter-clockwise 2D rotation transform perfectly described by a rotation matrix. $F$ is the FPRS aberration transform which must be constructed by interpolation of the results of existing simulations of the FPRS optics.

2.4.2 POA pipeline

The subpath specific to the POA pipeline relays a small section of the technical field in the surroundings of the POA’s head of the technical field to the POA’s detector plane, and is represented by $T_{\text{POA}}^s(\theta, \phi)$. This subpath is dependent upon the specific POA configuration, which is completely determined by its axis angles $\theta, \phi$. An object in the relayed focal plane whose coordinates equals to those of the center of the POA’s head at a given $\theta, \phi$ configuration should appear (in the error-free case) exactly in the center of the detector plane ($x_d = 0, y_d = 0$). As the detector is expected to move together with the arm, the detector plane will experience a rotation of angle $\phi - \theta$ around its center with respect to the relayed focal plane (Fig 7).
The optics responsible for relaying the light between the POA’s head and the POA detector belong to a component named Low-Order Optical Bench (LOB). For the sake of this study, it is assumed that the magnification ratio of the LOB optics is 1. This assumption could be revisited as the LOB design evolves in future versions of the model. Random contributions to the pointing error are summarised in Tab. 3.

| Symbol | Description | Sampling event |
|--------|-------------|----------------|
| $B_\theta$ | Bit resolution of the primary axis encoder | Fixed |
| $B_\phi$ | Bit resolution of the primary axis encoder | Fixed |
| $q_\theta$ | Quantisation error of the primary axis encoder | Positioning |
| $q_\phi$ | Quantisation error of the secondary axis encoder | Positioning |
| $\Delta \theta$ | Angular error of the primary axis servomotor | Positioning |
| $\Delta \phi$ | Angular error of the secondary axis servomotor | Positioning |
| $e_R$ | Arm length tolerance | Manufacturing |
| $\Delta R$ | Arm length instability (integrated) | Positioning |
| $\Delta \rho$ | Isotropic positioning errors (integrated) | Positioning |

Table 3. Fixed and random variables in $T^s_{POA}(\theta, \phi)$

Figure 7. Modelisation of the POA path

$T^s_{POA}(\theta, \phi)$ can be further decomposed into the following transforms:

$$T^s_{POA}(\theta, \phi) = S(m) \circ R(\tilde{\phi} - \tilde{\theta}) \circ T(-\tilde{x}_{\theta, \phi})$$  

(5)

where $R$ is the 2D counter-clockwise rotation transform and $T$ the translation transform. $S$ is a scale transform that depends on the magnification ratio $m$ from the LOB optics. $\tilde{\theta}, \tilde{\phi}$ are the simulated angles of the POA axes, which are connected to the requested $\theta, \phi$ angles by:

$$\tilde{\theta} = \frac{2\pi}{n_\theta} \left( \frac{n_\theta \theta}{2\pi} + q_\theta \right) + \Delta \theta$$  

(6)

$$\tilde{\phi} = \frac{2\pi}{n_\phi} \left( \frac{n_\phi \phi}{2\pi} + q_\phi \right) + \Delta \phi$$  

(7)

with $n$ the number of discernible encoder intervals of each axis. This number is related to the bit resolution ($B$) of each encoder by $n = 2^B$. On the other hand, the simulated POA’s head center $\tilde{x}_{\theta, \phi} = (\tilde{x}, \tilde{y})$ is obtained from $\tilde{\theta}, \tilde{\phi}$ as:

$$\tilde{x} = \tilde{R} \cos \tilde{\theta} - R \cos (\tilde{\phi} - \tilde{\theta}) + \Delta \rho \cos \alpha$$  

(8)

$$\tilde{y} = \tilde{R} \sin \tilde{\theta} + R \sin (\tilde{\phi} - \tilde{\theta}) + \Delta \rho \sin \alpha$$  

(9)

with $\alpha$ following a uniform distribution between 0 and $2\pi$ and $\tilde{R} = R + e_R + \Delta R$.

2.4.3 Pointing error pipeline

This pipeline is used to generate simulated measurements of the pointing error in the technical field. By taking advantage of the fact that the magnification ratio between the telescope’s focal plane and the relayed focal plane is 1, the pointing error can be defined as:

$$\varepsilon = T_{e}[x] - x$$  

(10)

which can be used to train an appropriate corrective model for the systematic part of the pointing error.

The subpath specific to this pipeline is represented by the transform $T^s_e[x]$ and simulates the uncertainties of measuring the location of an object in the relayed focal plane (Fig. 8). The calculation of $T^s_e[x]$ is a multi-step process that involves:
1. Calculating the nominal arm configuration $\theta, \phi$ that centers the head around $x$.

2. Simulating $\tilde{x}_{\theta,\phi}$ from the requested $\theta, \phi$ as described in Equations 8 and 9, and

3. Simulating the measurement error by the fitting algorithm

![Modelisation of the pointing error subpath.](image)

The set of considered error contributions is the same as for the POA pipeline, plus the measurement error of the fitting algorithm. Nevertheless, as many details of the measurement process are yet to be decided, the exact error distribution of the fitting algorithm is currently unknown and therefore not included in the current iteration of the model. This reduces the definition of $T_s^\varepsilon$ to:

$$T_s^\varepsilon [x] = \tilde{x}_{\theta(x),\phi(x)} \quad (11)$$

And, in the case we are evaluating the performance of certain corrective model $C$ such that $x_c = C[x]$:

$$T_s^\varepsilon [x] = \tilde{x}_{\theta(x_c),\phi(x_c)} \quad (12)$$

### 2.5 Kinematics of the POA

The third objective of the model was to enable comparisons between different calibration strategies. Since one of the figures of merit of a calibration strategy is how fast it completes the measurement of a set of points, we need to take the kinematics of the POA into account.

The estimation of the calibration time involves knowing the exact sequence of POA configuration requests and the behavior of the servomotor of each axis. The current iteration of the model assumes that the kinematics of the POA is purely governed by the two error contributions detailed in Tab. 4.

| Symbol | Description | Sampling time |
|--------|-------------|---------------|
| $\omega_\theta$ | Maximum sweep speed of the primary axis servomotor | Positioning |
| $\omega_\phi$ | Maximum sweep speed of the secondary axis servomotor | Positioning |

Table 4. Random parameters of the POA kinematics

As a first order approximation, the current model assumes that both servomotors can rotate immediately at their maximum sweep speed until they reach their desired angle. This implies that the completion time of a given $\Delta \theta, \Delta \phi$ sweep is given by:

$$t = \max \left( \frac{\Delta \theta}{\omega_\theta}, \frac{\Delta \phi}{\omega_\phi} \right) \quad (13)$$

A noteworthy consequence of this behaviour is that, since both servomotors run independently and in parallel, one axis may reach its desired angle way before the other. When this happens, the sweep transitions from two servomotors running to only one, causing a sudden variation both in sweep speed and direction. The transition point should be reflected as a small cusp in the POA path (Fig. 9).
Figure 9. Traversal of 10 calibration points. The POA path from the first point \( p_1 \), marked in red, to the last \( p_{10} \) exhibits cusps as one axis reaches its destination angle before the other. This effect is particularly noticeable in path sections \( p_1 - p_2 \), \( p_5 - p_6 \) and \( p_7 - p_8 \).

Although we do not expect that the true kinematics deviate much from this description, a complete model should include startup times, the effects of the POA’s inertia, angle correction near the desired position and potential memory effects that may depend on past histories of arm displacements.

2.6 Corrective model

The goal of any calibration strategy is to gather the necessary data to fit a corrective model, whose performance analysis also falls within the scope of this study. This model consists of a certain 2D transform \( C \) that attempts to mimic the behavior of \( T_\varepsilon \) all over the technical field, minimising the residual \( r = \| T_\varepsilon[x] - C[x] \| \). In a similar way as we did for \( T_\varepsilon[x] \) in Equation 10, \( C \) can be expressed in terms of the pointing error model \( \varepsilon'(x) \) as:

\[
C[x] = x + \varepsilon'(x)
\]  
(14)

and therefore the residual can also be written as:

\[
r = \| \varepsilon'(x) - \varepsilon(x) \|
\]  
(15)

In order to exploit the maximum resolution achievable by the telescope (\( \sim 8 \text{ mas}, \sim 28 \mu\text{m} \) in the focal plane), the corrective model must ensure a safe residual goal of \( r < 10 \mu\text{m} \) for all points in the relayed focal plane.

In real life operation, \( C \) is found during calibration by measuring the position of the GCU mask dots with the POA detector. As the time available for calibrations is limited, a good corrective model \( C \) should not only minimise the residual, but also the time required to gather the necessary data, which is directly connected to the number of points used for calibration. Consequently, we will favour models that require fewer points.

With these considerations in mind, we decided to express the pointing model \( C \) in terms of the complex Zernike expansion of the pointing error.\(^5\) The rationale behind this approach is twofold:

- Zernike expansions are familiar to optical engineers when it comes to describing optical aberrations, and
- the first radial orders describe aberrations that can be connected to common opto-mechanical effects (misalignments, rotations...).
There are some additional benefits derived from using Zernike polynomials: as they conform a complete orthogonal basis on the unit circle,\(^6\) coefficients of a Zernike expansion can describe properties of the aberration without redundancy or overlap of information between them (which provides insight on the nature of the aberrations). On the other hand, the algebra of a 2D Zernike expansion is greatly simplified when complex coefficients are used. In particular, if we identify 2D vectors with complex scalars as:

\[
x = \Re[z] \hat{e}_x + \Im[z] \hat{e}_y, \quad z \in \mathbb{C}
\]

Then, the pointing error model can be expanded as:

\[
\hat{\epsilon}'(z) = \sum_{n=0}^{N} \sum_{i=0}^{n} a_n^{-n+2i} Z^{-n+2i} \left( \frac{z}{R} \right)
\]

where \(a_n^m\) is the complex coefficient multiplying the complex Zernike polynomial \(Z_n^m\) and \(N\) is the maximum radial order of the expansion. As Zernike polynomials are defined on the unit circle, \(z\) must be normalised prior to the evaluation by the technical field radius \(R\). Note that for a maximum radial order \(N\) we need to fit \(J = \frac{(N + 1)(N + 2)}{2}\) complex coefficients.

When Zernike polynomials are used to build certain vectors and matrices, it is customary to refer to the first \(J\) polynomials. In these cases, it is convenient to map the two indices \(m, n\) of the Zernike polynomials to a single integer index \(j \geq 0\). For the sake of standarisation, we chose OSA/ANSI indices,\(^7\) which relate \(j\) and \(m, n\) by:

\[
j = \frac{n(n + 2) + m}{2}
\]

### 2.6.1 Model fit

Fitting the aforementioned corrective model requires deciding both the maximum radial order of the model \((N)\) and the calibration pattern \(P = \{p_0, p_1, \ldots\} \subseteq G \subset \mathbb{C}\), with \(G\) the set of locations of the GCU mask dots in complex form. Ideally, we would need as many points as free parameters: \(Q = J = \frac{(N + 1)(N + 2)}{2}\), with \(Q = \text{Card}(P)\). However, since measurements are affected by noise, it could be convenient to introduce certain redundancy by choosing more than \(J\) calibration points.

Once the error measurement vector \(\tilde{\epsilon} = (\tilde{\epsilon}(p_0), \tilde{\epsilon}(p_1), \ldots, \tilde{\epsilon}(p_{Q-1}))^T \in \mathbb{C}^Q\) has been obtained, one can pose the model fitting problem as finding \(a = (a_0, a_1, \ldots, a_{J-1})^T \in \mathbb{C}^J\) such that:

\[
\begin{pmatrix}
Z_0(p_0^2) & Z_1(p_0^2) & \cdots & Z_{J-1}(p_0^2) \\
Z_0(p_1^2) & Z_1(p_1^2) & \cdots & Z_{J-1}(p_1^2) \\
\vdots & \vdots & \ddots & \vdots \\
Z_0(p_{Q-1}^2) & Z_1(p_{Q-1}^2) & \cdots & Z_{J-1}(p_{Q-1}^2)
\end{pmatrix}
\begin{pmatrix}
a_0 \\
a_1 \\
\vdots \\
a_{J-1}
\end{pmatrix}
= 
\begin{pmatrix}
\tilde{\epsilon}(p_0) \\
\tilde{\epsilon}(p_1) \\
\vdots \\
\tilde{\epsilon}(p_{Q-1})
\end{pmatrix}
\]

The coefficient matrix of this system of equations is referred to as the collocation matrix,\(^8\) and it is denoted by \(Z(P)\). When \(Q > J\) the system is generally inconsistent due to the measurement noise. Nevertheless, a solution can still be found if we reformulate it as an optimisation problem:

\[
a = \underset{\alpha}{\text{argmin}} \| Z(P)a - \tilde{\epsilon} \|^2
\]

The quality of the calibration pattern \(P\) is given not only by the number of points in it, but also by how evenly it samples the technical field. This is directly reflected in the reduction of the condition number of \(Z(P)\), indicating higher robustness against measurement noise.
2.6.2 Calibration strategies

A calibration strategy is the choice of a calibration pattern $P$ along with the order in which it should be traversed that ensures a calibration residual below 10 µm. A good calibration strategy is a trade-off between the quality of the calibration pattern and the time required to measure it. Since finding this trade-off is non-trivial, different calibration patterns and orderings should be tested. In the current iteration of the model, we propose three calibration patterns:

- Random (i.e. calibration points are chosen randomly from $G$).
- Spiral (closest GCU points to the spiral pattern described in $9$).
- Optimal Concentric Sampling (OCS, closest GCU points to the pattern described in $8$).

Finding the ordering that minimises the calibration time is a particular case of the Traveling Salesman Problem and can only be addressed via exhaustive search of all the $Q!$ possible paths. Instead, we propose two heuristics: as-is (traversing the points in a fixed arbitrary order), and closest neighbour (the next point to be traversed is the closest to the last one in infinity-norm distance).

3. DESIGN AND IMPLEMENTATION

The resulting application must be understood as a prototype that will eventually be used in the observatory during operation. As such, it should not only address the functional requirements selected for this first iteration, but also foresee future extensions of the model by other contributors until then. The goal of the project is to provide the observatory with a useful tool to optimise the pointing calibration strategies of the instrument. Its architecture should be easily upgraded to account for a more realistic description of the “as-built” instrument.

3.1 Software requirements

The simulator will consist of a series of programs which, upon successful execution, will produce different simulation products in the form of data files and graphs. From the perspective of functional requirements, the prototype must produce the following results:

1. Heatmaps of the pointing error in the relayed technical field.
2. Priors of the coefficients of the corrective model via Monte Carlo sampling.
3. Curve of residuals for different calibration strategies.
4. Heatmaps of the pointing error residual after applying the corrective model.
5. Time plots of POA’s operation for different calibration strategies.

In addition to the goals described in Section 2, the design must be aware of the incompleteness of the model. In particular, we must assume that other programmers with varying sets of skills will eventually be involved in the project. Also, as other use cases may be identified in the future (e.g. integration with the observatory’s control software) proper component decoupling will be necessary. All this translates to the following set of non-functional requirements (NFR):

1. The programming language of the project must be the result of a trade-off between performance and popularity.
2. The architecture of the application must be decided beforehand.
3. The code of the entry point of the simulator (i.e. the executable programs) must be decoupled from the simulator’s logic.
4. Implementation must prioritise maintainability before performance.
5. Simulator must work out of the box, even if no parameters are provided.
6. Exhaustive code and application documentation must be provided.

NFR 1 is satisfied by choosing Python 3 + NumPy. This is a popular combination among physicists and engineers that enables high-performance computing with tensors of arbitrary rank. Additionally, the functional requirements of the simulator can be fulfilled by a component-based architecture (NFR 2) which must be properly documented (NFR 6) prior to the coding phase.

3.2 Simulator architecture

NFRs 2 and 3 motivated a component-based architecture. The goal of this paradigm is to decompose the design of the application into individual functional or logical components that represent well-defined interfaces containing classes. It is the object-oriented equivalent of the well-known programming principle of divide-and-conquer.

In order to ease future code reuse, terminal components (i.e. those representing executable scripts) are separated from non-terminal components, with the latter being placed inside a Python package named harmoni.pm (Fig. 10).

![Component diagram of the project.](image)

Figure 10. Component diagram of the project. Non-terminal components (containing class implementations and support code) are colored in blue, while terminal components are colored in yellow. The high-level of the instrument model as a composition of several transforms is included in the Optomechanics component.

Terminal components PointingSim and SGSim refer to the scripts that may be executed by the user. PointingSim will be in charge of producing the results enumerated in 3.1, while SGSim will be the prototype of detector simulator that will be used in the future to characterise the behavior of the POA.

3.3 Source code

The source code of the simulator can be accessed from its public GitHub repository at https://github.com/BatchDrake/harmoni-pm, while the documentation of the internal API can be accessed at https://actinid.org/harmoni-pm. The usage of Python 3 as the project’s programming language should ensure portability to other platforms. However, at the current stage of development, it is recommended to run the code in Unix-like operating systems.
Project files are structured so that every component of the simulator is kept in a separate directory under the `harmoni.pm` package, while executable scripts are kept at root level of the project directory.

In order to run the different executable scripts, the user must ensure the following `pip` dependencies are met: `chardet`, `matplotlib`, `numpy`, `pandas`, `Pillow`, `Pint`, `seaborn`, `scipy`, `skyfield`, `uncertainties`

4. RESULTS AND APPLICATIONS

Although the opto-mechanical model is expected to be refined prior to its final integration in the observatory, some of its simulation products have direct application in terms of calibration strategy. In particular, we provide a thorough comparison of the performance of different calibration strategies, providing strong evidence of the superiority of the OCS pattern with closest neighbour traversal. These results are expected to be included in the technical documentation of the instrument with the aim of guiding the design of the calibration control software.

Model-limited results are also included (like pointing error heatmaps before and after applying the corrective model). Although these results would be revisited in future phases of the instrument development, they already provide useful insight on the effects of using corrective models of different radial orders.

4.1 Analysis of the corrective model

4.1.1 Uncorrected error heatmap

We evaluated $\epsilon$ in the whole relayed focal plane as seen by the POA using `PointingSim`'s subcommand `errormap`, obtaining a maximum error of around 200 $\mu$m (with typical values around 100 $\mu$m). This value surpasses the goal of 10 $\mu$m by more than one order of magnitude, and it has been tracked to the optical aberrations inside the FPRS (Fig. 11).

![Error heatmap (without model)](image)

Figure 11. Uncorrected pointing error error heatmap.

4.1.2 Radial order of the corrective model

Following, we wanted to know whether complex Zernike polynomials are a good choice for both optical and mechanical aberrations, and what is the minimal radial order ($N$) required to describe them with an error below 10 $\mu$m. For this purpose, we simulated a full calibration (i.e. comprising all possible test points) fitting the coefficients of several complex Zernike expansions with $N$ ranging from 1 to 4. The total number of coefficients ($J$) fitted in each case is related to the radial order by $J = (N + 1)(N + 2)/2$.

The heatmaps in Fig. 12 represent the residual as defined by Equation 15. Note that for the case $N = 1$, only displacements, rotations and changes in aspect ratio can be modelled.
We see that although the goal accuracy is achieved when $N = 2$ ($J = 6$), we still observe structure in the heatmap residual. This structure disappears at higher orders, and the residual is found to be lower-bounded by a noise floor at $r = 1.5$ μm. In the current iteration of the simulator, this floor is dominated by the angular error of the POA encoders ($\sigma_\Delta \theta = \sigma_\Delta \phi = 1$ as).

We therefore conclude that there are two candidate configurations of the corrective model, one with $N = 2$ and other with $N = 3$. Radial orders below 2 are too simple to meet the $r < 10$ μm goal, while radial orders above 3 do not provide significant improvements of the residual. This tie can be broken by making a trade-off between model accuracy and calibration time.

### 4.2 Calibration pattern

One of the results with immediate application comes from the performance analysis of all calibration patterns with different corrective model orders and numbers of points. This comparison was performed by fitting corrective models for each pattern (random pattern, spiral sampling and OCS) with radial orders 2 ($J = 6$), 3 ($J = 10$) and 4 ($J = 15$), different numbers of calibration points between 1 and 30 and 50 different realisations of each pattern. The quantity used for comparison was the median of the 50 mean residuals calculated for each pattern realisation. Error bars in Fig. 15. represent the maximum and minimum observed mean error observed in the 50 realisations of the pattern.
We observed that OCS exhibits a faster reduction of the mean error, closely followed by the spiral pattern (although with higher dispersion). However, there is still the possibility of tie for OCS and spiral patterns when the number of calibration points matches the number of complex coefficients of the corrective model.

One remarkable feature of OCS-based calibrations is the presence of valleys in the median error when the number of calibration points is triangular (3, 6, 10, 15...). This is expected: OCS patterns always have the number of points of a complete set of Zernike polynomials up to certain radial order (which is triangular), and the only way to obtain an intermediate number is by removing points of an existing complete pattern, affecting its symmetry and therefore its performance.

It is worth noting that the number of points required to perform a calibration that meets the $r < 10$ µm goal (order 10) is much smaller that the number of available dots in the GCU mask (order $10^5$). Once a full
opto-mechanical model is developed, this result may be used to optimise the design of the GCU mask in order to provide only those dots actually used for calibration.

### 4.2.1 Spatial distribution of the calibration residuals

To break the apparent tie between the OCS and spiral patterns, we carried a more detailed study of the residual. To that end, we obtained residual heatmaps for both the spiral and OCS pattern when $Q = J$ (i.e. when the number of calibration points matches the number of coefficients of the corrective model), with $J = 6$ and $J = 10$. In the $J = 6$ case, we observed that the lack of symmetry of the spiral pattern always causes one side of the relayed field to be undersampled, while OCS calibrations tend to sample the plane more homogeneously (Fig. 16).

![Figure 16. Calibration with spiral pattern (left) and OCS (right) up to radial order $N = 2$ ($J = 6$). The heatmap represents the Euclidean norm of the 2D residual after calibration.](image)

This effect is exaggerated at higher orders, as the scarcity of sampling points in the edges of the field causes the corrective model to extrapolate in uninformed ways. In these regions, the pointing error reaches values comparable to those found in the uncorrected case (Fig. 17).

![Figure 17. Calibration with spiral pattern (left) and OCS (right) up to radial order $N = 3$ ($J = 10$). The heatmap represents the Euclidean norm of the 2D residual after calibration. The spiral pattern reduces the error near the calibration points but increases it significantly near the undersampled region. On the other hand, the OCS pattern reduces the error almost everywhere, to the extent of making the non-systematic contributions of the error noticeable.](image)

From these results, we conclude that OCS performs better than the spiral pattern when the number of calibration points equals to the number of coefficients of the corrective model. Additionally, when a model with radial order $N = 3$ is used, calibrations are far from the $r < 10 \mu m$ goal due to the presence of highly undersampled regions. This renders OCS the only viable alternative. Nevertheless, the optimality of this result could still be challenged by error contributions not yet considered in the instrument model.

### 4.3 Sampling strategy

Next, we addressed the problem of finding the best calibration strategy, i.e. finding the path that traverses all calibration points in the shortest possible amount time. Since this problem is a particular instance of the
Travelling Salesman Problem (TSP)\textsuperscript{10} with a distance given by the time the POA takes to move from the departure configuration $(\theta_1, \phi_1)$ to the destination configuration $(\theta_2, \phi_2)$, it is NP-complete and the optimal solution implies testing all $Q!$ point permutations which, for the $Q = J = 10$ case, amount to more than $3.5 \times 10^6$.

Instead, we tested a heuristic based on finding the closest uncalibrated point (in $L_\infty$ distance of its $\theta, \phi$ coordinates) iteratively. In the $Q = 10$ case, we observed a speedup from 600 s to 350 s (which supposes a reduction of 40\% in calibration time, see Fig. 18). The simplicity of the heuristic, along with its dramatic reduction of the calibration time justifies its usage when $Q \gg 10$.

Figure 18. Time plots for a typical non-optimised OCS calibration with 10 calibration points (left) and after applying the closest uncalibrated point heuristic (right).

4.4 Priors of the corrective model

\textit{PointingSim} is able to compute histograms of the coefficients of the corrective model by simulating multiple calibrations using the maximum number of calibration points. The resulting histograms of the coefficients of the corrective model can be used to model the priors of the bayesian formulation of the calibration problem:

$$p(a_j, \ldots | \tilde{\varepsilon}_i, \ldots) \propto p(\tilde{\varepsilon}_i, \ldots | a_j, \ldots) p(a_j, \ldots)$$

where $a_j, 0 \leq j < J$ are the model coefficients, $\tilde{\varepsilon}_i, 0 \leq i < Q$ the measured errors and $p(\tilde{\varepsilon}_i, \ldots | a_j, \ldots)$ the likelihood function, whose exact formulation requires a good understanding of the measurement process.

For illustrative purposes, we simulated a total of $2 \times 10^4$ calibrations over individual realisations of manufacture-time values, using all calibration points in the simulated GCU mask. The results of the simulation shows high statistical independence between coefficients, which suggests that the bayesian problem may be simplified down to $J$ smaller problems:

$$p(a_j | \tilde{\varepsilon}_i, \ldots) \propto p(\tilde{\varepsilon}_i, \ldots | a_j) p(a_j), 0 \leq j < J$$

5. CONCLUSIONS

We addressed the problem of modelling the contributions to the pointing error as observed by HARMONI. Due to the inherent complexity of the instrument, we selected a subset of contributions that we considered to be the main drivers of this error. This resulted in a preliminary corrective model whose formalisation foresaw future refinements. These refinements may take the form of better characterisation of the current contributions or the inclusion of contributions not yet considered in the model.

The model was put into practice by a set of Python packages and scripts that simulated different aspects of the instrument, including uncalibrated pointing errors, the performance of different calibration strategies and corrective model configurations and corrective model residuals.

Although many properties of the instrument components were not considered in this stage, the current software architecture facilitates the extension of the model as more details of the design are constrained. This did not prevent the simulator from producing results that can be applied in the future stages of development of the instrument:
• Assuming the FPRS model is realistic, we conclude that a corrective model with \( N = 2 \) (6 complex coefficients) fulfills the positioning accuracy of 10 \( \mu \)m. If a corrective model with \( N = 3 \) (10 complex coefficients) is fitted, the residual plummets below the amplitude of the mechanical instabilities. The choice between both alternatives should be the result of a trade-off between positioning accuracy and calibration time.

• We provided evidence for the superior performance of the OCS pattern over the traditional spiral pattern when the number of calibration points matches the number of complex coefficients of the corrective model, especially at higher radial orders.

• We also provided priors for the coefficients of the corrective model which, in a worst case scenario, will provide information on the orders of magnitude of the aberration as projected in each Zernike polynomial.

• Finally, we concluded that we can speed up the calibration process not only by reducing the number of calibration points to the number of coefficients of the corrective model, but also by reordering points in a way that does not involve testing all possible permutations of the calibration set.

REFERENCES

[1] ESO, “The e-elt construction proposal.” https://www.eso.org/sci/facilities/eelt/docs/e-elt_executivesummary.pdf (2010). Accessed: 2021-08-26.

[2] Thatte, N. A., Clarke, F., Bryson, I., Shmetler, H., Tecza, M., Fusco, T., Bacon, R. M., Richard, J., Mediavilla, E., Neichel, B., Arribas, S., Garcia-Lorenzo, B., Evans, C. J., Remillieux, A., Madi, K. E., Herreros, J. M., Melotte, D., O’Brien, K., Tosh, I. A., Vernet, J., Hammersley, P., Ives, D. J., Finger, G., Houghton, R., Rigopoulou, D., Lynn, J. D., Allen, J. R., Zieniwniski, S. D., Kendrew, S., Ferraro-Wood, V., Pécontal-Rousset, A., Kosmalski, J., Laurent, F., Loupias, M., Piqueras, L., Renault, E., Blaizot, J., Daguisé, E., Migniau, J.-E., Jarno, A., Born, A., Gallie, A. M., Montgomery, D. M., Henry, D., Schwartz, N., Taylor, W., Zins, G., Rodríguez-Ramos, L. F., Cagigas, M., Battaglia, G., López, R. R., Suárez, E. H., Gigante-Ripoll, J. V., López, J. P., Martin, M. V., Correia, C., Pascal, S., Blanco, L., Vola, P., Epinat, B., Peroux, C., Vigan, A., Dohlen, K., Sauvage, J.-F., Lee, M., Carlotti, A., Verinaud, C., Morris, T., Myers, R., Reeves, A., Swinbank, M., Calcines, A., and Larrieu, M., “The E-ELT first light spectrograph HARMONI: capabilities and modes,” in [Ground-based and Airborne Instrumentation for Astronomy VI], Evans, C. J., Simard, L., and Takami, H., eds., 9908, 595 – 605, International Society for Optics and Photonics, SPIE (2016).

[3] Estrada, A., Piqueras, J., Argelaguet, H., Morris, T., Dubbeldam, M., Bardou, L., Dunn, A., Deshmukh, R., Dohlen, K., Mignot, D. L., Clarke, F., Arribas, S., and Pereira-Santaella, M., “HARMONI at ELT: Pick-Off Arm module design status and prototype results,” International Society for Optics and Photonics (2022).

[4] Clarke, F., “Harmoni wcs knowledge,” Internal document.

[5] Biesheuvel, R. S., Janssen, A. J. E. M., Pozzi, P., and Pereira, S. F., “Implementation and benchmarking of a crosstalk-free method for wavefront Zernike coefficients reconstruction using Shack-Hartmann sensor data,” OSA Continuum 1, 581 (Oct. 2018).

[6] Navarro, R., Rivera, R., and Aporta, J., “Representation of wavefronts in free-form transmission pupils with Complex Zernike Polynomials,” Journal of Optometry 4, 41–48 (Apr. 2011).

[7] Thibos, L. N., Applegate, R. A., Schwiegerling, J. T., Webb, R., and VSIA Standards Taskforce Members, “Standards for Reporting the Optical Aberrations of Eyes,” in [Vision Science and its Applications], S652–60, OSA, Santa Fe, New Mexico (2000).

[8] Ramos-López, D., Sánchez-Granero, M., Fernández-Martínez, M., and Martínez-Finkelshtein, A., “Optimal sampling patterns for Zernike polynomials,” Applied Mathematics and Computation 274, 247–257 (Feb. 2016).

[9] Navarro, R., Arines, J., and Rivera, R., “Direct and inverse discrete Zernike transform,” Optics Express 17, 24269 (Dec. 2009).

[10] Lawler, E. L., ed., [The Traveling salesman problem: a guided tour of combinatorial optimization], Wiley-Interscience series in discrete mathematics, Wiley, Chichester [West Sussex] ; New York (1985).