Considerations for EAGLE from Monte Carlo adaptive optics simulation

Alastair Basden,* Richard Myers, and Timothy Butterley

Department of Physics, Durham University, South Road, Durham DH1 3LE, UK
*Corresponding author: a.g.basden@durham.ac.uk

Received 15 January 2010; revised 12 April 2010; accepted 3 May 2010; posted 11 May 2010 (Doc. ID 122837); published 27 May 2010

The EAGLE instrument for the European Extremely Large Telescope is a multi-object integral field unit spectrograph that uses a multi-object adaptive optics (AO) system for wavefront correction of interesting lines of sight. We present a Monte Carlo AO simulation package that has been used to model the performance of EAGLE, and provide results, including comparisons with an analytical code. These results include an investigation of the performance of compressed reconstructor representations that have the potential to significantly reduce the complexity of a real-time control system when implemented. © 2010 Optical Society of America

OCIS codes: 010.1080, 110.1080.

1. Introduction

The next generation of optical ground-based Extremely Large Telescopes (ELTs) is currently in the design phase, with plans for primary mirror diameters of over 30 m [1,2]. Once built, these facilities will allow astronomers to probe the universe with unprecedented sensitivity and very high resolution. A suite of instruments for these telescopes is planned, allowing many different observation goals to be met.

The EAGLE instrument for the planned 42 m European ELT (E-ELT) is currently in the design phase [3]. It is a multi-object integral field unit spectrograph using adaptive optics (AO) with a multi-object AO (MOAO) system to correct incoming wavefronts in open loop, using wavefront sensors that do not sense the corrections made to the science fields. Baseline designs for EAGLE include up to 11 wavefront sensors, using laser and natural guide stars. It is envisaged that there will be 20 science field pick-offs, allowing good AO correction in 20 separate fields, each 1.5 arc sec in diameter, simultaneously across a 5 arcmin field. The use of a MOAO system allows good atmospheric correction to be achieved for selected objects across a wide field of view.

Part of the design phase for EAGLE includes extensive simulation and modeling of the AO performance, since AO is an essential part of the instrument design. The simulation and modeling are carried out in two phases. First, an analytical code is used to obtain an order-of-magnitude performance estimate, covering a large parameter space relatively quickly. However, many fine details that are essential to include for ELT scale designs are not included. A Monte Carlo code is then used to fill in details to give a more reliable performance estimate, including nonlinear effects and noise sources. However, Monte Carlo simulation has far greater computational requirements, so a reduced parameter space is considered.

The Durham AO simulation platform (DASP) is a Monte Carlo code that can be used for the simulation of any common form of AO system (including classical AO, laser tomographic AO, multiconjugate AO, and MOAO [4]), and has been developed specifically with ELT simulation in mind [5]. It is an end-to-end time-domain code and is parallelized, allowing it to be used across a computing cluster using the Message Passing Interface (MPI) library to reduce computation time. It includes detailed models of telescope and AO systems, allowing high-fidelity models to be produced.
The development of a real-time control system for EAGLE is a challenge. There are expected to be of the order of $10^5$ wavefront slope measurements, and of the order of $10^4$ deformable mirror (DM) actuators to control per science path (of which there will be about 20). These wavefront slope measurements will be used to update the DM actuators at about 250 Hz. It is likely that EAGLE will use a conventional matrix-vector-based wavefront reconstruction, although other techniques, such as iterative algorithms, have not been ruled out. In this paper, we concentrate only on the matrix-vector-based wavefront reconstruction and consider some details that may make this easier to implement in hardware. To access all the elements of the control matrix for each new set of slope measurements, a data rate of the order of 20 Tbytes s$^{-1}$ is therefore required (assuming 4 bytes per matrix element). This will require a very advanced control system, and so any achievable simplifications are desirable.

Here, we present some recent results obtained from the simulation of EAGLE using the DASP. Some of these results are compared with those from an analytical code, produced independently by another member of the EAGLE consortium, where appropriate. However, the analytical code is unable to include non-linear effects, and so can only be used for rough performance estimates. The simulations presented here include an investigation of compressed wavefront reconstructor algorithms that could simplify real-time control system design. The technical difficulty of the real-time control system design is such that it should be considered even at the early design phases. We also discuss the issue of Shack–Hartmann sensor non-linearities for open-loop systems.

In Section 2, we describe the simulations that have been carried out, in Section 3 we give results, and conclusions are made in Section 4.

2. Simulation Description

There are several possible designs for EAGLE, with different laser guide star (LGS) and natural guide star (NGS) requirements, based on trade-offs among cost, performance, and sky coverage. Here, we concentrate on a design with nine LGSs equally spaced around a ring with a 7.3 arc min diameter and a single NGS with 16 $\times$ 16 subapertures used for low-order-mode correction (tip, tilt, focus, and astigmatism), which LGS sensors easily measure usefully. We assume that the LGSs are center launched and have an elongation of 5 arcsec at the edge of the telescope pupil (maximum elongation). No measures to mitigate this elongation are made and wavefront slope computation uses a center-of-gravity algorithm. The telescope diameter is assumed to be 42 m, and the wavefront sensors (WFSs) have 84 $\times$ 84 subapertures, each with 20 $\times$ 20 pixels, unless otherwise stated, requiring a WFS detector with 1680 $\times$ 1680 pixels. Each science field pick-off has its own deformable mirror (85 $\times$ 85 actuators) with wavefront control optimized along the line of sight for this field. Unless otherwise stated, the results presented here are for a target at the center of the field, i.e., in the middle of the LGS ring. The science wavelength is H-band (1.65 $\mu$m). We use a virtual DM formulation for wavefront control, placing virtual DMs conjugate to the height of strong turbulence. This allows us to reconstruct the atmospheric turbulence at the positions of these virtual DMs, and to use this knowledge to determine the shape that should be given to the physical DM for this science field. In the simulations here, the shape given to each MOAO physical DM is the sum of the virtual DMs projected along the line of sight for this science field. The virtual wavefront reconstruction is performed using a standard truncated least-squares matrix-vector algorithm with the vector containing the latest wavefront slope measurements, and the matrix being the pseudoinverse of the system interaction matrix (the measured WFS response to perturbations induced on the virtual DMs). The simulations presented here all assume two discrete layers of atmospheric turbulence, and two virtual DMs unless otherwise stated. It should be noted that the estimated performance reported here should be seen as optimistic due to the simple nature of a two-layer profile. The atmospheric outer scale is taken as 50 m, and Fried’s parameter is 10.6 cm (at 500 nm), corresponding to a seeing of 0.95 arcsec. These values are as used in simulations carried out by Fusco et al. for the EAGLE consortium [6], which our Monte Carlo simulations are used to verify. The update rate of the AO loop is 250 Hz and, unless otherwise stated, a delay (latency) of 4 ms between wavefront detection and correction is simulated.

The deformable mirrors are operated in open loop, i.e., the WFSs do not sense changes made to the DMs. In these simulations, atmospheric phase screens are translated across the telescope pupil assuming a frozen-flow turbulence model [7]. The sections of these screens relevant to a given line of sight at a given time are then selected (with subpixel interpolation) and summed (with interpolation for a source at finite distance, e.g., a LGS). These line-of-sight pupil phases are then used as input to Shack–Hartmann WFS models, which produce a simulated noisy Shack–Hartmann image, and to generate science camera images, before and after correction of the phase using a DM. The DMs are controlled by a wavefront reconstructor, which uses the slope measurements taken by the WFSs to compute the correction to be applied. We use a center-of-gravity algorithm for wavefront sensing. Sodium laser spots (as produced by a LGS) are assumed to form at 90 km, with a depth of about 10 km and a Gaussian distribution. The performance of the AO system can be measured as a function of time, and the average long-exposure performance is also obtained. These simulations include many noise sources, including detector noise, photon shot noise, LGS elongation, and WFS non-linearities. The simulation code is therefore suited to the high-fidelity modeling of AO systems.
A. Parameter Space

We have covered a large parameter space during these simulations and, with our available hardware, we are able to cover about seven parameter points per day, including the generation of interaction and control matrices. Parameters that have been explored include:

1. DM misconjugation,
2. control matrix representation (investigating reductions in control matrix size to simplify real-time control system development),
3. WFS linearization,
4. DM misalignment,
5. wavefront sensor pixel scale,
6. zenith angle,
7. LGS power,
8. WFS read-out noise,
9. secondary mirror support obscuration,
10. woofer–tweeter configuration, and
11. sodium layer profile.

Here, we consider further the first three parameters. DM misconjugation can occur when knowledge of the atmosphere is not perfect, and it introduces an additional error into the correction applied to the DM. Control matrix representation is an important consideration for the design and development of real-time control systems, allowing designs to be simplified and costs reduced when the memory required to store a control matrix is reduced. WFS linearization is necessary for open-loop systems because Shack–Hartmann-based WFSs have a slightly non-linear response to incident wavefront slope, and attempts to calibrate and correct this non-linearity can improve AO performance.

B. Deformable Mirror Misconjugation

A multiconjugate AO system (including a MOAO system such as EAGLE, using virtual DMs) requires information about the strength and position of turbulent atmospheric layers to operate most effectively. Deformable mirrors are then conjugated at the locations of the most dominant layers. However, if misconjugation occurs, for example, because layer positions are not well known, the AO system performance will be degraded.

Figure 2 shows how the performance of an AO system (correcting at H-band) is degraded by misconjugation. Here, dominant turbulent layers were placed at 0 and 10 km, and two virtual DMs were placed at 0 km, and at varying heights between 8 and 12 km. This figure demonstrates that it is necessary to be...
able to conjugate DMs to within a few hundred meters of dominant turbulence. Analytical results provided by Fusco et al. [6], which replace the LGSs with NGSs (the analytical code cannot model cone effect or spot elongation) for a system with $110 \times 110$ subapertures per wavefront sensor (long dashes) are shown to be slightly optimistic for perfect conjugation when compared with equivalent Monte Carlo results (solid curve), with $110 \times 110$ subapertures (for comparison purposes) for each wavefront sensor. When a cone effect due to the laser spot being at a finite distance (meaning only a cone of turbulence is sampled by the WFS) and spot elongation caused by the three-dimensional nature of sodium emission are included in the simulations (dotted curve), performance is seen to fall by about 10%. These effects are not modeled in the analytical results. A reduction of WFS order to $84 \times 84$ subapertures is shown to further reduce performance (the lower two curves). Performance is also shown to be dependent on the sodium profile.

It should be noted that these simulations are a simplification of the true situation, where there will be many more turbulent layers, each with finite thickness. However, for all these cases, the general trend with misconjugation is clear, implying that a DM should be conjugated to dominant turbulence with an accuracy of a few hundred meters. This places constraints on the design of turbulence profiling systems.

Figure 3 shows how the simulated AO system performance falls as a function of number of atmospheric layers in these simulations. Here, we have not sought to optimize the wavefront reconstruction in any way, using a simple truncated least-squares wavefront reconstructor. A virtual DM has been placed conjugate to each layer, with an actuator spacing calculated to minimize fitting error, following [8]. The ideal number of virtual DMs, their conjugate heights, and the actuator spacings to use to optimize MOAO system performance is a subject of ongoing research. Here, we do not consider the effect of DM misconjugation when there are more than two atmospheric layers. The table in Fig. 4 shows the parameters used for these multiple layer simulations, as provided by Fusco et al. [6]. A global Fried parameter of 10.6 cm and an outer scale of 50 m were used. It should be noted that wavefront reconstruction uses a least-squares algorithm. The use of a minimum variance wavefront reconstruction may improve performance. However, this shows that the performance of EAGLE is likely to fall when the atmospheric turbulence is heavily layered.

C. Reconstructor Representation
The control matrix for a single EAGLE science path is likely to contain of the order of $10^9$ elements and must be accessed at a rate of 250 Hz, requiring a memory bandwidth of $1 \text{Tbytes s}^{-1}$, assuming 32 bit floating point format storage. When considering that EAGLE is likely to have up to 20 science paths, the memory bandwidth requirement increases by a factor equal to the number of science paths, up to $20 \text{Tbytes s}^{-1}$ for EAGLE.

Reducing this memory bandwidth requirement is important to reduce the real-time control system complexity. Assuming a field programmable gate array (FPGA-based wavefront reconstruction unit, the memory bandwidth will be determined by the FPGA clock rate, the memory-to-FPGA bus width, and the number of FPGAs used for processing. By reducing the total size of the control matrix, the number of FPGA scans be reduced, leading to a cheaper, simpler, more reliable design. We now consider several techniques that can be applied to reduce the control matrix size.

1. Sparse Representation
Sparse matrix representation of AO system control matrices has been studied [9], and for multiconjugate systems (or most systems without a specific WFS-to-DM alignment), sparse matrix techniques are known to perform poorly [10] due to poorly sensed modes and LGS tip–tilt uncertainty. Figure 5 verifies this, showing that a highly non-sparse representation is required to maintain the AO system performance. The sparse matrices used here are created by removing the least influential parts of the control matrix, i.e., elements closest to zero. Typically, 70% of the original matrix must be present, as demonstrated in Fig. 5. However, when stored in sparse format, each matrix element must be stored accompanied by its position, resulting in twice as much storage (assuming 32 bit floating point for the matrix element, and a 32 bit integer for position), thus consuming more memory than the original control matrix. Therefore, sparse matrix representation is not a solution for EAGLE.

2. Fixed Point Representation
Fixed point representation is often used in hardware (for example, FPGAs) as it is simpler to use than
floating point representations, and has a lower computational complexity. Here, values are stored in two’s-complement integer format, with a known scaling factor. By using a fixed point control matrix representation, as shown in Fig. 6, it is possible to reduce the control matrix storage requirements by a factor of 2, using 16 bit fixed point values rather than 32 bit floating point values, while still maintaining the AO system performance. To compute the fixed point control matrix, the minimum and maximum elements were found, and were used to compute an offset (equal to the minimum value) and scaling factor (equal to the range), unique for a given control matrix. The fixed point control matrix elements are computed by subtracting the offset and dividing by the range before being scaled by $2^b$, where $b$ is the number of bits used to store the fixed point representation.

3. Compressed Floating Point Format

A control matrix is far from homogeneous, with a large range of values. This suggests that fixed point representation may not be ideal, since the relative resolution of small values is low and so will influence the wavefront error to a greater extent. We therefore consider a compressed floating point representation, which is able to cover the full range of 32 bit floating points, but with a reduced precision. Standard IEEE 32 bit floating point values have 8 bits dedicated to the exponent, 23 bits dedicated to the mantissa, and a single sign bit. A compressed floating point format that reduces the precision of the mantissa can be investigated. In AO, wavefront slope measurements are commonly computed using a center-of-gravity measurement, which, in good conditions (high light level, low detector noise), is typically assumed accurate to, at best, a 100th of a pixel, and in practice, is far less accurate. With, say, 20 × 20 pixels per subaperture, we can assume that there are 2000 measurable spot positions across the subaperture, which can be encoded in 11 bits. Therefore, we can predict that a mantissa of a compressed floating point number need be no more than 11 bits wide.

By running AO simulations with a range of bit widths for the mantissa, we find (Fig. 7) that AO system performance is not degraded until fewer than 10–12 bits are used for the mantissa, which is represented by a compressed floating point number requiring between 19 and 21 bits in total. However, this is a greater storage requirement than we have shown to be required when using a fixed point representation.

<table>
<thead>
<tr>
<th>Number of layer</th>
<th>Layer heights</th>
<th>Layer strengths</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0, 12800</td>
<td>0.92, 0.08</td>
</tr>
<tr>
<td>3</td>
<td>0, 1800, 12800</td>
<td>0.77, 0.17, 0.06</td>
</tr>
<tr>
<td>4</td>
<td>0, 1800, 4500, 12800</td>
<td>0.67, 0.15, 0.13, 0.05</td>
</tr>
<tr>
<td>5</td>
<td>0, 300, 1800, 4500, 12800</td>
<td>0.53, 0.21, 0.12, 0.10, 0.045</td>
</tr>
<tr>
<td>6</td>
<td>0, 300, 900, 1800, 4500, 12800</td>
<td>0.47, 0.18, 0.11, 0.1, 0.09, 0.04</td>
</tr>
</tbody>
</table>

Fig. 4. Layer heights and relative strengths used for multiple layer simulations.

Fig. 5. Predicted EAGLE AO system performance as the sparsity of the control matrix is altered with the sparsity factor representing the fraction of the original control matrix present. Uncertainties are about 2% in Strehl ratio.

Fig. 6. Predicted EAGLE AO system performance with a fixed point control matrix representation. Uncertainties are about 2% in Strehl ratio.
4. Variable Precision Floating Point Format

By compressing the exponent, as well as the mantissa of a floating point number, we can further reduce the storage requirement for the control matrix. We represent a floating point number in the form

\[( -1)^s \times b \times a^e \times \left( \frac{a}{2} + m \right), \tag{1} \]

where \( s \) is the sign (1 bit), \( a \) is the base (2 for standard floating point representation), which is constant for a given control matrix, \( b \) is a scaling factor (constant for a given control matrix), \( e \) is the exponent value, and \( m \) is the stored mantissa value. As with standard floating point representation, the mantissa is stored without an implicit integer part, which can be assumed (if it was not there, the exponent value can always be changed to shift the mantissa), and this is represented in the equation by the addition of the mantissa (fixed point with a value of less than \( \frac{a}{2} \) with \( \frac{a}{2} \)). The exponent \( e \) is in standard two's-complement integer format.

To convert a standard control matrix into this format, the minimum and maximum values required for storage are first obtained. We then set requirements that the mantissa for the maximum value is all 1s, and the mantissa for the minimum (nonzero) value is all zeros, except for the final bit, which is set. The exponent for the maximum value has all bits set, and the exponent for the minimum value has all bits unset. A value of zero is represented by having all bits of the mantissa and exponent unset. These conditions allow us to find two unknown values, \( a \) and \( b \), which will allow us to store this control matrix with highest precision. We then proceed to convert the standard control matrix into the variable precision representation.

As can be seen from Fig. 8, using a 4 bit exponent and mantissa is sufficient for good AO system performance, i.e., a total of 9 bits per control matrix element (including a sign bit). Similarly, a 5 bit exponent and 3 bit mantissa, and a 6 bit exponent and 2 bit mantissa also provide similar performance (taking 9 bits per control matrix element). The memory storage requirement has therefore been reduced by almost a factor of 4. We have not investigated the effect of using a greater number of atmospheric layers and virtual DMs, although, at most, this will increase the number of bits required slightly.

By using variable precision floating point format for storage of the control matrix, the memory bandwidth requirement can be reduced by a factor of almost 4, which will greatly simplify the design of a real-time control system for EAGLE. Only a quarter of the FPGAs used by an uncompressed system would be required, with simplifications also made by reducing the number of inter-FPGA connections, and an increased reliability due to a reduced number of components.

5. Implementation in FPGA

Implementation of variable precision floating point format in a FPGA is trivial: A \( 2^{14} \times 4 = 256 \) element lookup table can be used to translate the stored 9 bit control matrix values (using the mantissa and exponent for the index into the lookup table) into standard 32 bit floating point values, to which the sign can then be inserted. A standard floating point multiplication routine can then be used during the matrix-vector multiplication.

The Virtex-6 family of FPGAs is the latest offering from the company Xilinx, one of the major manufacturers of these devices. This range includes devices with up to 1200 input/output pins, with 37 MB internal memory in the FPGA [similar to a central processing unit (CPU) cache], and a clock rate of up to 1.6 GHz. The design of a real-time control system could be carried out using internal memory only. In this case, to store 20 control matrices (one for each line of sight) each 4 GB in size (1 billion 32 bit floating point values), would require over 2000 FPGAs. Using
variable precision floating point can reduce this requirement to just over 600 FPGAs, although this is still an undesirably large number.

Alternatively, we can use external memory connected to the FPGA pins. With standard 32 bit floating point storage, and a 1024 bit wide memory bus (1024 FPGA pins connected to memory), we can access 32 values in each memory read. The remaining pins are reserved for the address bus and inter-FPGA communications. Assuming that memory can be accessed at the full FPGA clock rate (1.6 GHz), we will achieve a memory bandwidth of about $5 \times 10^{10}$ values per second. The requirement for EAGLE is a minimum of $5 \times 10^{12}$ values per second, so 100 FPGAs would be required. If, however, variable precision floating point storage is used, we could access 114 values in each memory read (with a 1026 bit wide FPGA bus), equating to $1.8 \times 10^{11}$ values per second, requiring 28 FPGAs for EAGLE, a far more attractive proposition to develop.

In practice, the memory bandwidth requirement may be increased to reduce the AO system latency. Here, we have assumed a latency of 4 ms, equal to the frame time (at 250 Hz). To achieve a latency of 1 ms, we would require an improvement in memory bandwidth by a factor of 4, which, in turn, would require 112 FPGAs to meet this requirement.

D. Wavefront Sensor Calibration

Shack–Hartmann-based wavefront sensors are slightly nonlinear due to the pixilated nature of the detector, meaning that position information is lost: the measured slope is not proportional to the actual wavefront slope across the subapertures. For closed-loop AO systems, this is not a problem because the degree of nonlinearity is small and, because the measured wavefront slopes are minimized by the DM, a linearity approximation works well. However, for typical open-loop systems, this is more problematic because large uncorrected wavefront slopes can be measured. Therefore, the corrected wavefront (uncalibrated) will have some additional error due to this nonlinearity. This error is enough to lead to reduced performance of the AO system, and is present regardless of the slope measurement algorithm used if this algorithm is linear (e.g., center-of-gravity, matched filter, and correlation algorithms). However, a suitable calibration of the WFS can be carried out, measuring the WFS estimated response to a set of known incident wavefront slopes (introduced by a flat mirror on a tip–tilt stage). During AO system operation, the uncalibrated measured wavefront slope can then be used to infer the true (calibrated) wavefront slope by interpolating from the calibration data. This calibrated measurement can then be used to perform a more accurate wavefront reconstruction.

We have performed Monte Carlo simulations using this technique for WFS calibration using a center-of-gravity slope measurement algorithm, and have investigated the number of calibration steps required for good AO performance. These simulations are based around the aforementioned EAGLE simulations. We have used Shack–Hartmann subapertures with $20 \times 20$ pixels each, and a pixel scale of 0.8 arcsec per pixel at a wavelength of 589 nm and, as before, Fried's parameter is 10.6 cm. The WFS calibration is performed over the entire subaperture field of view. This large field of view is due to the need to detect the elongated LGS spots, and due to the higher dynamic range of the open-loop WFS (spots are measured in open loop, so are not necessarily close to the center of subapertures, as is usually the case for a closed-loop system). Figure 9 demonstrates the degree of nonlinearity in the simulated Shack–Hartmann wavefront sensor, showing the deviation of measured wavefront slope from the true slope as the true spot position moves across the subaperture. A true (physical) sensor would display even more nonlinearity due to imperfect optics. Figure 10 shows the performance improvements achieved with increasing calibration accuracy (number of calibration steps), demonstrating that this linearization calibration is an important part of open-loop AO system operation. We see that in this case, at least 50 slope calibration measurements are required to achieve the best performance, each step corresponding to a spot shift of less than half a pixel. By performing this calibration, the Strehl ratio (relative to uncalibrated performance) is increased by over 25%, and so the design of an open-loop real-time control system should, therefore, incorporate this calibration step. It should be noted that the optimal number of calibration steps is dependent on the WFS spot size on the detector, and so will vary with instrument and atmospheric conditions.

E. Thirty Meter Telescope comparisons

The Thirty Meter Telescope project also has plans for a multi-object spectrograph with AO, called IRMOS [11, 12]. The results presented here show that the estimated performance of these systems (taking into account the many unknowns in the designs), both estimating 50%–60% ensquared energy in 50 mas.
It should be noted that the results presented in this paper have been for energy within 75 mas. When we use our simulation models to measure energy within 50 mas, this is typically about 1%–2% lower than the energy within 75 mas. This serves to strengthen the assumption that modeling of AO systems can yield reliable performance estimates.

4. Conclusion

We have performed full end-to-end Monte Carlo simulations of an AO system for EAGLE. Investigations reported here show that the atmospheric turbulence profile must be well known, with the heights of turbulent layers known to within a few hundred meters. We have also reported on an investigation of compressed reconstructor representations and find that it is possible to reduce control matrix memory requirements by almost a factor of 4 in the cases investigated, significantly reducing the complexity of an FPGA-based real-time control system. An investigation into the effect that the nonlinearity of Shack–Hartmann-based wavefront sensors has on AO system performance has also been carried out, demonstrating that a linearity calibration should be included in an open-loop real-time control system to improve performance.

This work is funded by the Science and Technology Facilities Council (STFC).

References