

THE HISTORY AND DEVELOPMENT OF THE ESO ACTIVE OPTICS SYSTEM

THE ACTIVE OPTICS SYSTEM IS THE FUNDAMENTAL OPTICAL CHARACTERISTIC OF THE ESO NEW TECHNOLOGY TELESCOPE (NTT) AND VERY LARGE TELESCOPE (VLT). THE NTT PIONEERED THIS SYSTEM. WITHOUT IT, THE VLT, WITH ITS THIN, VERY FLEXIBLE MIRROR, COULD NOT GIVE A USABLE OPTICAL IMAGE AT ALL.

R. N. WILSON

THE TERM “ACTIVE OPTICS” is normally, and I believe correctly, associated with the ESO system developed for telescopes with monolithic primaries and applied in the ESO NTT and VLT telescopes. Technical systems based on the same principles, also with thin meniscus primaries, are used in the other very large telescopes (8 m) of GEMINI (2x) and SUBARU (1x). Other important telescopes with monolithic primaries, but using stiffer lightweighted blanks, are also actively controlled, the most notable being the WIYN (3.5 m) telescope, the three 6.5 m telescopes of the MMT upgrade and the two Magellan telescopes, and the two 8.4 m telescopes of the Large Binocular Telescope (LBT). The other major branch of modern optical telescope development, that using segmented mirrors, was pioneered and exemplified by the two Keck 10 m telescopes. This has its own system of active control of the segmented primary. Although the aim is the same, the technologies involved in the control systems of monoliths and segmented mirrors are essentially different. This is the case because the flexure function of a monolith has no discontinuities, which are fundamental to the nature of segmented mirrors. The ESO active optics system is “closed loop” in the sense that correction is made by measurements in real time of the quality of a star image. This is not the case with the Keck telescopes which correct the primary with an internal active system. “Closed loop” in the ESO sense is also not necessarily used in the other active systems for monolithic primaries: some of them rely on “precalibration” of the flexure effects (see page 9).

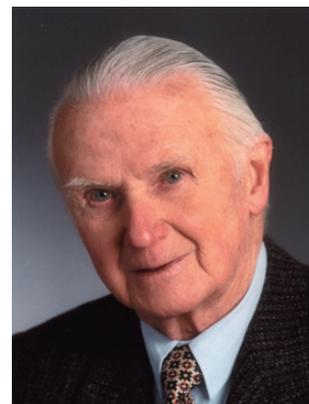
In this article, I shall confine myself to active optics with monoliths. When was the idea of active optics for monoliths first conceived? We must remember that all “classical” telescopes had monolithic

mirrors, normally made as thick and stiff as possible to avoid flexure. The idea of segmenting goes back to Lord Rosse in 1828, but was first realised in practice by Horn d’Arturo in the 1950s using a fixed primary. Keck I, finished in 1992, was the first such telescope with a normal 2-axis mounting. So it was logical that earlier ideas of active optics should have been limited to monoliths. The history of the development of mirror support systems and of active optics for monoliths is given in relatively complete form in my second book, *Reflecting Telescope Optics II* (Wilson 1999; W99 in the following). The present article is a simplified and much abbreviated version, with more emphasis on the personalities involved.

A friend of mine at ESO, who comes from the rich French tradition in telescope optics, recently suggested to me that active optics (and the Ritchey-Chrétien (RC) aplanatic telescope) might have been invented by one of my great French heroes in optics: Jean Bernard Léon Foucault. Reference is made in this connection to his largest – 80 cm – Newton telescope completed in 1862 (Wilson 1996). With all respect to the great genius Foucault, I believe that neither of these inventions would have been possible at that time. For active optics, no technology existed for *measuring* in a systematic way

the errors in a star image, although Foucault’s invention of the knife-edge test enabled a very sensitive *qualitative* assessment. The detector available, the eye, was sensitive but highly non-linear; photography was terribly slow and insensitive, and also non-linear. Third order aberration theory, due to Seidel, was only published in 1856 and existed only for spherical surfaces, not for a Newton parabolic primary. The complete theory for telescopes was only published in 1905 by Karl Schwarzschild. Lack of theory was also the reason Foucault could not have invented the *aplanatic* (RC) telescope. Both Schwarzschild (1905) and Chrétien (1910) used the Abbe “sine condition” (Wilson, 1996) as the basis for setting up aplanatic (i.e. free not only of spherical aberration, but also of field coma) telescope forms, unknown to Foucault. No, realistically Foucault, a scientific and technical genius, conceived and adjusted his mirror supports to get the best image he could. But this was *not* active optics: it was a procedure which had been used throughout the history of the reflecting telescope. It had been used empirically by James Short, William Herschel, Lord Rosse and others before the invention of modern support forms by Lassell in 1842 (astatic lever) and T. Grubb, also about 1842 (whiffle tree) (see W99).

I should like to dedicate this article to GERHARD SCHWESINGER (b. 08.01.1913 in Krappitz, Upper Silesia, d. 03.11.2001 in Heidenheim, Württemberg: see photo), who developed the first complete Fourier theory for the support of primary mirrors of telescopes and thereby also stimulated my thinking on active optics; and to LO WOLTJER whose vision and support led to the NTT and VLT based on my active optics concept.



We must conclude that active optics was anyway a concept of the 20th century. I believe the first ideas of a systematic process came from the great French optician Couder in 1931 and (probably independently) from the great Russian optician Maksutov in 1948* (W99). Couder recognised the high sensitivity to the aberration astigmatism of mirrors which were inadequately supported and suggested that such “regular” errors might be corrected by “a system of forces suitably applied”. He concluded that astigmatism left over from manufacture could be thus corrected. But as with Foucault, Couder had no means of *measuring* astigmatism in a star image. It could only be done *qualitatively*, off-line, in a slow process (effectively dc in modern terms of active optics). A rapid, repetitive, quantitative correction process was technically out of the question. Similarly, Maksutov proposed the adjustment of Lassell type astatic support levers to correct such errors observed with an ocular or a Foucault knife-edge, recognising too that the result was only valid for one zenith angle of the telescope. Since he had no mathematical algorithm for applying the correction, he suggested trial-and-error. Again, as with Couder, such a procedure was so cumbersome that its application could normally only be a once-off process at initial set-up (i.e. trial-and-error, qualitative, dc active optics). Understandably, although these suggestions were highly perceptive, they never led to any practical results or to serious further thinking. The time was still technically not ripe!

In 1968 I was already working, at Carl Zeiss in Oberkochen, indirectly for ESO, on a study for the optics of the ESO 3.6 m telescope. This made me aware of the *centering problem* of such Cassegrain telescopes. A lateral decentering tolerance of the axis of the secondary mirror to that of the primary, in order to maintain the optical specification of this telescope, would have to be set at well under 1 mm. In discussions with my colleagues in mechanics, it was clear that this was virtually impossible in practice, bearing in mind exchange operations of the top-end units. It became clear to me why the aberration produced (decentering coma, a particularly unpleasant, asymmetrical degradation of the star images) was the main curse of the Cassegrain telescope in practice: I called it “Cassegrainitis”. It struck me then that, *if one could measure its amount and direction*, its correction on-line would be a relatively simple mechanical operation requiring only a small lateral shift of the secondary to correct the error. This was the mental start of a com-

plete system of correction, which later became the ESO active optics system. In the course of further discussions with Dr Gerhard Schwesinger, a brilliant engineer and mathematician who also had excellent knowledge of optical aberration theory, I became aware of his general Fourier theory of the flexure aberrations introduced by support errors in primary mirrors. I realised that the circular nature of the mirror led to polynomial functions which were completely equivalent to those defining optical aberrations, although the mathematical boundary conditions are not the same. A light went on in my head! It would be perfectly possible to interpret all the flexure effects in terms of the classical optical polynomials of Hamilton or Zernike. Again, *if one could measure their amount and direction*, one could correct all such flexure terms by appropriate force changes of the supports, calibrated from the Schwesinger theory. Above all, this would apply to the lowest order term, by far the most important as Couder had recognised, astigmatism. This was particularly interesting, as Schwesinger’s calculations showed that maintenance of the absolute tolerances for the astigmatism specification was just as impossible in practice as the maintenance of an absolute decentering tolerance.

It followed that, when I formally joined ESO in September 1972, I had had the whole theoretical basis of active optics in my head for several years. However, not only the central problem of image measurement and analysis remained to be solved, it was also necessary to convince other colleagues and astronomers of the immense possibilities. In spite of the interest and vital information given by Sch-

wesinger, Zeiss showed no interest in pursuing it. I think an industrial concern, however brilliant and engaged in the matter, would anyway have been too far removed from practical telescope use. In other words, an observatory concerned with practical telescope development for a functioning observing site was essential. Of course, I had this in mind when I joined ESO. However, I soon learned that ESO had other problems far more urgent: the successful realisation of the 3.6 m telescope, on which its reputation and, indeed, its future existence depended. This was a conventional telescope following the line of the “Bowen-class telescopes” (Wilson, 1996) and had no significant innovative features. Nevertheless, at its completion, the 3.6 m telescope gave me the opportunity, during its optical set-up, alignment and test in 1976, to simulate the whole theoretical basis of an active optics system. The test system used was “classical Hartmann”. This was a painfully slow and exhausting process, measuring photographic Hartmann plates in a semi-automatic mode. But it enabled our team (essentially Francis Franza, Maurice Le Luyer and myself) to do a rigorous aberration analysis of the finished telescope. This led to my definition of the “Intrinsic Quality” (IQ) (see Fig. 1) of a telescope as that optical quality which would be achieved in principle if all the correctable terms measured could also be corrected in practice. In the 3.6 m telescope, there was no means of doing this: the primary mirror was too thick and rigid to allow it, even if a suitable support system had been available. Furthermore, a far simpler, rapid and on-line image analysis system than that used would be essential.

The essential elements of my active op-

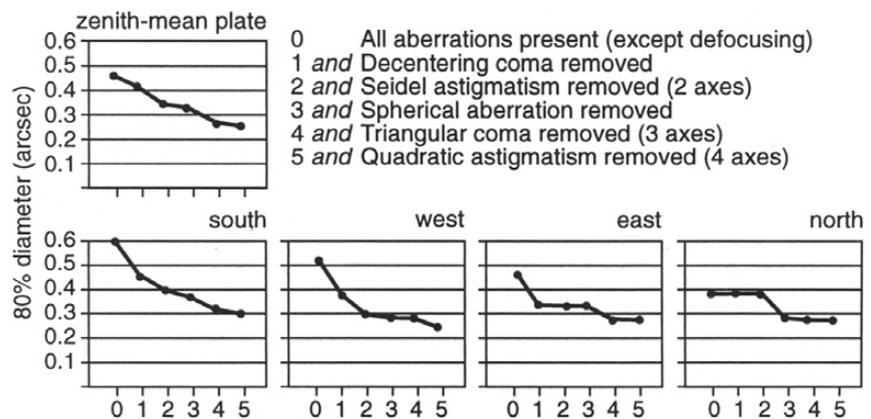


Figure 1: Results of classical Hartmann tests of the conventional ESO 3.6 m telescope in 1976, illustrating the theoretical improvement after successive removal of polynomial terms. The mean right-hand point of the functions gives the Intrinsic Quality (IQ) of the telescope. (W99)

*In W99, I expressed my gratitude to D. Enard and K. Bahner for drawing my attention to these proposals by Couder and Maksutov, respectively.

tics system were first published formally in Wilson (1977). At that time I called the concept a “feedback” telescope. The term “active optics” was used in a more explicit publication in Wilson (1982), which essentially gave the whole system of the NTT except for the practical details of the image detection and analysis. At the previous verbal presentation in 1981 at an optical conference in Graz, an American in the audience asked me at the end: “Have I understood you correctly, that you propose a telescope in which the optical system continuously checks itself and optimizes itself fully automatically?” I replied: “Yes, I congratulate you, you have understood it perfectly”. He replied: “Well, my feeling is that such a system will never be realisable in practice”. I hope he has since followed developments and registered what has emerged with the NTT and VLT.

As is always the case with radically new developments, parallel thinking had been going on independently by other groups. In 1970, a paper was published by Creedon and Lindgren in the American journal “Automatica”, a journal hardly known in the astronomical or optical communities. This work had been commissioned by NASA in connection with the 2.4 m Hubble Space Telescope (HST) and was reported in detail in secret NASA reports (W99). I only became aware of this work about 1985 through Oberto Citterio. The authors, who were brilliant control engineers but not versed in optical aberration theory, proposed a very complex mathematical scheme for the active control of the HST. In a (secret) NASA study of 1973, Howell and Creedon developed this approach further, improving it fundamentally by proposing a *modal* approach, also a fundamental feature of my own concept as thought out at Zeiss in 1968 with Schwesinger, from his own modal theory. However, Howell and Creedon’s algorithm was extremely complex and required the optimization of the *support geometry* according to the errors measured. In contrast, the ESO system algorithm can be applied directly to *any* normal passive support, whatever its geometry. Since the Howell and Creedon proposal was completely impractical, it was rejected by NASA. Because the HST primary has a very stiff, lightweighted primary, the forces required for active optics correction of the initial spherical aberration error, discovered after launch in 1990, would have been unrealisable in practice. With a thin, relatively flexible primary similar to the NTT, the ESO active optics system could have corrected it immediately. Although unknown to the general

telescope community at the time and not suitable for practical application, these studies had some valuable theoretical features, notably modal control with so-called “natural modes”. The measurement system proposed was not “closed-loop” in the ESO system sense, using a natural star image in real time, but an experimental precalibration of deformations of the primary for given forces and determined by interferograms. In any event, the astronomical community in the United States, with a few exceptions such as Aden Meinel, showed little awareness or interest in the potential of active optics until the late 1980s when the ESO NTT produced its first results. Before the NTT development at ESO, the same was true in Europe. In the early 1970s, a very cheap low quality 4 m IR spectroscopic telescope was built in France by Connes, Chevillard et al. (1989). This used a primary with 36 square segments, similar in principle to the later Keck 10 m project. However, unlike the Keck, it used a closed-loop feedback control system based on measurements of a natural star image, in principle like the ESO system for monoliths. The detector was a circular aperture passing flux to a photo-multiplier for each segment. Image analysis in the ESO system sense would hardly have been possible with this detection system. The image quality aimed for was very low, 10 arcsec, but the segments were so poor that the image was more like 10 – 20 arcsecs. The project was apparently known to some ESO astronomers, but was not communicated to the engineers: apparently it generated no interest. Finally, it was abandoned in 1975 because of total lack of interest and support in the French astronomical community.

THE ORIGINS AND DEVELOPMENT OF THE ESO NTT AND VLT

The ESO Technical report No. 8 in 1977, concerned with the image analysis of the newly set up 3.6 m telescope, gave the first account of the theoretical basis of an actively controlled telescope with a monolithic mirror with the ESO system. It was followed in the same year by the formal paper (Wilson, 1977) at the ESO Conference on “Optical Telescopes of the Future”. At this conference, Prof. Woltjer also gave a paper assessing the great merits, above all for spectroscopy, of larger ground-based telescopes and thereby justifying astronomically the construction of a 16 m telescope. He left it open whether this should be a single 16 m telescope, or an array of smaller telescopes. Such forward thinking was legitimate and necessary for ESO at that time following the

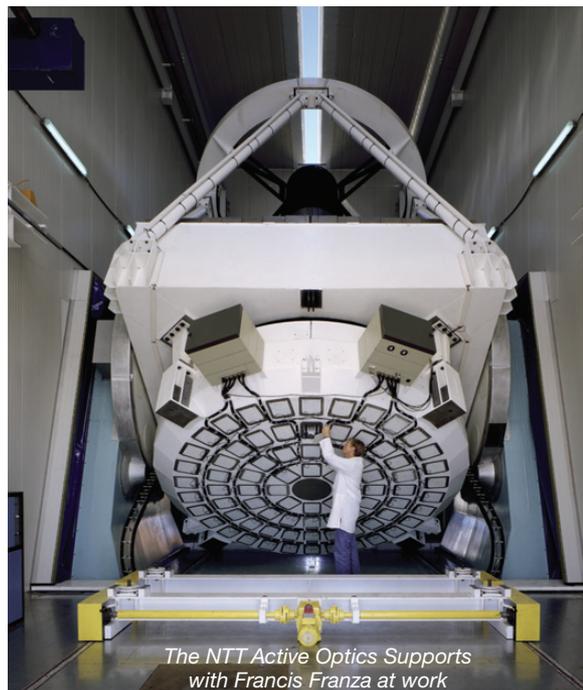
successful completion of the 3.6 m – undoubtedly essential for the future existence of the organisation. Immediately following the conference, Woltjer asked Richter, Chief Engineer in charge of the 3.6 m, to coordinate a study of three options: 1 x 16 m, 4 x 8 m, 16 x 4 m. Although Richter favoured the 1 x 16 m option, there was general agreement at ESO in favour of 4 x 8 m. I myself was strongly in favour of this option. A 16 m telescope was technologically too big a step in size, while 16 x 4 m represented *no* step in size and was politically too banal not to suffer reduction in the number of telescopes to save money. However, at that time, there was no clear idea what optical solution would be available for 8 m unit telescopes, let alone a 16 m. The concept of the 10 m Keck was just published, but design studies were only just starting. The Angel technology of lightweighted blanks (a further development of Palomar) had not then taken off, and the Multi-Mirror Telescope (MMT) was only completed in 1979 and comprised only small telescopes. Already in these discussions on an ESO VLT, I had my active optics solution in mind. But a trial on a smaller-size telescope seemed to me essential. Shortly after these discussions, the extension of ESO by the membership applications of Italy and Switzerland provided a marvellous new perspective at just the right time: the entrance fees could perhaps fund a new test telescope to try out new technology. Woltjer asked Richter to make a cost estimate for such a 3.5 m New Technology Telescope (NTT). At that stage, the only “new technology feature” envisaged was an alt-az mounting, which had been pioneered by the Russian 6 m telescope completed in 1976, but no western project had had the courage to take this fundamental step. This cost estimate enabled political support for the NTT to be marshalled.

At this time, I was spending a most instructive year on La Silla, learning about the practical (maintenance) problems of the many (and varied) telescopes at our ESO observatory. Everything I saw convinced me further that active optics was the only answer to the problem of maintaining optimum optical quality with telescopes in practice. Back in Europe in the summer of 1980, the big event for ESO was the move from Geneva to Garching in Germany. Since Richter left ESO at this point, the position in charge of the Telescope Group was free and Woltjer asked me to fill it. I was happy to do this, for my central interest had always been the telescope optics side, although my final position in Geneva had been in charge

of the Instrumentation Group. For the Telescope Group in Garching, although other tasks were not negligible (e.g. the building for the 2.2 m telescope of the MPIA or the achromatic plate for the Schmidt telescope), the NTT was the central project and was exactly the project I had longed for. I think I can truly claim that I determined every major characteristic of this telescope (the alt-az mount was taken over from Richter and was self-evident for a New Technology Telescope) – not always to the pleasure of all the astronomers. For example, I rejected not only a prime focus but also a Cassegrain focus, equipping the telescope only with two symmetrical Nasmyth foci. This meant that complex changes of foci as in the 3.6 m were totally avoided. With an act of great courage and expression of confidence in my knowledge of optics, Woltjer accepted my active optics concept, but imposed one entirely reasonable and prudent condition: that the NTT should work in the *passive mode* (i.e. without active optics) to the same specification as the classical (passive) 3.6 m telescope. This condition forced me to increase the thickness of the NTT primary from the 1:18 ratio I had envisaged to 1:15. This reduction in flexibility reduced the dynamic range of the active correction – see below concerning “First Light”. With the proven success of the active optics, 1:18 would have been better, but I still accept Woltjer’s imposed condition as correct at the time. The optics team was minimal in 1980 and consisted essentially of Francis Franza and myself with valuable assistance from Bernard Delabre on the optical design side. Quite early on I decided that it was too risky to go ahead with the final active optics system without a smaller scale experiment. This led to the 1 m-mirror experiment in the optics lab, in which Paul Giordano played a fundamental part. However, there was a serious lack in our optics group: a physicist to deal with the image analysis side and the necessary software development. This gap was filled by the engagement of Lothar Noethe, who came from Siemens. His application for the ESO job was one of the greatest pieces of good fortune in our whole active optics development. The presence of Francis Franza was the other essential pillar in the NTT development. Francis was engaged in 1973 in Geneva and from that time on we developed a perfect working symbiosis.

The 1 m-mirror experiment represent-

ed much work and was finally successful in demonstrating the practicability of the ESO active optics system. This was also our first trial of the image analyser based on the Shack-Hartmann principle. I had learned of this invention of Roland Shack (whom I had met at the Imperial College of London University in the early 1960s) about 1979, when he was a professor at the Optical Sciences Center in Tucson. I visited him with Francis and he was delighted at last to find someone who was deeply interested in his system of optical quality measurement and who wished to apply it immediately. The astronomical community in Tucson had shown no interest at all. Shack gave me an “S-H



The NTT Active Optics Supports with Francis Franza at work

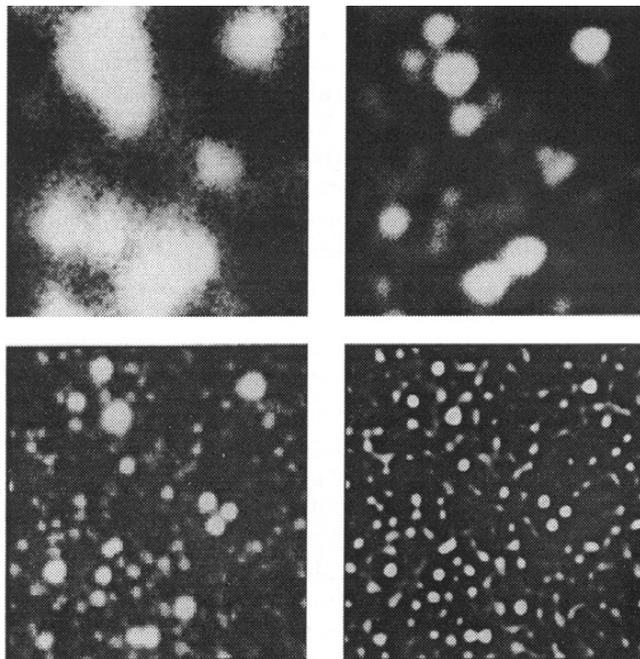
screen”, a raster of lenslets, which he had made mechanically on a lathe. This had a serious phase defect, but was usable and enabled us to operate our image analyser. Initially, the detection was with photographic plates and, to get rapid results, I favoured doing the whole 1 m-experiment with plates. But Lothar Noethe wished to initiate CCD technology and worked also with a CCD which had just become available. The final results were only published in 1988 (Noethe et al. 1988, W99), but already earlier they had given us full confidence to complete the NTT active optics system. The CCD as detector was fundamental: there was no serious alternative for our closed-loop system working rapidly in real time. The 1 m-experiment was also fundamental for the VLT study, initiated by Woltjer in the 1980s after the financial approval of the VLT as an ESO project and technically

led by Daniel Enard. He decided to follow completely the active optics, thin monolithic primary concept of the NTT. However, the VLT would require a much bolder approach than the (correctly) cautious approach of the NTT regarding the flexibility of the primary. The basic technical decisions on the VLT had therefore to be taken before the completion of the NTT First Light in March 1989. Essentially, the 1 m-experiment enabled those decisions to be taken with confidence.

Before the “Astronomical First Light” of the NTT we had what I called the “Technical First Light”. This was the first time I saw a star image in the newly erected telescope. No adjustments had then been made: this phase of the operation was now starting and included the so-called “dc phase” of active optics, i.e. the fixed, once-off corrections – see pages 6-8. The star image I saw in a hand-held eyepiece did not please me. I looked at the defocused image inside and outside focus, a classical test procedure I had used for decades on many telescopes. The appearance indicated strong spherical aberration, a defect which became world famous two years later, when this aberration, indicating a “matching error” in the forms of the primary and secondary mirrors, was revealed in the Hubble Space Telescope. I hoped for the best, thinking perhaps there was a strong thermal effect in the local air of the building at that time. However, with the tests of the image that followed, my fears were confirmed: there was indeed a strong spherical

aberration present. We were able to prove it was an error in the form of the primary. Exhaustive investigation, also with the manufacturer Carl Zeiss, showed that a spacer error had been made, as with the HST, in one of the so-called “null-test” systems. In fact, this error had been detected by a careful check of the system at Zeiss. However, owing to a misunderstanding of the sign of the spacer error, the error was corrected in the wrong direction, thereby doubling the resulting spherical aberration instead of eliminating it! The amount of spherical aberration was about the same as that found later in the HST. Our active optics system was able to correct it completely, although it used up about 80% of the dynamic range of correction available – see pages 8-9. This correction, saving a very costly reworking of the primary, was a marvellous demonstration of the power of active op-

Figure 2: CCD pictures obtained at “First Light” with the ESO 3.5 m NTT in March 1989, compared with previous records of the same field (globular cluster ω Centauri). Upper left, ESO 1 m Schmidt; upper right, ESO 3.6 m telescope; lower left, ESO NTT raw image; lower right, NTT processed image. See text for details. From West, R. (1989).



tics to extend vastly the manufacturing tolerances of correctable aberrations.

The “Astronomical First Light” results of the NTT were so fantastic that they established it as optically the best telescope in the world at that time (1989). As is well documented (W99), the conditions were extraordinarily good both for external seeing and dome seeing: we were remarkably fortunate. This was also due to another new technology feature of the NTT, taken over and improved, from the MMT. This was the building concept, whereby the building rotates with the azimuth movement of the telescope. We improved the MMT building concept by removing the back wall, thereby allowing ventilation to pass laminarily through the whole building “slit” for the telescope. This feature has been very important for the excellent optical quality and, in somewhat modified form, has been taken over for the VLT.

Only one new technology feature which I envisaged was not realised. This was a second primary with an aluminium blank. This was finally abandoned for cost and time-scale reasons, but I believe that this decision was an error. The NTT would have been a perfect telescope to test the viability of aluminium as a blank option. Excellent and reasonable offers existed both for the manufacture of the blank and its “Canegen” coating and for the optical figuring. The existing blank in Zerodur is, of course, excellent: but the extreme zero expansion property of Zerodur (or ULE fused quartz) is no longer necessary for actively controlled

telescope optics. The finite expansion coefficient of aluminium is largely compensated by its excellent thermal conductivity and active optics can easily handle residual expansion effects. Although interest in aluminium has since been shown, above all in France, no telescope of significant size has been equipped with an aluminium primary since the brilliant pioneer work of Mottoni for the Merate 1.37 m telescope (1969) in Italy (W99). This is unfortunate and demonstrates once again the inherent technical conservatism of the astronomical community: the refusal to abandon glass corresponds exactly to the inverse refusal to abandon metal (speculum) in the 1860s and to introduce chemically silvered glass! This led to the disaster of the Melbourne reflector, set up in 1869 (Wilson, 1996).

It follows from the above account that three very important aspects of the NTT technology came from the USA: the CCD detector, the Shack-Hartmann image analyser and the building concept. But the active optics concept for thin meniscus monoliths was a purely European development, which, apart from Roland Shack and Aden Meinel, was ignored or actively rejected in the USA until the “First Light” success in 1989.

The success of the ESO active optics concept has been wonderfully demonstrated by the best Full-Width-Half-Maximum (FWHM) star images recorded for the NTT and VLT. With “First Light” in March 1989, the NTT revealed a best star image for a CCD frame in the globular cluster Centauri of 0.33 arcsec FWHM –

a world record at that time for a ground-based telescope. Richard West identified the field (only 12 x 12 arcsec because of the small size of the CCD used directly at the Cassegrain focus) and set up a beautiful comparison (West 1989, W99) which is reproduced here (Fig. 2). Upper left shows the field, suitably magnified, taken from a plate by the ESO 1 m Schmidt telescope in 1984 under modest seeing conditions (ca. 2 arcsec). Upper right is from a plate, considered excellent by normal standards, by the passive 3.6 m telescope with seeing about 1 arcsec. Already with this improvement, the “clumps of cotton wool” representing star images in the Schmidt plate have vastly improved. At the bottom left is the “First Light” frame with the NTT. The five separate images of the Schmidt became about 15 with the 3.6 m and number almost 100 with the NTT. (The bottom right frame shows further improvement by off-line processing; but this cannot be compared with the other three which are “raw images”). The enormous gain in resolution is striking and well-illustrated by the triple star right of centre. The Schmidt shows no resolution, with the 3.6 m the triple nature can be inferred without resolution, while the NTT resolves the three components completely. But at least as significant as the gain in resolution is the gain in *light concentration per star image*, giving a huge increase in depth penetration for the same exposure time.

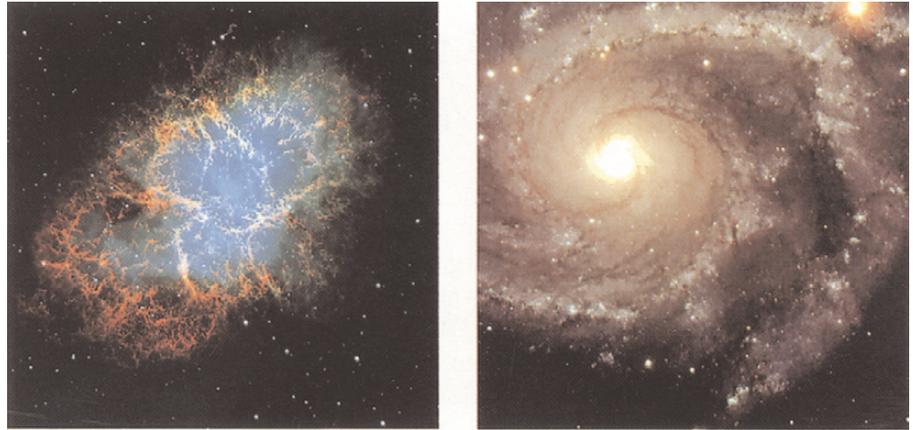
Figure 3 reproduces the frontispiece of my second book (W99) and shows the UT No. 2 (Kueyen) of the VLT together with two beautiful photographs. The photo at the top right, a three-colour composite of the Spiral Galaxy NGC 2997, was taken with UT No. 1 (Antu) and the FORS 1 instrument on 5 March, 1999. In the near IR band, the FWHM of the best star images was 0.25 arcsec, a record at that time.

SOME BASIC PROPERTIES OF THE ESO ACTIVE OPTICS SYSTEM

(a) Automatic optical Maintenance

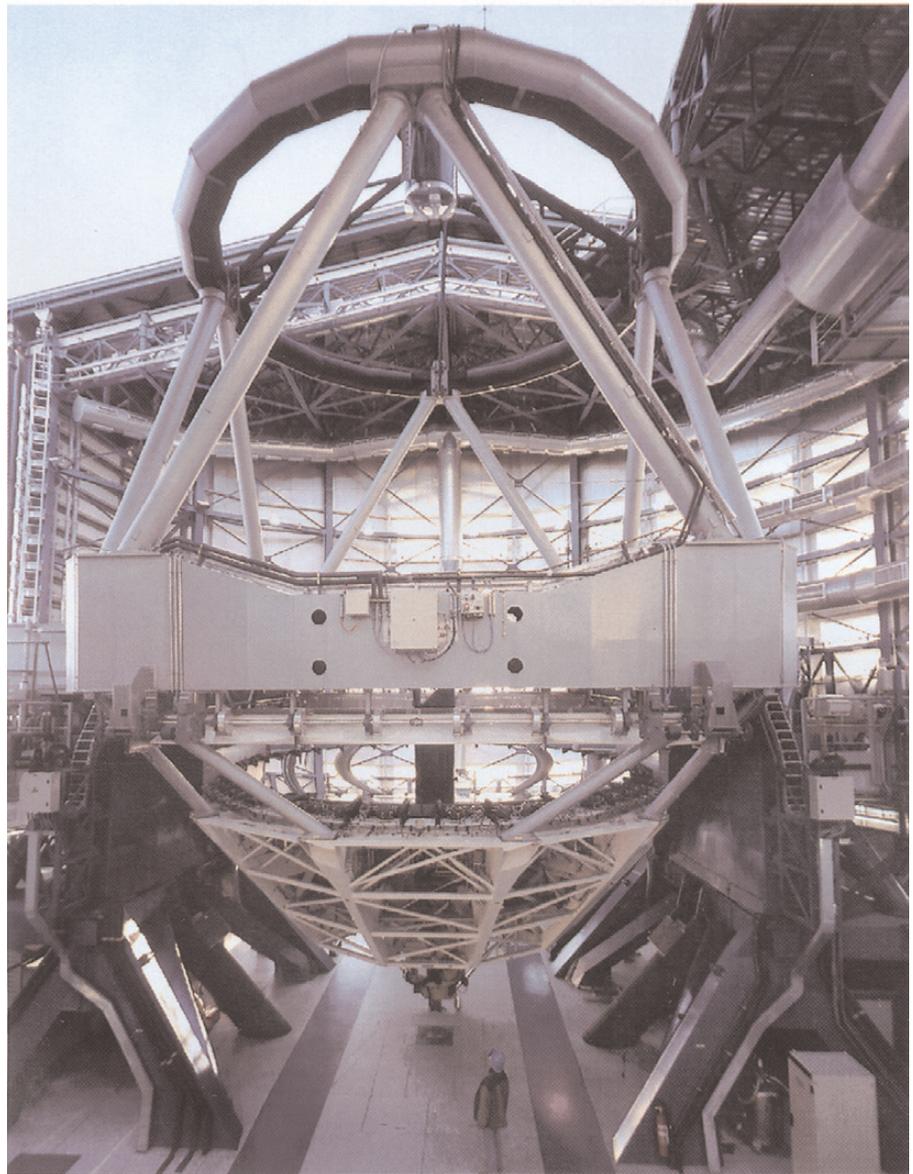
As indicated in the first section, the ESO active optics system was conceived to remedy what seemed to me the most intractable problem of the “passive” telescopes built up till about 1980: the problem of optical maintenance of the finished telescope. The optical specifications of such telescopes (e.g. the ESO 3.6 m) were much inferior to those which have become normal for “active” telescopes, but they were still good. Furthermore, they were largely met by the manufacturers. The problem was that they could rarely be maintained in practical observatories. Often, they could be re-established

Figure 3: ESO VLT Unit Telescope No. 2 photographed in March 2000 by Hans-Hermann Heyer. The photo upper left shows the Crab Nebula taken with UT No. 2 and the FORS2 instrument on 10 November 1999. The photo upper right was taken with UT No. 1 and the FORS1 instrument on 5 March 1999 and shows the Spiral Galaxy NGC 2997. The best star image quality (in the near IR band) had a FWHM of 0.25 arcsec.



by careful adjustment in a complex off-line operation, but most frequently they declined again long before the next such operation. The aim of active optics was therefore to *automate* the whole optical maintenance procedure. In the NTT, the design aim was to re-optimize the quality automatically with a cycle time of *about 10 min*. This automation was never realised in my time (it was still initiated by hand). The NTT is very robust in its design (see previous section), and this manually initiated procedure could also ensure good quality. However, it depended on procedures being followed and, in the real world, this does not always happen. The VLT telescopes are not at all robust in this sense, because the primaries are about *50 times more flexible* than that of the NTT. Without its active optics, the VLT cannot produce a usable image. Fully automatic operation is thus essential, and optimization is performed *every 40 s*. Therefore, the period over which the optical performance can decline is reduced from what used normally to be weeks for a passive telescope to 40 s. Furthermore, the optimization is always complete, recovering fully the maximum potential of the telescope, whereas the old-fashioned, off-line procedures were rarely fully effective simply because the telescope was inevitably out of commission and there was always great time pressure. Also, the telescope designs were rarely “maintenance friendly” for the optics. The optimization cycle of the VLT means essentially that the optical quality must maintain itself for the change of zenith distances involved in tracking for 40 s.

Originally, the NTT software had no provision for automation of the active optics correction because we knew we had to learn from experience how this could best be realised. The initial huge success after First Light was therefore achieved by purely manual operation. By the end of 1990, we had sufficient practical experience to make an attempt at automation possible, but organisational changes prevented any further advance in practice. Finally, the wise decision was taken to use the NTT as a test bench for the new VLT software, which was installed about 1996.



This included, of course, the fully automated active optics correction cycle. Recently, the VLT software for the active optics has been further improved to eliminate aberration effects of the air and to make the choice of reference star more flexible (i.e. easier) because the bright-

ness and colour are less critical. This was from the start a problem with the NTT, for which the availability of sufficiently bright stars was uncertain. This problem was exacerbated by the raster of the Shack-Hartmann detector, which was laid out cautiously with 40×40 sub-aper-

tures. With more experience, this was reduced to 20×20 sub-apertures for the VLT, which anyway gives more light because of the larger aperture. Although the S.-H. raster is unchanged, it has been possible (according to current information from Olivier Hainaut) to operate a partial automation of the NTT active optics for astronomical exposure times longer than ca. 2 – 3 minutes and under stable conditions for the residual aberrations. Typically, about 5 – 7 image analyses are averaged out, giving then a correction every 5 – 10 minutes. This procedure is therefore close in cycle time to my original proposal.

(b) Frequency bandpasses of active and adaptive optics

Table 1, taken from W99, shows the basis of all this thinking in terms of the “Bandpass” or frequency of all the sources of image degradation. The most important conclusion from this Table is that all the error sources are dc or of bandpass $<10^{-2}$ Hz except (8), (9) and (10) and partly (7). This is of central importance, because it is roughly the frequency limit of *normal* active optics correction in *closed-loop* and implies that two thirds of all the errors listed are amenable to it. The definition of the *correction bandpass* is an essential feature of any control system. The situation with my definition is shown in Fig. 4.

The normal active optics bandpass A, as defined for the NTT, goes from dc to 1/30 Hz. The limit of 1/30 Hz simply corresponds to the well-known fact that, in the presence of good astronomical external seeing, an integration time of 30 s is sufficient to “integrate out” the external seeing completely, giving a round image corresponding to the external seeing

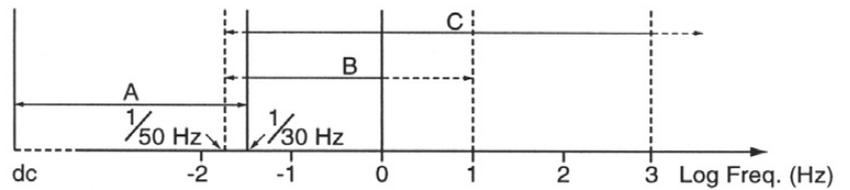


Figure 4: The bandpasses for active and adaptive optics correction. From original publication by Wilson and Noethe 1989 and W99.

quality, the classical definition of “seeing”. (If the integration time is inadequate, the image analyser will give completely erratic and wrong results since random aberrations of the external seeing are included). For a frequency higher than 1/30 Hz (for inferior seeing at somewhat lower frequencies), we enter into the *adaptive optics bandpass* C, going to beyond 10^3 Hz, for the external seeing. In this bandpass, we are confronted with the phenomenon of the *isoplanatic angle* Θ (W99), that angle over which the phase of the error introduced by atmospheric seeing is essentially constant. Θ is a function of seeing quality, wavelength and the frequency. For visible light and an extended frequency band, the value of Θ is only a few arcsec and even at the lowest frequencies of bandpass C amounts to only one or two arcmin at most. For *closed-loop operation*, for which a reference star within the isoplanatic angle is required, this is a very serious limitation *compared with bandpass A for active optics*, for which there is no isoplanatic angle limitation at all and a reference star at any convenient point in the field can be used. The bandpass B, which I call the *extended active optics bandpass*, goes from 1/30 Hz to about 10 Hz. This is particularly important for Error 7 of Table 1, of which the higher fre-

quency component can only be actively corrected within the limits of the isoplanatic angle. For the external seeing in general, unlimited correction will only be possible using artificial, laser-generated reference stars.

(c) Modal control

A *modal* concept (i.e. successive terms of some polynomial with increasing powers of the parameters involved) has always been normal practice in optical design based on the theory of optical aberrations. As explained in the first section, I realised from discussions with Dr Schwesinger at Carl Zeiss in 1968 that the flexure terms in his theory of elastic flexure of circular mirrors could be interpreted in a similar way. The whole theoretical basis of active optics was already clear to me. A modal basis was thus clear from the start. It is a fundamental property of physics, linked to thermodynamics, that so-called “higher order terms”, involving higher powers of the polynomial parameters (essentially the radius, the thickness and the azimuthal orientation in cylindrical mirrors) require more energy for their generation and are more stable than “lower order terms”. This is embodied in the principle of St Venant, fundamental to this application of elasticity theory. The

conclusion is of great importance for active optics: low order terms such as defocus and astigmatism can occur in telescopes readily and vary rapidly and require relatively low forces to generate them, such as the gravity effects due to telescope movement. Beyond a certain (high) order, conversely, gravity effects produce effects which are optically negligible. The corollary is a very simple basic axiom of active optics: if forces of the order of the gravity forces on the supports can produce an optical error of significance, then correcting forces *of the same order of magnitude* can correct it, if we can determine how and where to apply them! Conversely, a higher or-

Table 1: The ten sources of error giving degradation of image quality in ground-based telescopes, and their corresponding bandpasses. Diffraction, which is inevitable and continuous, is excluded since (for a given signal wavelength) it cannot be influenced. In space, the three errors dependent on air vanish (W99).

SOURCE OF ERROR	BANDPASS (Hz)	
(1) Optical design	dc	(fixed)
(2) Optical manufacture	dc	(fixed)
(3) Theoretical errors of: - Mirror supports - Structure (focus, centering)	dc \rightarrow 10^{-3} 10^{-3}	(fixed \rightarrow minutes) (minutes)
(4) Maintenance errors of the structure and mirror supports	$10^{-6} \rightarrow 10^{-5}$	(weeks \rightarrow days)
(5) Thermal distortions - Mirrors - Structure	$10^{-5} \rightarrow 10^{-4}$ 10^{-3} 10^{-7}	(days \rightarrow hours) (minutes) (years)
(6) Mechanical distortion of mirrors (warping)	$10^{-4} \rightarrow 10^2$	(hours \rightarrow 0.01 s)
(7) Thermal effects of ambient air (telescope, dome and site “seeing”)	$10^{-2} \rightarrow 10^1$	(minutes \rightarrow 0.1 s)
(8) Mirror deformation from wind gusts	$2 \cdot 10^{-2} \rightarrow 10^3+$	(50 s \rightarrow $< 10^{-3}$ s)
(9) Atmospheric turbulence (external “seeing”)	$5 \rightarrow 10^2$	(0.2 s $\rightarrow 10^{-2}$ s)
(10) Tracking errors		

der error which cannot be corrected by forces of this magnitude will also not be generated, i.e. it will not be present. In other words, any error present which is due to elasticity and gravity can be corrected. Later, I realised that this modal approach is also mathematically essential for finding a practical solution. Mathematically, it would seem simple and elegant to measure the total aberration error at many points of a rectangular raster over the pupil. If calibrations exist for the aberrations produced by known force distributions, then a so-called matrix inversion would give a solution reducing the optical error at all the raster points to zero. However, this procedure would include *all* the aberration orders, including higher order effects which are negligible in practice, *but not zero*. These terms would only be correctable by impossibly high forces. The result would be a solution matrix with an inaccessible solution: mathematically a solution matrix with enormous eigenvalue ratios. The modal approach avoids all such problems, provided the modes are reasonably determined. In the NTT, because the primary was relatively stiff, seven modes were sufficient. In the VLT, with a primary about 50 times more flexible, 16 modes are corrected. Fig. 5 shows the nature of the deformations produced by these modes. The use of these so-called “natural vibration modes” is one of the great contributions of Lothar Noethe (Noethe 1991, W99). The obvious optical alternative of Zernike polynomial modes, orthogonal modes commonly used in optical design, is quite feasible. But the dynamic range of correction for a given range of forces is an optimum for the natural modes, a very important advantage.

(d) Closed-loop operation

Reference has been made several times above to the fact that the ESO active optics system is a *closed-loop* system performing corrections at frequent time intervals by measuring the errors in a star image in real time. The development of this image measurement system, based on the Shack-Hartmann detection principle, was not trivial, but has long been standard technology at ESO through the NTT and VLT. In a VLT unit telescope, up to 1000 image analyses might be made in a single winter night! In the first section, it was indicated that an alternative approach is by precalibration of aberrations as a function of zenith distance in an alt-az mounted telescope, and that this approach is used in some other projects. The ESO viewpoint is that precalibration can be a reasonable approximation in many cases,

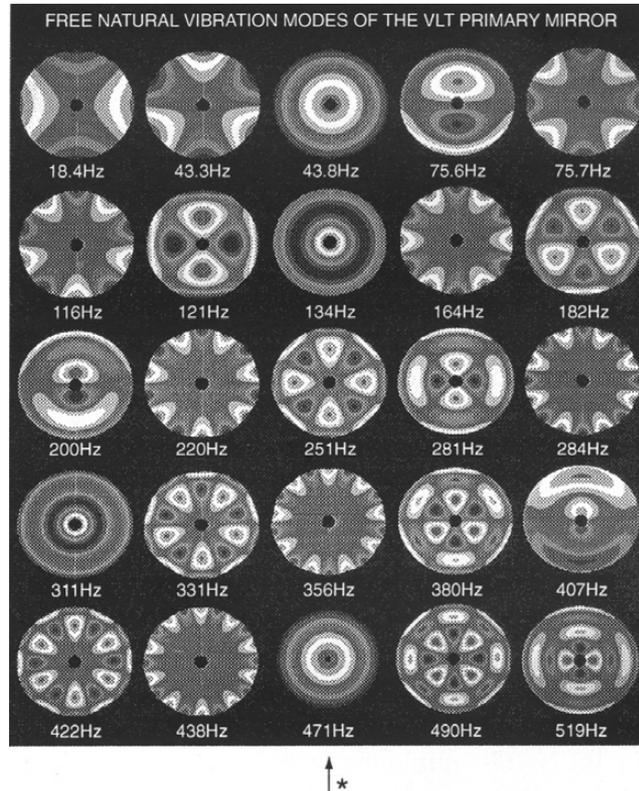


Figure 5: The first 25 natural vibration modes of the VLT primary. The first 16 of these are corrected by the active optics system. The modes shown were calculated for a thickness of 200 mm, whereas the final thickness of the mirror is 175 mm (W99).

but can never rival the repeated direct measurement of the actual aberration in the telescope image. Since there is no isoplanatic angle problem and there is no unsolved practical problem of applying image analysis using CCDs in big telescopes, my view is that the *closed-loop* system with image analysis is the optimum way of performing active optics. This was my intention from the start of my complete theoretical concept of active optics in 1968, although I knew of no technical solution for real-time image analysis at that time.

CONCLUSION

As with all technical developments departing radically from accepted technology, it took a long time, 21 years, between my first theoretical basis of active optics in 1968 to its final practical confirmation with the NTT in 1989. Without the confidence and support of Prof. Woltjer for the NTT and VLT, who knows whether it would have been tried in practice to this day? The significance of active optics seems to me, in hindsight, greater today than in those early years. Together with the segmented technology of the Keck telescopes, it enabled the breakthrough of both the technological quality barrier and the cost barrier presented roughly by the 5 m Palomar telescope. Future huge telescope concepts, such as the 100 m OWL of ESO (Dierickx et al. 2003), would be inconceivable without these modern

active optics technologies, both for their optical function and for their cost. Finally, for the reader interested in a full account of the current status of active optics, the best reference is a recent review article by Lothar Noethe (2002).

ACKNOWLEDGEMENTS

I am most grateful to Prof. Woltjer for discussions and clarification regarding the early years of the NTT and VLT developments, and to Lothar Noethe for technical discussions. Thanks are also due to Olivier Hainaut for information on the current status of the NTT. The editor of *The Messenger*, Peter Shaver, has also been most helpful regarding the content. My gratitude is due to him for suggesting the article in the first place.

REFERENCES

- Connes, P., 1989, SPIE Conf. Active Telescope Systems, F. Roddier ed *SPIE* vol. 1114, 238
- Dierickx, P. et al., 2003, *SPIE* vol. 4840, 151
- Noethe, L. et al., 1988, *J. Mod. Optics*, 35, 1427
- Noethe, L., 1991, *J. Mod. Optics*, 38, 1043
- Noethe, L., 2002, Active optics in modern large optical telescopes, *Progr. Optics* 43, E. Wolf ed, 1
- West, R., 1989, *The ESO Messenger*, 56, 3
- Wilson, R.N., 1977, *Procs ESO Conf. Optical Telescopes of the Future*, ESO, Geneva, 99
- Wilson, R.N., 1982, *Optica Acta*, 29, 985
- Wilson, R.N., 1996, *Reflecting Telescope Optics I*, Springer-Verl. Berlin & Heidelberg, 2nd corr. repr. 2002
- Wilson, R.N., 1999, *Reflecting Telescope Optics II*, Springer-Verl. Berlin & Heidelberg, corr. repr. 2001(W99)