The revolution in telescope aperture

Matt Mountain & Fred Gillett

The 40 years that have elapsed since the opening of the Kitt Peak National Observatory have seen several revolutions in astronomy. One such revolution that is often not recognized is the exponential growth in the total collecting area of large telescopes. This revolution, although ultimately driven by curiosity, has come about because of advances in computers, materials and fabrication techniques—these advances have facilitated the construction of telescopes that are not just bigger, but better as well. The enormous growth in ground-based observational capabilities that will become available over the next decade, combined with future generations of space-based telescopes, should contribute to dramatic changes in our understanding of the Universe. The first hints of this new understanding are already beginning to emerge, giving new directions to future telescope builders.

From the time of Galileo, 400 years ago, we have gained a deeper appreciation of the Universe and our place in it by observing and analysing the appearance of the night sky with telescopes. Seeing the moons orbiting Jupiter convinced Galileo to place the Sun, not the Earth, at the centre of the Solar System. Observations made with the Mt Wilson 50-inch telescope showed that the Sun is in the outskirts of our own Galaxy. Measurements made with the world’s first 100-inch and 200-inch telescopes showed that even the visible matter of galaxies was merely the tip of an iceberg. Vast quantities of dark matter seem to dominate the dynamics of all galaxies and probably the Universe itself. Today’s observers, using the Keck 10-m telescopes and the new generation of 8-m telescopes, are analysing the constituents of early galaxies within a few billion (10⁹) years of the creation of the Universe.

This ground-based capability for observing and studying the Universe at optical and infrared wavelengths has grown from Galileo’s first telescope, with a collecting area of about 1/2000 m², to an ensemble of telescopes that, by the end of this century, will have a total collecting area of over 1000 m² (Fig. 1). Of this growth in collecting area, 80% will have occurred in the last 10 years of this century, and will be dominated by the emergence of a new class of 8-m to 10-m diameter telescopes. This tremendous growth in telescope collecting area has been driven by three factors. The first, and most fundamental, is the curiosity-driven willingness to invest in astronomy, inspired by the discoveries of the past century. Most of this recent investment is in national and international facilities, and continues the trend to internationalize astronomy.

The Associated Universities for Research in Astronomy (AURA; the parent body of the National Optical Astronomy Observatories) must take much of the credit for starting these international partnerships. The second factor is the profound revolution in computer and materials technology that has occurred over the past few decades. This has allowed the fabrication of large, high-quality mirrors with short focal lengths, and sophisticated computer-controlled telescopes and instruments. Lastly, advances in adaptive optics at near-infrared wavelengths should soon allow us to build large telescopes that are essentially diffraction limited. This will result in sub-arcsecond resolution, like that of the Hubble Space Telescope (HST) combined with the increased collecting area of an 8–10-m telescope.

The challenge of building large telescopes

Before the development of high-performance charge-coupled-device detectors, increasing the light-collecting area of telescopes was the main goal of telescope builders. The reason for this is simple: if too few photons per unit area are reaching the detector (which is least 20 times fainter than a conventional 4-m telescope delivering 1.0 arcsec images. Will we see such gains in the next generation of telescopes? MMT, Multiple Mirror Telescope; CFHT, Canada–France–Hawaii Telescope; WHT, William Herschel Telescope; NTT, New Technology Telescope; HET, Hobby–Eberly Telescope; VLT, Very Large Telescope; Gem-N, Gemini North Telescope.

Figure 1 The growth in cumulative telescope collecting area over the past 400 years, with each point representing a completed ground-based telescope. The combination of the ability to manufacture and support large mirrors combined with adaptive optics has grown the new generation of large telescopes tremendous scientific gains over the previous 4-m telescopes. For example, an 8-m telescope delivering images of 0.1 arcsec can observe point-like objects at...
more pronounced as spatial or spectral resolution is increased), the noise of the detector; in this limit, the signal-to-noise ratio is proportional to the collecting area of the telescope. Consequently, the main challenge that most telescope builders addressed was how to manufacture and support ever-larger primary mirrors to collect more light.

The larger the mirror, the greater is the need for mechanical stiffness to counteract the effects of glass deformation due to its own weight and wind forces. This classically requires more massive mirrors and supports, which in turn result in more structure solely by the noise of the detector; in this limit, the signal-to-noise ratio can vary considerably as is shown in the table (the lower the number, the smaller the thermal distortion resulting from a given temperature transient).

### Materials used for current and possibly future large telescope mirrors

<table>
<thead>
<tr>
<th>Material</th>
<th>Thermal susceptibility to distortion $\alpha_{\text{C}/K}$</th>
<th>Normalized to glass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass</td>
<td>15.68</td>
<td>1.00</td>
</tr>
<tr>
<td>Pyrex (borosilicate)</td>
<td>3.67</td>
<td>0.30</td>
</tr>
<tr>
<td>Fused quartz (fused silica)</td>
<td>0.97</td>
<td>0.04</td>
</tr>
<tr>
<td>Ultra-low-expansion glass</td>
<td>0.04</td>
<td>0.002</td>
</tr>
<tr>
<td>Zerodur</td>
<td>0.15</td>
<td>0.010</td>
</tr>
<tr>
<td>Aluminium</td>
<td>0.36</td>
<td>0.023</td>
</tr>
<tr>
<td>Beryllium</td>
<td>0.10</td>
<td>0.011</td>
</tr>
<tr>
<td>Silicon carbide (CVD)</td>
<td>0.22</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Typical values for these materials are taken from Yoder and the precision ratio depends on the exact type and grade of material under consideration. Aluminium, beryllium and silicon carbide have been included for completeness.

### Box 1 Mechanical and thermal deformations

#### Mechanical.
Analytically, the deflection $\delta$ (normalized to an observing wavelength $\lambda$) of a circular plate due to its own self-weight supported by three discrete points is described by:

$$\delta \lambda \propto \frac{D^4\rho}{E t^4} \lambda$$

where $\rho$ is the density, $D$ is the diameter, $E$ is the Young’s modulus and $t$ is the plate thickness.

For the case when a uniform pressure $P_0$ (such as wind) is applied to a circular plate the normalized deformation becomes:

$$\delta \lambda \propto \frac{P_0 D^4}{E t^4} \lambda$$

In both cases as the diameter of a mirror grows, the thickness and hence the mass of the mirror must grow even more quickly to maintain the same stiffness against gravity or wind. Because this approach soon becomes prohibitively difficult and expensive, modern telescope designers have moved towards using thinner and lighter mirrors, using a complex array of supports to maintain the mirror figure against both gravitational and wind forces.

The ratio $\rho/E$ changes by only $\pm 15\%$ for a wide variety of glassy and glass–ceramic materials used in today’s large telescope mirrors. In the future, if the costs can be brought down, large telescope mirrors could be made of beryllium or silicon carbide. The $\rho/E$ ratios of these materials are one-fifth to one-sixth those of glass or glass–ceramics, allowing far stiffer and lighter mirrors than is possible with the less expensive materials used today.

#### Thermal.
The susceptibility of the mirror figure to thermal distortion is determined by its thermal properties such as the coefficient of thermal expansion $\alpha$ ($10^{-5} \text{ K}^{-1}$), its specific heat capacity $C$ ($\text{J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$) and its conductivity $K$ ($\text{W} \cdot \text{m}^{-1} \cdot \text{K}^{-1}$). For a given change in temperature (a thermal transient), the thermal distortion experienced by a mirror is proportional to the ratio:

$$\alpha C_0 P_0 K$$

where $\alpha$ is the density of the mirror ($10^3 \text{ kg} \cdot \text{m}^{-3}$).

Unlike the mechanical properties of glasses and glass–ceramics, this ratio can vary considerably as is shown in the table (the lower the number, the smaller the thermal distortion resulting from a given temperature transient).

### Optical telescopes 8–10 m in diameter

The key design feature for the 8–10 m class telescopes was to greatly increase the ratio $D/t$ of the primary mirrors to reduce the mass of the mirror, the supporting telescope and the surrounding structure. The availability of analytical models opened the way to understanding
the detailed performance of these lighter mirrors and telescope structures and the active systems needed to support the mirrors. Computers also enabled the precise control of the altitude—azimuth telescope mounts, techniques pioneered by the radio astronomy community. These techniques were successfully extended to optical telescopes by the UK’s 3.8-m Infrared Telescope (UKIRT) and University of Arizona/Smithsonian Institution’s Multiple Mirror Telescope (MMT) in 1979 (ref. 17). Telescope designers were released from the Palomar legacy.

In the 1980s, KPNO and the Universities of Arizona, California and Texas initiated the study of a number of new telescope approaches. The University of California at Berkeley began investigating the fabrication of 10-m primary mirrors from individual segments of low-expansion glass, which would lead to the two Keck telescopes. Edge sensors attached to each mirror segment would feed corrections to 180 actuators controlling the 36 polished hexagonal segments, ensuring that the entire ensemble would behave as a single mirror with an effective aspect ratio for the glass of 133:1. The University of Texas at Austin adopted a fixed, segmented primary and a movable secondary for its rather radical Hobby–Eberly telescope, after initially investigating a meniscus design for a 300-inch telescope.

The University of Arizona began developing large, structured mirrors made from borosilicate glass. At 16 tonnes an 8.4-m-diameter structured borosilicate mirror weighs almost the same as the Kitt Peak 4-m mirror and the Keck 10-m primary mirror (14.4 tons of glass). However, the modern demand for superb image quality still means that these types of mirror require a distributed and active support system. Arizona has adopted a matrix of 164 active, computer-controlled supports for the 8.4-m primary mirrors of its Large Binocular Telescope.

At KPNO, in Europe and in Japan, studies looked at the feasibility of using thin meniscus mirrors with aspect ratios ($D/m$) of 40:1 or higher, using new materials such as Corning’s Ultra Low Expansion Glass (ULE) or Schott’s glass-ceramic, Zerodur, which have essentially a zero coefficient of thermal expansion at room temperature (less than 1 in $30 \times 10^{-6}{^{\circ}\text{K}^{-1}}$). Today the four 8.1-m VLT telescopes at the European Southern Observatory (ESO), the 8.3-m Subaru telescope and the two 8-m Gemini telescopes all use meniscus mirrors weighing ~20 tons. These mirrors are supported against gravity and wind deflections on a matrix of between 120 and 264 active supports (see ref. 27 for a good review of mirror support techniques used in large telescopes).

Despite this diversity in mirror technologies, a common theme is the move to faster—that is, smaller—focal ratios. (The focal ratio, or $f$-number, of the primary mirror is the focal length of the mirror divided by its diameter.) Shortening the focal length of the primary mirror shortens the telescope tube length and reduces the size, and hence the cost, of the telescope domes (Fig. 3). For example, the domes enclosing the Keck 10-m telescopes and Gemini 8-m telescopes are comparable in size to the domes housing the Kitt Peak 4-m telescope and the Yerkes 40-inch refractor.

However, the difficulty of polishing and testing a large mirror so that it can produce good images is a strong function of its departure from a spherical section; the larger the departure, the greater the fabrication difficulties. Two-mirror (primary and secondary) telescopes typically use either parabolic or hyperbolic primary mirrors. For the Palomar f/3.3 mirror, the parabolic surface differed from that of a perfect sphere by only 32 $\mu$m, which was within the capabilities of polishers in the 1930s (ref. 9). But as the mirrors get larger, a similar departure from a sphere would require a longer focal length, greatly increasing the size and cost of the telescope. Modern polishers have developed a range of computer-controlled polishing and testing techniques for producing faster primary mirrors of very high precision. The University of Arizona is pushing this technique to produce large mirrors with focal ratios of f/1.2 or less. The deviation from a sphere for an f/1.2 8-m mirror is measured in millimetres (1.19 mm) rather than micrometres.

In addition, the active support systems used in the telescope can correct any large-scale errors, and polishers can now concentrate on local smoothness. The result is 8-m diameter mirrors that can perform near the diffraction limit at optical wavelengths. To get some sense of how smooth these mirrors will be on their support systems, if one were to enlarge the mirror to the size of the Atlantic Ocean, the largest wave would be no higher than a foot (30 cm).

With the faster telescopes has come another requirement for modern computer analysis and control techniques. For the 8-m f/1.8 optical systems to remain diffraction-limited at 2.2 $\mu$m, the
secondary mirror has to remain in alignment on the optical axis of the primary mirror to within \( \sim 50\ \mu\text{m} \). To remain in focus, the primary to secondary spacing has to be maintained to within \( \sim 2.5\ \mu\text{m} \) over a distance of 12 m (ref. 29). These alignment tolerances become even smaller as the focal ratio of the primary mirror is reduced. Using computer models and image measurements to control the position of the secondary mirror and the shape of the primary mirror precisely, gravitational deflections, thermal distortions and wind buffeting of the telescope can all be actively corrected, keeping an 8-m \( f/1.8 \) telescope diffraction limited as it tracks across the sky.

To give some indication of how well modern telescopes can 'break the cost curve', the Kitt Peak 4-m telescope cost $10M at its completion in 1973 (ref. 13). In 1998 dollars this is approximately equivalent to $26M. With an expected cost growth scaling law\(^{32,33} \) of \( D^{3.5-D^{2.8}} \) this would translate into an 8-m telescope costing between $147M and $181M. Keck and Gemini have spent approximately $88M on each of their 10-m and 8-m telescopes (these costs exclude instruments and ignore differences due to inflation), so these new technologies have resulted in significant cost savings.

**Improving image quality**

Improving the image quality of telescopes allows us to observe fainter point-like sources. Telescopes are used to observe faint astronomical objects, which are typically much fainter than the instantaneous sky background. For some infrared measurements, sources can be several orders of magnitude fainter than the background resulting from the thermal emission of the telescope. Because of this limit, the ultimate signal-to-noise ratio (for point sources) is related to the physical parameters of the telescope and instrument by the following equation:

\[
\frac{S}{N} \propto \frac{(\text{telescope area})^{1/2}}{\text{delivered image diameter}^{1/2}} \sqrt{\frac{\eta}{\epsilon_b}}
\]

where \( S \) is the signal from a faint astronomical object, \( N \) is the noise in the background in a given integration time, \( \eta \) is the total system throughput (including atmospheric, telescope and instrument transmissions as well as the detector quantum efficiency), and \( \epsilon_b \) is the effective background emission. It is the combination of a large aperture delivering small images that makes the new generation of telescopes so powerful. In addition, the HST, with its 0.1-arcsec or better image quality (at optical wavelengths), has revealed great richness of spatial detail in almost every object that it has observed, from star-forming regions to galaxies within a few billion years of the Big Bang. Historically, most of this structure has been blurred by our atmosphere and resulted in images with five to ten times poorer spatial resolution.

The image quality of a ground-based telescope also has two major atmospheric 'seeing' components. The first, realized through pioneering work at the MMT\(^{34} \), the Canada–France–Hawaii telescope (CFHT)\(^{35} \) and by the Subaru\(^{36} \) group, is that substantial temperature differentials between either the mirror surface or telescope structure and ambient air lead to thermally induced turbulent eddies, which can substantially degrade the 'local' seeing in the environment of the telescope. Modern enclosures are now actively flushed during observing to remove any warm air accumulating around the telescope. During the day the structures are cooled with air conditioning to keep the entire enclosure near the night-time ambient temperature. In addition, a variety of strategies are now used to keep the mirror surface near the ambient temperature\(^{37} \). The success of these new approaches in removing the effects of the enclosure, telescope and mirror seeing effects is shown in Fig. 4, from the remarkable 'First Light' results obtained by the ESO LT-1\(^{38} \).

The second component is the intrinsic atmospheric seeing due to turbulent cells that pass through the telescope's line-of-sight (typically at an altitude of \( \sim 10\ \text{km} \)). The emerging techniques of adaptive optics can substantially reduce the effects of atmospheric turbulence and offer the promise of another revolution.
in ground-based astronomy. By using a bright reference star, and more recently an artificial star created by a laser beam from the telescope, fast sensors can detect the phase distortion imprinted on the incoming wavefront by the turbulent atmosphere and apply a correction using a deformable or segmented mirror. The corrected wavefront can then be reflected into the detection apparatus mounted on the telescope (see ref. 39 for a good review of adaptive optics techniques). The potential scientific gains from adaptive optics have been demonstrated using the CFHT.

Through careful design and the use of adaptive optics, large telescopes can be built that will be essentially diffraction-limited at near-infrared wavelengths, combining the resolution of the HST with the collecting area of an 8–10-m telescope. This will permit detailed infrared imaging and spectroscopy of the objects observed with the HST. Equation (1) shows why the combination of large-aperture telescopes with adaptive optics promises such great gains in sensitivity. If an 8-m telescope can deliver 0.1-arcsec images, the S/N advantage compared with the typical 4-m telescope can be over 16; alternatively, this advantage can be expressed as making the same measurement in $(1/16)^2 = 1/256$ of the observing time.

Looking to the future

The enormous growth in ground-based optical and infrared capabilities becoming available to astronomers over the next decade will dramatically advance our understanding of the Universe. However, hints of the possible direction that future observations might take are already beginning to emerge. For example, recent infrared and submillimetre-wavelength observations suggest that the formation of stars and the creation of the first heavy elements in the earliest galaxies might be largely hidden from optical observations. Much closer to home, there has been an invigorated interest in searching for planets around other stars, and for the detailed investigation of extrasolar planetary systems. For the first time in the 400-year history of observational astronomy there is a growing realization, both in the scientific and public arenas, that we might be able to study “the evolution of the universe in order to relate causally the physical conditions during the Big Bang to the development of RNA and DNA.” What will this require from future telescopes?

To study the earliest formation of stars and the formation of the early precursors to the galaxies we see today will require very sensitive observations. The Hubble Deep Field represents the most sensitive probe of distant galaxies yet achieved. After over two years of work, the Keck 10-m telescope has managed to obtain spectroscopic redshifts of about 65 objects, only a small fraction of the 2,400 identified sources in the Hubble deep field. If we could increase the sensitivity by another factor of 10 or 20, about 50% of the remaining objects, many of which could be nescent galaxies, would become accessible to spectroscopy and detailed analysis.

As twenty-first-century astronomers attempt to exploit the recent discoveries of extrasolar planets, modelling indicates that even modern 8–10-m telescopes will struggle to obtain spectra of Jupiter-like objects at a distance of only 10 pc. To make progress in this area requires an effective aperture of at least 30–50 m. The challenge for the next generation of telescope builders who build the ELT will be once again to take on both the engineering and cost obstacles that such large telescopes will clearly require from future telescopes?

Equation (1) suggests that the only way forward in ground-based observations is to increase greatly the ratio of $(\text{telescope area})^{1/2}$ to $\eta e_b$ rapidly approaching its limit. The options are either a very large telescope delivering diffraction limited images, or an interferometric array of large telescopes. Although optical/infrared interferometers seem an attractive approach, it is hard to achieve diffraction-limited images of the required sensitivity with a sparse array. To achieve $S/N$ gains with factors of 10–20, the next large ground-based telescope needs to be an ‘Extremely Large Telescope’ (ELT) with an effective aperture of at least 30–50 m. The challenge for the next generation of telescope builders who build the ELT will be once again to take on both the engineering and cost obstacles that such large telescopes will clearly require.
large telescopes will present. Scaling the approach from today’s innovative 8-m and 10-m telescopes (using a cost growth ratio proportional to $D^2$), a single 30–50-m telescope could cost between $3.5$ billion and $8.0$ billion. Unless the costs of a ground-based ELT can be brought to within $1$ billion, it might never be built. Consequently, the astronomers and engineers who are contemplating these ELTs, both in Europe and in the USA$^{25-34}$, are trying to push today’s technological developments even further. Possibilities include exploiting new mirror materials such as silicon carbide or beryllium and perhaps making mirrors faster than $f/1$.

An alternative is to put a large telescope in space where the sky background levels ($e_b$), especially in the infrared, can be reduced by factors of between $10^2$ and $10^6$ (Fig. 5). For a space telescope of equivalent aperture and imaging performance to a ground-based telescope, this can produce $S/N$ gains of between $10$ and $1,000$. It is this dramatically increased sensitivity that is the principal science driver for the ‘Next Generation Space Telescope’ (NGST), a 6–9-m telescope that NASA is contemplating launching into space$^{35}$.

If NASA can bring the NGST to within a total mission cost (telescope, launch and subsequent operations) of $1$ billion, it will compete favourably with any future ground-based facility with at least equivalent scientific capabilities$^{31}$. The approach that NASA is hoping to take might deliver a space-based 6–8-m telescope for about one-fifth of the cost of the 2.4-m HST$^{36}$; this approach might also yield the materials and technologies for the next generation of ground-based telescopes. Perhaps early in the twenty-first century, space-based telescopes and ground-based ELTs will not look so very different, and apart from a large difference in aperture they will probably cost roughly the same.

Should ground-based telescope builders consider abandoning their terrestrial mountain-top sites and head for space? Perhaps not yet. Many of the scientific issues of the future, such as determining the mass distributions of galaxies, tracing the formation of elements, or the characterization of sub-stellar objects such as brown dwarfs or extrasolar planets, will demand observations with high spectral resolution of at least $\lambda/\delta\lambda \approx 10,000$ (ref. 57). For studies of this sort, collecting area will remain at a premium, and the effect of background emission (or lack of it) will be much reduced compared with detector performance. When realistic detector properties such as read noise and dark current are included in the comparisons, the complementary nature of ground-based and space-based telescopes begins to emerge$^{38}$ (Fig. 6).

**Figure 6** The performance of ground-based 20-m and 50-m telescopes compared with an 8-m next-generation space telescope. The relative $S/N$ gain of a 20-m telescope is shown in blue and the 50-m telescope in red, using the methods described in ref. 66. The $S/N$ gain for a ground-based telescope is plotted as a function of wavelength, assuming that NGST makes an observation of a point-like source in an integration time of 4,000 s achieving an $S/N$ of 10. We assume that the objects we need to observe are very small (less than 0.01 arcsec), such as a star-forming cloud or stellar cluster at high redshift, or a planet around another star. Two cases are shown, a comparison for imaging observations with $R(\lambda \lambda \lambda) = 5$ and the relative performance for a spectroscopic observation with $R(\lambda \lambda \lambda \lambda) = 10,000$. For imaging, NGST has a clear advantage over a 20-m telescope in the infrared. For spectroscopic observations, a diffraction-limited 20-m ground-based telescope has an $S/N$ of 2–3 over an 8-m NGST making the same observation. In any wavelength up to 3.8 $\mu$m. Once the wavelength of observation becomes greater than 10 $\mu$m, a 50-m ground-based telescope becomes comparable to an 8-m NGST.

### References

33. van Horners, S. NLRST memorandum no. 2 (NRAO, 1987).
52. Mountain, C. M. What is beyond the current generation of ground-based 8-m to 10-m class telescopes and the LT-IF. Proc. SPIE 2871, 597–606 (1996).

Acknowledgements. We thank Marianne Takamiya for producing Fig. 1, and Ruth Kneale for finishing the figures and for work on the text and the references. We benefited from discussions with Larry Barr, Earl Pearson, Larry Stepp and Jim Oschmann.

Correspondence and requests for materials should be addressed to M.M. (e-mail: mountain@tartarus.gemini.edu).