The WASP Project and the SuperWASP Cameras

D. L. Pollacco,[1] I. Skillen,[2] A. Collier Cameron,[3] D. J. Christian,[1] C. Hellier,[4] J. Irwin,[5] T. A. Lister,[3,4]
R. A. Street,[1] R. G. West,[6] D. Anderson,[4] W. I. Clarkson,[7] H. Deeg,[8] B. Enoch,[7] A. Evans,[4] A. Fitzsimmons,[1]
C. A. Haswell,[7] S. Hodgkin,[5] K. Horne,[3] S. R. Kane,[3] F. P. Keenan,[7] P. F. L. Maxted,[4] A. J. Norton,[7] J. Osborne,[6]
N. R. Parley,[7] R. S. I. Ryans,[1] B. Smalley,[4] P. J. Wheatley,[6,9] and D. M. Wilson[4]

ABSTRACT. The SuperWASP cameras are wide-field imaging systems at the Observatorio del Roque de los Muchachos on the island of La Palma in the Canary Islands, and at the Sutherland Station of the South African Astronomical Observatory. Each instrument has a field of view of some 482 deg^2 with an angular scale of 13″.7 pixel^{-1}, and is capable of delivering photometry with accuracy better than 1% for objects having $V \sim 7.0$–11.5. Lower quality data for objects brighter than $V \sim 15.0$ are stored in the project archive. The systems, while designed to monitor fields with high cadence, are capable of surveying the entire visible sky every 40 minutes. Depending on the observational strategy, the data rate can be up to 100 Gbytes per night. We have produced a robust, largely automatic reduction pipeline and advanced archive, which are used to serve the data products to the consortium members. The main science aim of these systems is to search for bright transiting exoplanet systems suitable for spectroscopic follow-up observations. The first 6 month season of SuperWASP-North observations produced light curves of ∼6.7 million objects with 12.9 billion data points.

Online material: color figures

1. INTRODUCTION

In recent years, interest has grown in relatively small aperture and inexpensive wide-field imaging systems essentially composed of large CCDs mounted directly to high-quality wide-angle camera optics. The first prominent success of such an instrument was the spectacular discovery of the neutral sodium tail of comet Hale-Bopp (Cremonese et al. 1997), with a temporary purpose-built camera system (CoCAM). Since then, similar cameras have resulted in imaging of a gamma-ray burst during the burst period (Akerlof et al. 1999), and the first detection of the transits of an extrasolar planet in front of its parent star, HD 209458 (Charbonneau et al. 2000). Such in-

struments are ideal for projects requiring photometry of bright but rare objects (see Pinfield et al. 2005).

1.1. Extrasolar Planetary Transits

The first extrasolar (exo)planets were discovered in 1992 by pulsar timing experiments (Wolszczan & Frail 1992). While this technique is sensitive to the detection of terrestrial-sized planets, its limited applicability has restricted its use to just a few objects. Mayor & Queloz (1995) discovered the first exoplanet, 51 Peg, from optical radial velocity studies, and since that time the field has been dominated by this technique. One of the surprises of these surveys (e.g., Marcy et al. 2005) is the existence of a significant population of solar-type stars accompanied by relatively rapidly orbiting Jupiter-sized planets. However, spectral measurements alone cannot determine unambiguously the true mass or radius of the planet, as the orbital inclination is unknown. These surveys have discovered Jupiter-sized objects in orbits out to 3 AU around 6% of the nearby Sun-like stars surveyed. Of these, some 30% are "hot Jupiters" situated in ∼4 day, 0.05 AU orbits, in which the equilibrium temperature is ∼1500 K. About 10% of hot Jupiters in randomly inclined orbits will transit their host star. Therefore, in random Galactic fields, roughly 1 in every 1000 solar-type stars should exhibit transits lasting roughly 2 hr with a period of a few days.

Given that the radius of Jupiter is $\simeq 0.1\ R_\odot$, these transits should result in a dimming of the parent star by $\simeq 0.01$ mag. The first transits were discovered in late 1999 (Charbonneau et al. 2000). The $V = 7.7$ star HD 209458 is dimmed by

---

[1] Department of Physics and Astronomy, Queen's University of Belfast, University Road, Belfast BT7 1NN, UK.

[2] Isaac Newton Group of Telescopes, Apartado de Correos 321, E-38700 Santa Cruz de La Palma, Tenerife, Spain.

[3] School of Physics and Astronomy, University of St. Andrews, North Haugh, St. Andrews KY16 9SS, UK.

[4] Astrophysics Group, Keele University, Keele, Staffordshire ST5 5BG, UK.

[5] The Wide Field Survey Unit, Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK.

[6] Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK.

[7] Department of Physics and Astronomy, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

[8] Instituto de Astrofísica de Canarias, Vía Láctea, s/n, E-38200 La Laguna, Tenerife, Spain.

[9] Department of Physics, University of Warwick, Coventry CV4 7AL, UK.

0.016 mag for 2 hr every 3.5 days, hence both proving the existence of the planet detected by a radial velocity search, and resulting in a precise measurement of its radius, mass, and bulk density. Because of the multiplexing advantage of imaging, this technique promises to be the fastest way of detecting exoplanets and could, over the next few years, dictate which candidates are followed up by radial velocity studies (rather than vice versa as at present).

Initially, groups trying to find transits of exoplanets reported disappointing results. For example, the Vulcan project (Jenkins et al. 2002) searched some 6000 stars, finding only seven transit-like variables. Follow-up observations of these showed them all to be stellar in origin. More recent surveys have had more success, with published transit detections reported by the Optical Gravitational Lensing Experiment (OGLE) survey (Udalski et al. 2002a, 2002b), TrES-1 (Alonso et al. 2004), HD 189733 (Bouchy et al. 2005), and, most recently, XO-1 (McCullough et al. 2006). Part of the reason for the apparent lack of transits stems from the difficulty in obtaining photometry of a sufficient number of solar- and late-type main-sequence stars. To increase the number of stars sampled, wide-field surveys have often concentrated on low Galactic latitude fields. However, while the number of observable stars is undoubtedly increased, we suspect that the stellar population in such surveys is dominated by more distant K giants. Brown (2003) showed that the number of bona fide exoplanet transits (as opposed to stellar impostors) is consistent with expected numbers of binary and multiple stellar and exoplanet systems.

Horne (2003) listed some 23 photometric transit projects either in operation or under construction at that time. While many of these are pencil-beam surveys from traditional telescopes, a number of them, no doubt encouraged by the relative cheapness of the equipment, are employing novel wide-field cameras.

1.2. The WASP Consortium

The Wide Angle Search for Planets (WASP) consortium was established in 2000 by a group of primarily UK-based astronomers with common scientific interests. In order to reduce the development cycle time and be on-sky rapidly, we use commercially available hardware and hence limit development work as much as possible. Our ethos is therefore quite distinct from other apparently similar projects, e.g., the Hungarian Automated Telescope (HAT) project (Bakos et al. 2004).

The WASP consortium's first venture was the production of the WASP0 camera. Our experience with these types of systems stems from the CoCAM series of cameras at the Isaac Newton Group of Telescopes from 1996 to 1998, one of which was responsible for the discovery of the so-called sodium tail in comet Hale-Bopp 1995 (Cremonese et al. 1997). WASP0 is composed completely of commercial parts, and uses a Nikon 200 mm, f2.8 telephoto lens coupled to an Apogee AP10 CCD detector. WASP0 was used in 2000 and 2001 in La Palma (Spain) and Kryoneri (Greece), respectively, and has been

shown to easily detect the extrasolar transit of HD 209458b, among other variables (Kane et al. 2004, 2005a, 2005b).

On the strength of the WASP0 success, the consortium was able to raise sufficient funding for a more ambitious project: the multidetector SuperWASP cameras. The limited development required is reflected in the aggressive project timescale: for the La Palma instrument, funding was approved in 2002 March and first light was achieved in 2003 November, while for the South African Astronomical Observatory (SAAO), instrument funding was secured in 2004 April and first light was achieved in 2005 December.

In this paper, we describe the SuperWASP facilities at the Observatorio del Roque de los Muchachos on La Palma (SuperWASP-N) and the recently commissioned system at the Sutherland Station of the SAAO (SuperWASP-S). Along with the hardware and data acquisition system, we outline the SuperWASP reduction pipeline and archiving system for the data products.

2. THE HARDWARE SYSTEM

In outline, each SuperWASP instrument consists of an equatorial mount on which up to eight wide-field cameras can be deployed. Each is housed within a two-room enclosure that incorporates a roll-off roof design. Figures 1 and 2 show the enclosure at SuperWASP-N and the detector system at SuperWASP-S. For both facilities, all observatory functions are under computer control, including data taking.

2.1. The Robotic Mount

Both systems employ a traditional equatorial fork mount constructed by Optical Mechanics, Inc. (Iowa; formerly Torus Engineering). The mount is manufactured within their Nighthawk telescope range. When properly configured, the mounts give a pointing accuracy of $30''$ rms over the whole sky, and a tracking accuracy of better than $0''.01$ s^{-1}. The mounts are easily capable of slewing at a rate of $10°$ s^{-1}. On site, the mounts are attached to a concrete pier.

For our project, we do not have a conventional optical tube assembly, but instead we employ a cradle structure to hold the individual cameras. The cradle allows limited camera movement in three dimensions for balance and alignment purposes.

2.2. The Enclosure

The rapid slewing of the mount and large field of view make a traditional dome impractical and inefficient; hence, we have a custom roll-off roof structure. In most deployments of this design, the roof is moved onto rails overhanging the building; however, in our design, the space under the rails is used as a fully temperature-controlled control and computer room, with the rails integrated into the roof. The building itself was constructed by Glendall-Rainford Products (Cornwall, UK) and is manufactured in laminated fiberglass strengthened with wood, making the structure extremely rigid. The likely absence of a crane during the building erection means that the size of the

Fig. 1.—SuperWASP-N enclosure at the Roque de los Muchachos observatory on la Palma. The weather station is in the background. [*See the electronic edition of the PASP for a color version of this figure*.]

roof panels was optimized to be liftable by three people. The moving roof is controlled by a hydraulically operated ram and associated electrics. In the case of SuperWASP-N, to fully retract the roof takes ∼19 s, and ∼54 s to fully close. The modular design of the building meant that the enclosure could be prefabricated by the manufacturer and then re-erected on site.

With the enclosure roof fully retracted, objects with declinations $-20° < \delta < 55°$ are visible for the entire period when their altitude is >30°. For $\delta > 55°$, the movable roof may obstruct visibility at some hour angles. For the SuperWASP-S instrument, we designed a longer enclosure to give better southern access.

2.3. The CCDs

The SuperWASP CCD cameras were manufactured by Andor Technology (Belfast, UK) and marketed under the product code DW436. The CCDs themselves are manufactured by e2v and consist of 2048 × 2048 pixels, each 13.5 μm in size. These devices are back-illuminated, with a peak quantum efficiency of >90%. Andor uses a five-stage thermoelectric cooler to reach an operating temperature of −75°C. At this temperature the dark current is ∼11 e pixel^{-1} hr^{-1}, which is comparable to cryogenically cooled devices. As our exposure times are

only 30 s, we do not require this level of performance, and hence we cool the devices to −50°C, at which point the dark current is ∼72 e pixel^{-1} hr^{-1}.

Andor Technology also provides a 32 bit PCI controller card that is used to control all CCD functions. These cards (one per detector) allow the devices to be read out at megapixel rates so that even after all overheads (e.g., header collection, disk write), a new image can be initiated within ∼5 s of the commencement of readout of the previous image. Even at this speed, the 16 bit images have good noise characteristics (gain of ∼2, readout noise of ∼8–10 electrons, and linearity better than 1% for the whole dynamic range). To simplify operations, we have not tuned detectors to controllers in any way.

The original design of SuperWASP-N conceived of using a renovated existing enclosure, with instrument control occurring from a nearby building. Hence, our shielded data cables are 15 m in length. Exhaustive testing showed that at this length, data collection was reliable and mains pickup rarely seen. The detector power supplies are stored next to the mount.

2.4. The Telephoto Lenses

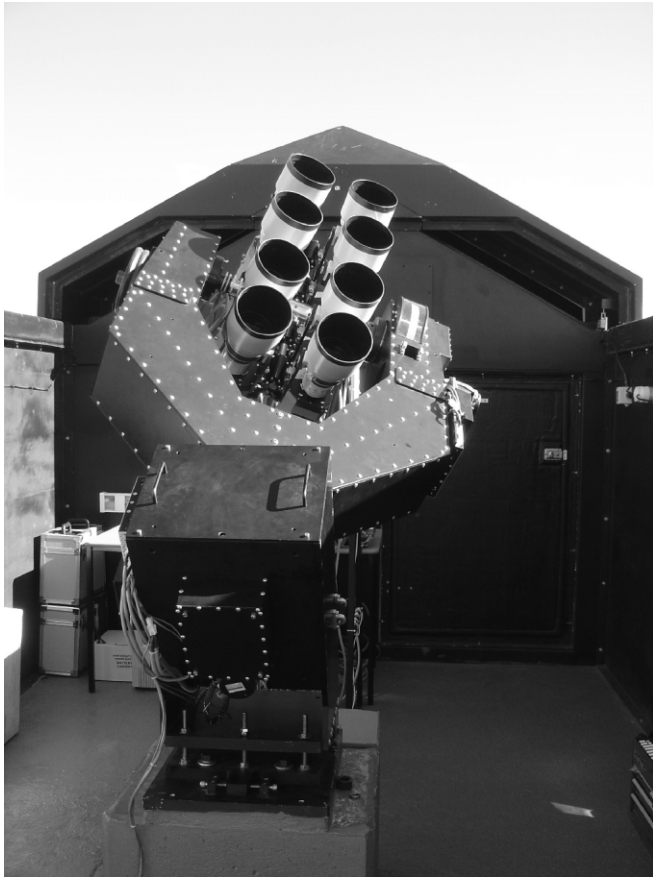In common with other similar projects, the SuperWASP cameras use Canon 200 mm, f/1.8 telephoto lenses. These lenses

Fig. 2.—SuperWASP-S instrument, with eight cameras mounted. The field of view of the instrument is ~482 deg². [*See the electronic edition of the PASP for a color version of this figure.*]

are among the fastest commercially available and have excellent apochromatic qualities. Funding constraints dictated that we initially purchase five lenses from a local supplier before this format became obsolete. We subsequently used eBay.com to track down the remaining units needed for both instruments. With the above detectors, they give a field size of ~64 deg² and an angular scale of 13″.7 pixel⁻¹. In the first year of operations for SuperWASP-N, our observations were unfiltered (white light) with the spectral transmission effectively defined by the optics, detectors, and atmosphere. Subsequently, we have deployed broadband filters at both facilities, which define a passband from 400 to 700 nm (see Fig. 3).

2.5. The Data Acquisition Computing Cluster

The easiest method to accommodate the high data rate from eight cameras is via a distributed data acquisition cluster, with each detector controlled by a dedicated DAS (data acquisition system) PC with local storage disks. Data taking itself is initiated by a central machine called the TCS (telescope control system), which also controls more general observatory func-

tions, such as pointing the mount and roof control. The TCS machine also has serial interfaces to a time service (supplied from a GPS receiver) and weather station. The DAS machines synchronize time with the TCS through the Network Time Protocol daemon. Overall, the relative time on the cluster is accurate to <0.1 s, while the GPS system ensures that the absolute time is accurate to better than 1 s. As the operation of the TCS is vital to the running of the instrument, a heartbeat system continually monitors the machine, with any break in communication initiating a close-down of the enclosure.

During the night, data are stored locally on each DAS machine. At the end of observing, the data are compressed and moved to a RAID system ready to be copied to tape (LTO2) for transportation back to the UK (recently, SuperWASP-N has begun sending data back via the Internet).

The weather station is provided by Vaisala (background in Fig. 1) and has sensors for internal/external temperature and humidity, wind direction and strength, precipitation, and pressure. A cloud sensor (IR-activated) is also used.

3. DATA ACQUISITION SOFTWARE

High-level software control of the entire SuperWASP system (robotic mount, CCD cameras, and roll-off roof) is provided by a modified version of the commercial Linux software Talon (now open-source), produced by Optical Mechanics, Inc. (OMI), for use with the Torus mount. Extensions to the software include support for multiple CCD cameras (developed by OMI) and some in-house modifications to add a command-line interface to supplement the standard graphical interface.

Talon supports two modes of operation: one for manual control with an observer present, using the graphical interface (or the new command-line interface); and the other for automatic observing, in which a dynamic scheduler takes control of the telescope and performs observations from a predefined queue.

In the first season of operation, the observer was responsible for taking bias and dark frames, opening and closing the dome, and taking twilight flat fields, automated using a driver script for the command-line interface. Science observations were taken using the `telrun` daemon within Talon, driven from a Perl script. A new dynamic scheduler, `waspsched`, developed in-house by the consortium and using the command-line interface, has recently been commissioned. This has increased observational efficiency by allowing continuous operation of the equipment (previously, a number of delays were required during observing in order to synchronize the interactions with the existing Talon scheduler, allowing about one 30 s integration per minute despite the 10° s⁻¹ slew speed of the mount). The dynamic scheduler also allows us to intersperse all-sky survey fields with the exoplanet fields. Support for alternative observing modes may also be added, in particular the ability to interrupt the scheduled observing to follow up transient events (e.g., gamma-ray bursts) without user interaction.

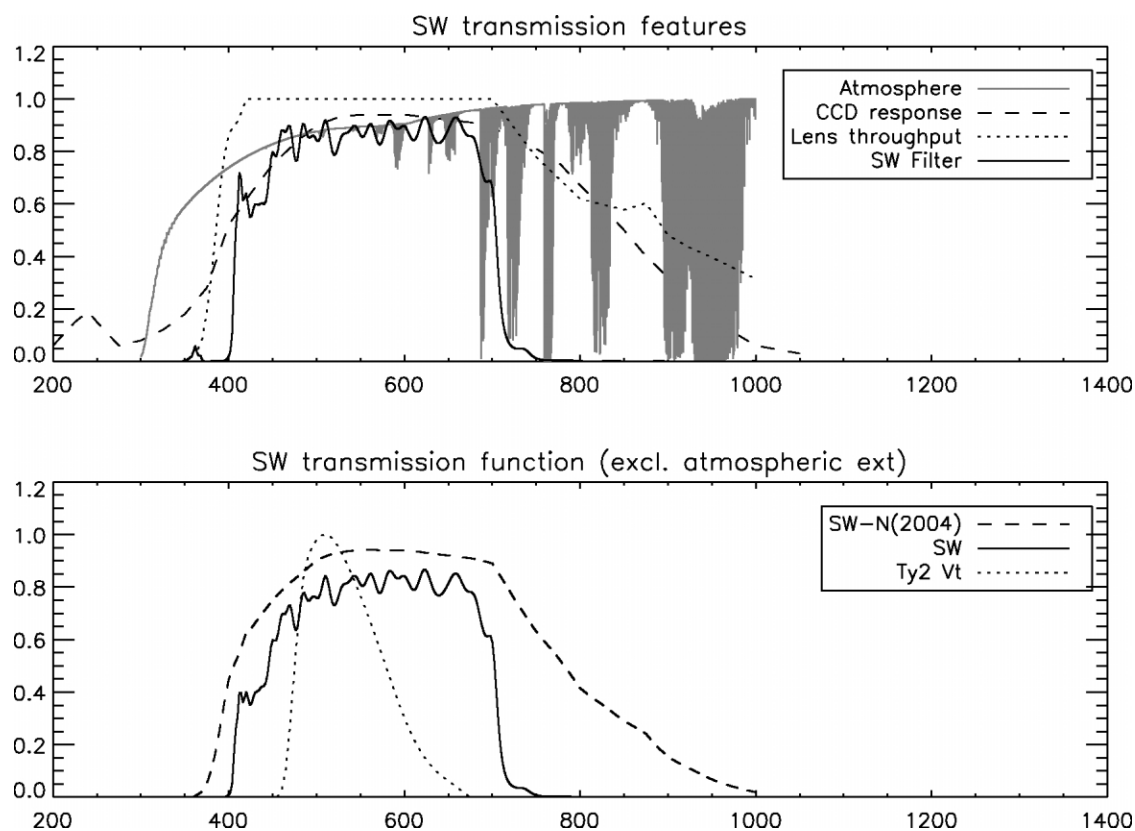Data from the site weather station are fed into the software.

Fig. 3.—Passband of the new SuperWASP filter (*top*) plotted alongside the atmospheric transmission, CCD response, and lens transmission. The bottom panel shows the original unfiltered system alongside those of the new filter and the Tycho-2 *V* filter.

Talon has a number of configurable conditions under which a weather alert is issued, for example, high wind or excessive humidity. On triggering a weather alert, the telescope is immediately slewed to a predefined park position (to avoid collisions with the roof), and the roof is then closed. After the alert condition finishes, the software waits for a short period (typically 20 minutes) and then opens the roof and continues observing if it is still dark. Other alerts are generated by cloud and lightning sensors. In the event of failure to operate the roll-off roof, an alert condition is reached, sending a radio signal to a receiver in a neighboring operator-attended telescope dome. As Andor Technology provides a Linux-based software development kit, all aspects of the CCD control and data collection are integrated within Talon.

4. THE REDUCTION PIPELINE

The SuperWASP data analysis pipeline employs the same general techniques described by Kane et al. (2004) for the prototype WASP0 project. We use the USNO-B1.0 catalog (Monet et al. 2003) as the photometric input catalog. We carry out aperture photometry at the positions of all stars in the catalog brighter than a given limiting magnitude. This has two important advantages for subsequent data retrieval and analysis:

all photometric measurements are associated with known objects from the outset, and the aperture for every object is always centered at a precisely determined and consistent position on the CCD.

4.1. Calibration Frames

Bias frames, thermal dark-current exposures, and twilight-sky flat-field exposures are secured at dusk and dawn on each night of observation. The pipeline carries out a number of statistical validity tests on each type of calibration frame, rejecting suspect frames before constructing master bias, dark, and flat-field frames.

Master bias frames and thermal dark-current frames are computed by taking iteratively sigma-clipped means of the 10–20 frames of each type taken on each night. The master bias frame is subtracted from all thermal darks, flat-field frames, and science frames. Temporal drifts in the DC bias level are removed using the σ-clipped mean counts in the overscan region. The overscan strip is then trimmed off the bias-subtracted frames. The thermal dark frame is scaled according to the exposure time and subtracted from all flat-field and science frames.

The twilight-sky flats are exposed automatically in a sequence of 15 preprogrammed exposures ranging in duration

from 1 to 30 s. They are timed so as to be uniformly exposed to a maximum of about 28,000 ADU in the frame center. The mount is driven to slightly different positions on the sky between exposures in order to facilitate removal of stellar images when the images are combined. The flat fields show a circularly symmetric vignetting pattern caused by a sequence of baffles of similar size within each lens. Gradients in the sky brightness across each flat-field image are removed by rotating each image through 180° about the center of the vignetting pattern, subtracting the rotated image, and performing a planar least-squares fit to the residuals. The gradient is then divided out from each flat-field exposure.

The sky brightness distribution on the short-exposure flat fields is slightly distorted by the finite travel time of the CCD camera shutters, which are of the five-leaved iris type. At each pixel position, we must determine both the correction $\delta t(x, y)$ to the exposure time and the normalized flat-field value $N_\infty(x, y)$ that would be obtained for an infinite exposure time. The normalized counts $N(x, y)$ in an image with exposure time $t_{\exp}$ are modified by the shutter travel correction $\delta t(x, y)$:

$$N(x, y) = N_\infty(x, y) \left[1 + \frac{\delta t(x, y)}{t_{\exp}} \right].$$

At each pixel position (x, y), we determine the combined flat-field and vignetting map $N_\infty(x, y)$ and the shutter time correction map $\delta t(x, y)$ via an inverse-variance–weighted linear least-squares fit to N versus $1/t_{\exp}$. An iterative rejection loop eliminates outliers, which are usually caused by a stellar image or cosmic ray falling on the pixel concerned in one or more of the frames. We then smooth the map of δt using a two-dimensional spline fit and use this to recover an improved map of

$$N_\infty(x, y) = \frac{N(x, y)}{1 + \delta t(x, y)/t_{\exp}}$$

for each individual exposure. We then average these corrected exposures, again using iterative σ-clipping to eliminate stellar images in individual frames. The shutter correction is applied to the science frames, as well as to the flat fields. In general, for a 30 s integration the shutter corrections are between $+0.02$ and -0.01, with $\sigma = 0.006$. While flat fields are obtained on a twice-daily basis (weather-dependent), we use an exponential weighting scheme to produce a daily master flat field (Collier Cameron et al. 2006).

4.2. Astrometry

In order to derive adequate astrometry, we must establish a precise astrometric solution for every CCD image. The celestial coordinates of the image center can be established to a precision of a few minutes of arc from the mount coordinates recorded in the data headers, and from the known offsets of the individual cameras from the mount position. Subsets of the Tycho-2 (Høg et al. 2000) and USNO-B1.0 catalogs are made and retained

for every preprogrammed pointing of the mount and for every camera. The subcatalogs cover a slightly larger region of sky than the images with which they are associated, to allow for pointing uncertainty.

We use the Starlink EXTRACTOR package, which is derived from SExtractor (Bertin & Arnouts 1996), to create a catalog of the 10^4 or so stellar images detected at 4 σ or greater significance on each frame. We project the Tycho-2 subcatalog on the plane tangent to the celestial sphere at the nominal coordinates of the field center. We attempt to recognize star patterns formed by the brightest 100 stars in both catalogs and establish a preliminary plate solution consisting of a translation, rotation, and scaling. Further stars are then cross-identified, and the solution is refined by solving for the barrel distortion coefficient and the location of the optical axis on both the sky and the CCD. The rms scatter of the EXTRACTOR positions, relative to the computed image positions of Tycho-2 stars on the CCD, is always close to 0.2 pixel.

Once the plate solution is established, the pipeline software creates a photometric input catalog from the list of all USNO-B1.0 objects brighter than magnitude $R = 15$ (in the USNO system), whose positions fall within the boundaries of the CCD image. Positional and rough magnitude matching yields USNO-B1.0 identifications for all but a few dozen of the objects found by EXTRACTOR. These "orphan" objects, some of which are likely to be transient variables or minor solar system bodies, are added to the photometric input catalog at their observed pixel locations. In addition, the positions of the bright planets are computed and, if they fall within the image area, added to the input catalog.

4.3. Aperture Photometry

The photometric input catalog gives the precise CCD (x, y) coordinates of up to 210,500 objects, together with their USNO-B1.0 magnitude estimates. We create an exclusion mask for fitting the sky background, by flagging all pixels within a magnitude-dependent radius about every object in the input catalog. A quadratic surface is then fitted to all remaining pixels in an iterative procedure. On the second iteration, the fit is refined by clipping outliers to remove cosmic rays and faint stars, and adding their locations to the exclusion mask.

Gradients and curvature in the sky-background illumination are removed by subtracting the quadratic sky fit from the image. Images are rejected if more than 50% of the pixels are clipped or have too high of a χ^2 value—usually indications of significant cloud cover affecting the observations. Aperture photometry is then performed in three circular apertures of radius 2.5, 3.5, and 4.5 pixels (these apertures were selected by inspection of images of known blended and unblended objects, at this spatial resolution). Since the aperture is centered on the actual star position, the weights assigned to pixels lying partially outside the aperture are computed using a Fermi-Dirac–like function. This is tuned to drop smoothly from 1.0 half a pixel inside the aperture boundary, to 0.0 half a pixel outside it. The weights

of these edge pixels are renormalized to ensure that the effective area of the aperture is πr^2, where r is the aperture radius in pixels.

The sky background is computed in an annulus with an inner radius 13 pixels and outer radius 17 pixels, so that the sky annulus has 10 times the area of the 3.5 pixel aperture. Pixels flagged in the exclusion mask as being occupied by stellar images or cosmic rays are excluded from the sky-background calculation.

For a given star, the ratios of the fluxes in the various apertures contain information on the point-spread function. We define two wing-to-core flux ratios: $r_1 = (f_3 - f_1)/f_1$ and $r_2 = (f_3 - f_2)/f_2$, where $f_1, f_2,$ and f_3 are the flux measurements in each of the three apertures defined above, respectively. Figure 4 is a plot of r_2 against r_1, which reveals that for unblended stellar images, the two wing-to-core ratios are related by a constant scaling factor. Stars whose wing-to-core flux ratios lie close to the main locus for unblended stars are flagged as such, while outliers are flagged as likely blends. The photometric measurements for blended images are very sensitive to small errors in the astrometric fit. Their light curves therefore tend to be substantially noisier than those of unblended objects.

4.4. Post-Pipeline Calibration—PPWASP

When the photometric input catalog file is created, each object is labeled with its air mass and catalog magnitudes. The heliocentric time is calculated on a per object basis (the heliocentric time varies significantly over the field of view of the instrument). The photometry modules add information on the sky background level, aperture radius, raw instrumental aperture fluxes and their associated variances, and blending information, outputting the results in FITS binary tables. These tables are then read into the post-pipeline calibration module PPWASP for reduction from raw instrumental to calibrated standard magnitudes.

PPWASP calibrates and removes four main trends in the photometry: the effects of primary and secondary extinction, the instrumental color response, and the system zero point. The nightly mean primary and secondary extinction coefficients are determined from an iterative least-squares fit to the variation of raw magnitude with air mass through the night for a sample of stars with colors defined in the Tycho-2 catalog. Bayesian priors are used to stabilize the fits on nights for which the air mass range is insufficient to yield a reliable least-squares fit. Stars showing excessive variance (determined by a maximum-likelihood procedure) about the mean trends are down-weighted by including the excess variance caused by intrinsic variability in the inverse-variance weights at each iteration. A time-dependent adjustment to the primary extinction coefficient is then computed for each frame by determining the mean deviation of the ensemble of calibration stars in each frame from the nightly extinction trend.

Once the instrumental magnitudes have been corrected to a standard air mass near the middle of the observed range, a
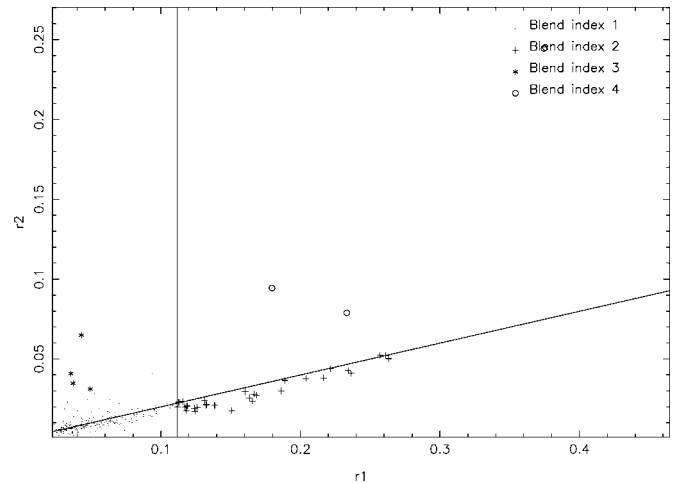

FIG. 4.—Graph showing that wing-to-core ratios defined from different-sized apertures are related by a constant scaling factor. This plot does not include all stars in a field. Instead, only those that are brighter than $V = 15.0$ and have reasonable values for the r_1 and r_2 indices (<0.5 for both) are included. The pipeline actually flags nine different types of "blended" objects, ranging from type 0 (unblended) to 9 (saturated star or a bad pixel). The types shown in this figure correspond to extremely red objects (1) or those indices that suggest nearby companions or distorted images (2–4). The vertical line shows when the effect of nearby companion stars is detectable in the r_1 index and was derived from real SuperWASP-N data in conjunction with Two Micron All Sky Survey (2MASS) and Digitized Sky Survey (DSS) data.

linear equation for the instrumental color response and zero point of each camera is used to transform the instrumental magnitudes to a system defined by the Tycho-2 V_t bandpass. Approximately 100 bright, nonvariable stars are adopted as secondary standards within each field. Their standardized magnitudes as determined over a few photometric nights are subsequently used to define the "WASP V" magnitude system for the field concerned. This allows the final zero-point correction to be determined to a precision of 1 or 2 thousandths of a magnitude for every frame on every night. This eliminates biases that would otherwise arise if one or more of the standards were rendered unusable by saturation or a cosmic-ray hit.

The calibrations are calculated and applied separately for the measurements from each of the three apertures applied by the aperture photometry program. Once postprocessing is complete, the calibrated fluxes are added to the binary FITS tables ready to be ingested into the archive.

5. THE ARCHIVE

The SuperWASP archive plays a central role in the efforts of the consortium, as it is the only long-term repository for the full set of the WASP photometric data. The archive interfaces have been specifically designed to facilitate the distributed, collaborative mode of working that typifies the efforts of the consortium.

Conceptually, the archive comprises three major classes of data: the bulk processed photometry, the raw images, and an

extensible catalog that can be augmented with the results of various analyses of the object light curves. Simple command-line tools are made available to the users of the archive to allow access to each of these three data classes.

5.1. Archive Server Hardware Configuration

The architecture chosen for the archive server is a storage cluster, i.e., a Beowulf-style cluster comprised of commodity compute nodes with gigabit Ethernet interconnect, wherein each cluster node is fitted with a large disk capacity (1.2 terabytes per node in the first incarnation). This model has a number of benefits, which include providing the compute power required for large-scale data-mining activities in a cost-effective manner. Perhaps as important, however, is that by embedding and distributing the storage throughout the cluster, all of the nodes act as both data servers and data consumers. There are multiple independent network paths between the servers and their potential consumers. This alleviates potential network bottlenecks associated with a more traditional model in which data is stored on and served by a small number of high-capacity storage nodes and consumed by a large number of client nodes. It also adds resiliency, in that the failure of a single node impairs archive performance for only as long as it takes to restore the relevant data to a different node. Finally, this model allows us to scale total capacity simply by adding more nodes, with each additional node contributing additional storage, compute power, and overall network bisection bandwidth.

5.2. Data Storage

Within the archive, the raw images and the processed photometric data are held in the form of FITS files on the archive server. These files are then indexed using a relational database management system (RDBMS) that allows the data relevant to a user query to be found in a fast and efficient manner. The RDBMS is also used to store and to support queries on the WASP catalog. MySQL was chosen ahead of PostgreSQL during the design phase based on performance measurements for typical queries anticipated at that time.

An early and important design decision made during the development of the archive was to separate the image and photometric data from the catalog and indexing information stored within the tables in the RDBMS. While in principle it would be possible to store images and light curves as BLOBs (binary large objects) within the RDBMS itself, we felt that for the data volumes to which the WASP archive is expected to scale (30 TB photometry and >100 TB images over 3 yr), such an implementation would be risky, with many potential performance and data integrity issues arising as the collection accumulated. The choice of FITS as a format to store the data was motivated by the wide acceptance of this format within astronomy, and the availability of stable and efficient third-party libraries and tools (for example, CFITSIO) to read and write these files.

The arrangement of the raw image data within the archive is relatively straightforward: the image files as processed by the pipeline are stored in their original form and are located within a directory hierarchy on a hierarchical storage management system. An entry into the RDBMS allows images matching a user query to be identified rapidly.

For the photometric data, the situation is rather more complex. The WASP pipeline at its heart operates on a single image at a time; the output of the pipeline is a collection of files, each of which contains the calibrated photometry for all of the stars detected in a given image. A typical delivery of data from the pipeline to the archive will be a collection of such files, representing the results of processing the data from a single night of observations. The end user of the archive will in the overwhelming majority of cases be interested in obtaining true light curves of objects of interest, i.e., a single file containing the photometry of a single object, collated from a large number of images taken over many weeks or months. For this reason, it is necessary for the archive to reorder the photometric data before they are presented to an archive user.

To create a light curve of an object from the pipeline products, it is necessary to be able to associate each photometric data point with a unique object; by matching object identifiers from image to image, it is then straightforward to build up a complete light curve. This process is greatly aided by the catalog-driven nature of the WASP pipeline: each data point can be uniquely and unambiguously assigned an identifier based on its place in the source catalog (USNO-B1.0). This obviates the need for complex and potentially error-prone cross-identification between photometric points based on positional coincidence, for example.

The collation of photometric data points into complete light curves takes place in two stages within the archive. The first stage occurs during the ingest phase and involves the reordering of the data from a given night (comprising several thousand files) into a smaller number of new files in which the photometric points for each object are stored contiguously within the rows of a FITS binary table. These files are also indexed internally by object name, which allows the table rows containing the photometry for a given object to be located rapidly. To ensure that the size of these bulk storage files is kept within manageable bounds, the archive divides the sky into "sky tiles," based on a plate carrée projection of equatorial coordinates. The size of these files is chosen to be $5° \times 5°$, as this will typically limit the size of the storage files to ~100 Mbytes to ~1 Gbytes. So the result of this reordering step in the ingest phase is a relatively small number of bulk storage files, each containing a single night of photometric data for all stars in a given sky tile (typically numbering thousands), presented in object order. These bulk files are then assigned a location on the storage cluster and indexed in the RDBMS.

For each sky tile, a meta-index is created by collating the internal indexes in the per night files associated with that sky tile. This meta-index lists, for each cataloged object, which files contain data pertaining to that object, and the numeric indices of the first and last rows in the table in each file holding

the data. The sky-tile meta-index is written to disk as a FITS binary table file; however, to improve access times, the meta-index files are also copied into a RAM disk on the archive server cluster nodes.

The second phase of light-curve collation occurs during the extraction process in response to a user query. A search of the RDBMS determines in which sky tile the object of interest falls. A scan of the meta-index for that sky tile then yields the names of the bulk photometry files containing photometry for the object of interest, plus the row offsets into the binary table. These files are then opened in turn, and the photometric data points are read and collated in memory into a FITS format light curve that, once complete, is presented to the user. This three-stage indexing process is highly efficient; tests on the system have demonstrated that a light curve comprising ~10,000 data points and spanning 130 days can be extracted in as little as 850 ms. This speed of access is of critical importance to many of the data-mining activities that will take place on the WASP archive, some of which will potentially require access to millions of light curves.

The pipeline processing of WASP images also yields detections of transient and/or uncataloged objects. These "orphan" data points are handled separately from the standard catalog-driven photometry and are stored in a table within the RDBMS for later analysis.

5.3. The Catalog

The second key element of the archive is the WASP catalog. At its core this provides very basic information about each object observed by SuperWASP, for example, equatorial coordinates and their relation to the entry in the USNO-B1.0 catalog that is used to drive the photometric pipeline. The catalog has been designed to be extensible, recognizing, however, that it is nearly impossible to define a priori the additional per object attributes that the catalog might be called upon to store as the understanding and exploitation of the WASP data proceed.

The catalog extensibility is implemented by storing newly derived attributes in separate tables in the RDBMS, alongside a WASP object identifier that acts as a key. By making use of table joins within the RDBMS at query time, the root catalog can appear to the end user to be extended on-the-fly to include these new attributes. The implication of this is that the results of any data-mining effort—for example, variability testing or period searches—can be loaded into a table in the RDBMS and immediately become visible to all users of the archive as additional per object attributes that can be used in queries in a transparent manner. In situations where different groups have different algorithms for attacking a problem (a notable example being planet-transit searches), these results can be ingested into different tables, and the choice of which set of results to use for any given query is left to the end user rather than the architect of the archive. This flexibility means that the archive will potentially become a very powerful tool in the exploitation

of the WASP data set, evolving to meet the needs of the end users.

Internal to the archive, queries of the RDBMS are constructed using the industry-standard Structured Query Language (SQL). To obviate the need for nonexpert users of the archive to formulate SQL queries directly (queries that can become very complex when involving table joins), we have developed a simpler language in which queries can be defined as a set of filters on object attributes using a much simpler syntax than SQL. This language (which we call WQL [WASP Query Language]) can be readily compiled into SQL queries by the archive software. Among other features, WQL allows users to perform cone searches by object coordinate (including on-the-fly name resolution using SIMBAD), to filter on object attributes, to select which attributes will be displayed in the output, and to sort the output on a selected attribute. The results from the query are returned either as formatted ASCII tables in a number of user-selectable flavors or as HTML. WQL is not directly comparable to other query languages, for example, the Astronomy Data Query Language (ADQL), which has an SQL-like syntax. While ADQL is undoubtedly more powerful and flexible than WQL, it suffers in that it is not very user friendly. The design goals of WQL were to create a query language that provides sufficient power and flexibility for most normal-use cases, but does so using a syntax that is amenable to use by astronomers without a computer science background. As a future development, the WQL compiler may be retargeted to generate ADQL rather than SQL queries.

5.4. The User Interface

The primary mechanism by which users can access the WASP archive is by means of three command-line tools that enable the querying of the catalog, the extraction of light curves, and the extraction of raw images. These tools communicate with the archive server using the HTTP protocol. HTTP was chosen because it potentially enables a wide variety of other clients (in particular, Web browsers) to be used in the future, and also because HTTP requests generally pass unmolested through network firewalls (which would not be guaranteed of a bespoke protocol on an arbitrary TCP/IP port).

The advantage of HTTP is that every WASP image and light curve has a unique and predictable URL, which can be accessed using any tool capable of retrieving a file via HTTP. Tools that are capable of opening FITS files directly from URLs (e.g., any application built on top of CFITSIO) can therefore request and open a light curve or image from the WASP archive across the Internet, with no more difficulty than if it was on local disk, simply by specifying a URL rather than a filename. In this respect, the WASP archive acts as a Web service.

6. OBSERVATORY PERFORMANCE

6.1. Hardware

The integrity of each observatory is checked remotely on a daily basis using interior and exterior network cameras, and
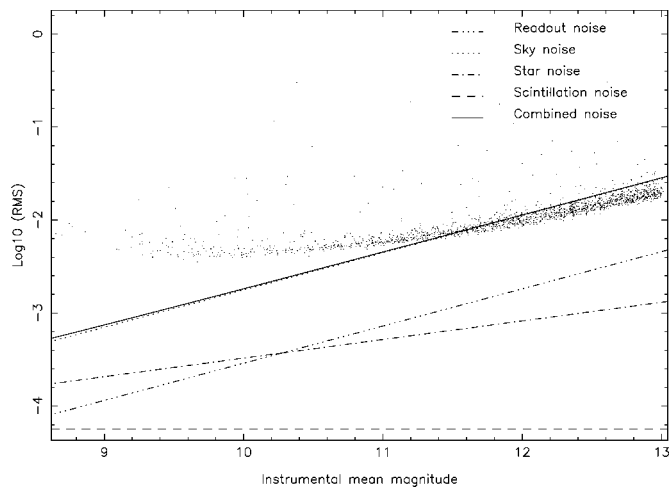
FIG. 5.—Diagram of rms error vs. magnitude for the 2004 season. Stars brighter than about $V \sim 11.5$ reach better than 1% accuracy. Also shown is a theoretical model that shows that the noise is dominated by the sky contribution (uncertainties in the sky contribution are apparent at fainter magnitudes).

manually on a weekly basis as part of a comprehensive preventative maintenance plan. Furthermore, during routine observing, the operation of the various subsystems can be monitored remotely using HTTP clients. No significant malfunctions have occurred to date. On La Palma, the enclosure has withstood three harsh winters, which included near 200 km hr^{-1} winds in a tropical storm, and 1.5 m of snow, without adverse effect. After the first season of operations, we found some indications of wear on the hour angle friction wheel, manifesting itself as vibration while slewing. After some investigation, we realized that during this first season some inexperienced observers had been moving the mount without disengaging the clamps. For subsequent seasons we have avoided this problem by redoing the pointing model and using the unused half of the friction wheel. Other problems arose in the declination axis: the original design of the mount expected a normal telescopic optical tube assembly, but the need to be able to insert the SuperWASP cradle meant that some movement was required. In the original design, the declination axis encoder was held at the end opposite to the drive. As this encoder is not a packaged product but a reader and glass, disk movement of the axis meant it was difficult to keep the encoder focused (or worse). For the second mount we therefore requested that OMI move the encoder to the drive end of the axis. For SuperWASP-N this problem only manifests itself when we are servicing the cradle unit. In normal operations, these mounts are made to work extremely hard. For SuperWASP-N the first season of operations regularly led to more than 600 movements per clear night. Similarly, the thermoelectrically cooled detectors have proved to be extremely reliable and operate stably at $-50°C$, even when the ambient nighttime summer temperature is in the mid 20°C range.

6.2. Pipeline

The pipeline has proved robust and runs in an unattended fashion. For example, an average night during the 2004 season produced around 550 frames per camera for 16 fields. After the postprocessing and data quality control, around 25 were rejected as outliers in the astrometric or point-spread profile fits (e.g., trailed images). From the 16 fields, around 710,000 starlike objects were accepted with good certainty and statistics, while an additional 220,000 were rejected, as they had a below-threshold number of data points (because either they were too faint or they saturated the detector). Each frame produced ~200 orphan objects (mostly cataloged objects in the USNO-B1.0 but with discrepant brightnesses). We find that on an average Pentium 4 (3 GHz processor, 1 Gbytes memory, SCSI disk system) class machine, one night of data from one camera takes about 11 hr to reduce. For the 6 months that SuperWASP-N was operational in 2004, the instrument produced light curves for ~6.7 million objects containing 12.9 billion data points.

6.3. Archive

A typical archive catalog query returns results within 1–10 s, depending on the complexity of the query and the number of matches. The RDBMS makes extensive use of in-memory caching of database tables, so repeating a query with small changes to the clauses often results in much improved performance for the second and subsequent queries.

Extraction of an arbitrary light curve typically takes ~1 s; however, the operating system in the storage cluster nodes takes the opportunity to cache some of the contents of the bulk photometry files in memory. Subsequent queries for light curves from nearby objects on the sky can often return much more quickly than the initial request. The clustered nature of the archive server allows dozens of independent requests to be served simultaneously without noticeable loss in performance.

Initially, the WASP archive will remain a private facility. However, if and when additional resources become available, we will make the data publicly available.

6.4. Photometric Performance

While our design goal was to reach 1% photometry for stars brighter than $V \sim 12.5$, in practice we reach this level of precision for stars brighter than $V \sim 11.5$. Stars brighter than $V \sim 9.4$ have a precision better than 0.004 mag. Equally important is the fact that the night-to-night variations are <0.002 mag. Objects with $V < 15.0$ are stored in the archive. Figure 5 shows the rms errors as a function of magnitude over the 2004 season and demonstrates that the complete system delivers good precision over an extended period. The faint end of our magnitude range is dictated by the sky brightness: in dark conditions, the integrated sky brightness in one of our pixels is ~16.5.
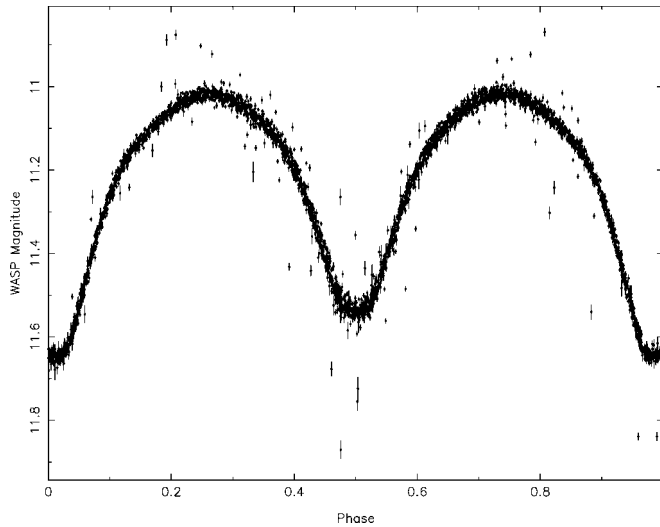
FIG. 6.—Example of an eclipsing binary star (GN Boo) from WASP data. This object has a period of 0.301601 days. The light curve is composed of 3873 data points obtained over the 2003 season in *all* weather conditions and has not undergone any postarchive processing.

6.5. Example Light Curves

For relatively large amplitude variations ($\Delta V > 0.05$), data from the archive are of sufficient quality for identification of variable stars, etc. Figure 6 is an example of a large-amplitude eclipsing system extracted from the archive. We stress that this figure is composed completely of raw extracted data and has not undergone any further processing. In this figure, there are 3873 data points taken over a 6 month period through all prevailing weather conditions. Close inspection shows that after ignoring outliers, the rms errors are around 0.015 mag at the quadratures, where it has a higher-than-normal noise for its magnitude.

Our primary design goal has always been to identify transit amplitudes with $\Delta V \sim 0.01$. In this regime, secondary effects can make the photometry less reliable. Rather than analytically trying to remove these, we use the `sysrem` algorithm discussed by Tamuz et al. (2005) to detrend the data. This algorithm removes time- and position-dependent trends from the light curves by computing the weighted average of magnitude residuals over all stars to calculate a time-dependent variation. We added an additional weighting to the magnitude uncertainties to down-weight known variable stars and poorer quality images. Note that the `sysrem` algorithm does not identify the causes of the trend, but removes it blindly, using all the stars available; for this reason, it should be applied with caution when finding genuine photometric variables is the desired outcome. After some experimentation, we remove four trends from the data. Figure 7 is an example of a transit candidate with and without the `sysrem` algorithm applied. This candidate is identified as 1SWASP J133339.36+494321.6 and has $V_t = 11.154$ mag. Out of transit, the rms scatter in the light
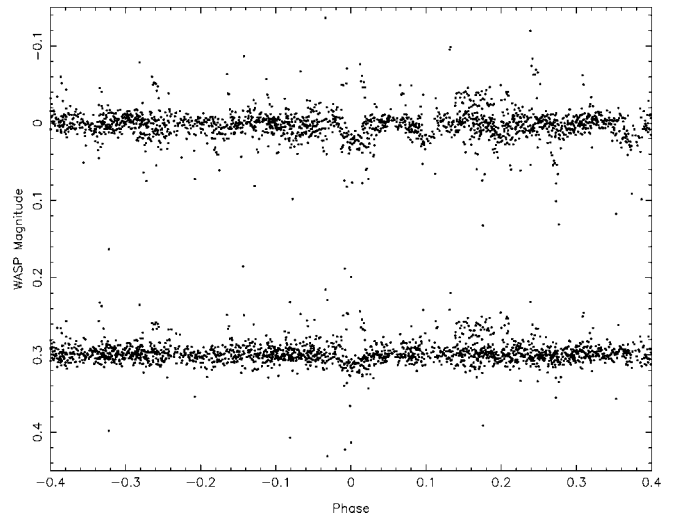

FIG. 7.—Typical transit candidate 1SWASP J133339.36+494321.6 found in SuperWASP data. The top light curve shows the archived data, while that below shows the same data after detrending through the `sysrem` algorithm. This candidate has a transit amplitude of ~1.5% and duration of 2.8 hr. The data are phased on a period of 3.63 days. Published data for this object available in the Tycho-2 and 2MASS catalogs suggest that the host has a spectral type of F7 (from the $V_{t\,\mathrm{(SW)}} - K_{\mathrm{2MASS}}$ color index; the 2MASS $J - H$ index gives a spectral type of F8). Consequently, the radius of the companion is $1.23R_J$. As these values are derived from photometry, they serve as a guide, but further observations of this target will be necessary to confirm its status.

curve is 0.0076 mag. Other exoplanet candidates from the first SuperWASP-N season of observations are now becoming available (e.g., Christian et al. 2006). For details of our `sysrem` implementation and transit candidate identification algorithm, see Collier Cameron et al. (2006).

7. SUMMARY AND FUTURE DEVELOPMENT

The original SuperWASP camera was constructed at Observatorio del Roque de los Muchachos on la Palma in 2003. Since then, we have also constructed a clone facility at SAAO, which has recently seen first light. Each system has a total field of view of 482 deg^2 and is capable of monitoring the sky down to ~15th magnitude every ~40 minutes. For the La Palma facility, in the first season of operation the instrument was used solely to monitor fields with high cadence (~10 minutes). In subsequent seasons, the development of the dynamic scheduler will allow all-sky monitoring interspersed with the high-cadence exoplanet fields (we expect complete sky coverage once per night).

The reduction pipeline is a continued development from the WASP0 instrument, now generalized and expanded to cater to both facilities. The rms plots show that the precision at the bright end of our magnitude range reaches 3–4 mmag over an extended period, and stars brighter than ~11.5 mag have errors less than 1%. An average night's data of 50 Gbytes produce some 6×10^6 records in the archive.

We have developed an archive system capable of storing tens to hundreds of terabytes of raw and processed data, and the results of high-level analyses of those data, but doing so in a manner that provides rapid access to individual light curves and images. The central catalog is designed to remain flexible and responsive to evolving user requirements despite the burgeoning data volume. The choice of HTTP as an interface protocol allows the end user a great deal of flexibility in the choice of access clients, and facilitates truly distributed, collaborative working within the consortium.

From the start of operations in 2006 at both facilities, we have begun an all-sky survey mode. Since the exoplanet program will continue, we have integrated these different survey modes (with only a small loss to the exoplanet survey efficiency), enabling once-a-night coverage of the entire visible sky. To do this effectively we will eventually deploy a reduction pipeline at the sites themselves. By compromising performance slightly we will be able to use this to grade the quality of the night, directly feeding back to the dynamic scheduler in order to modify the observing queue and hence matching the science to the prevailing weather conditions. In principle this could also provide a real-time alert system, and if sufficient software resources become available, we could use this to arrange follow-up observations on other nearby facilities.

The SuperWASP cameras are now surveying large parts of the sky, primarily in search of bright, transiting exoplanets. In the near future we will present large numbers of high-quality light curves suitable for follow-up and expect to at least double the numbers of known transiting exoplanets. Interests of the consortium members also cover other "secondary" science

areas, and publications in these areas are forthcoming. For nonmembers of the consortium, we also offer the opportunity to present ideas to the Project Steering Committee and hence achieve a Guest Observer status in this area.

REFERENCES

Akerlof, C., et al. 1999, Nature, 398, 400
Alonso, R., et al. 2004, ApJ, 613, L153
Bakos, G., Noyes, R. W., Kovács, G., Stanek, K. Z., Sasselov, D. D., & Domsa, I. 2004, PASP, 116, 266
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bouchy, F., et al. 2005, A&A, 444, L15
Brown, T. M. 2003, ApJ, 593, L125
Charbonneau, D., Brown, T., Latham, D. W., & Mayor, M. 2000, ApJ, 529, L45
Christian, D. J., et al. 2006, MNRAS, in press (astro-ph/0608142)
Collier Cameron, A., et al. 2006, MNRAS, in press (astro-ph/0609418)
Cremonese, G., et al. 1997, ApJ, 490, L199
Høg, E., et al. 2000, A&A, 355, L27
Horne, K. 2003, in ASP Conf. Ser. 294, Scientific Frontiers in Research on Extrasolar Planets, ed. D. Deming & S. Seager (San Francisco: ASP), 361
Jenkins, J. M., Caldwell, D. A., & Borucki, W. J. 2002, ApJ, 564, 495

Kane, S. R., Collier Cameron, A., Horne, K., James, D., Lister, T., Pollacco, D., Street, R. A., & Tsapras, Y. 2004, MNRAS, 353, 689
———. 2005a, MNRAS, 364, 1091
Kane, S. R., Lister, T., Collier Cameron, A., Horne, K., James, D., Pollacco, D., Street, R. A., & Tsapras, Y. 2005b, MNRAS, 362, 117
Marcy, G., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., & Jones, H. R. A. 2005, Prog. Theor. Phys. Suppl., 158, 24
Mayor, M., & Queloz, D. 1995, Nature, 378, 355
McCullough, P. R., et al. 2006, ApJ, 648, 1228
Monet, D. G., et al. 2003, AJ, 125, 984
Pinfield, D. J., Jones, H. R. A., & Steele, I. A. 2005, PASP, 117, 173
Tamuz, O., Mazeh, T., & Zucker, S. 2005, MNRAS, 356, 1466
Udalski, A., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., & Pietrzynski, G. 2002a, Acta Astron., 52, 115
Udalski, A., et al. 2002b, Acta Astron., 52, 1
Wolszczan, A., & Frail, D. 1992, Nature, 355, 145