*Astronomical Data Analysis Software and Systems XIII*
*ASP Conference Series, Vol. 314, 2004*
*F. Ochsenbein, M. Allen, and D. Egret, eds.*

# QLWFPC2: Parallel-Processing Quick-Look WFPC2 Stellar Photometry Based on the Message Passing Interface

Kenneth John Mighell

*National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, AZ 85719*

**Abstract.** I describe a new parallel-processing stellar photometry code called QLWFPC2 (`http://www.noao.edu/staff/mighell/qlwfpc2`) which is designed to do quick-look analysis of two entire WFPC2 observations from the Hubble Space Telescope in under 5 seconds using a fast Beowulf cluster with a Gigabit-Ethernet local network. This program is written in ANSI C and uses MPICH implementation of the Message Passing Interface from the Argonne National Laboratory for the parallel-processing communications, the CFITSIO library (from HEASARC at NASA's GSFC) for reading the standard FITS files from the HST Data Archive, and the Parameter Interface Library (from the INTEGRAL Science Data Center) for the IRAF parameter-file user interface. QLWFPC2 running on 4 processors takes about 2.4 seconds to analyze the WFPC2 archive datasets u37ga407r.c0.fits (F555W; 300 s) and u37ga401r.c0.fits (F814W; 300 s) of M54 (NGC 6715) which is the bright massive globular cluster near the center of the nearby Sagittarius dwarf spheroidal galaxy. The analysis of these HST observations of M54 lead to the serendipitous discovery of more than 50 new bright variable stars in the central region of M54. Most of the candidate variables stars are found on the PC1 images of the cluster center — a region where no variables have been reported by previous ground-based studies of variables in M54. This discovery is an example of how QLWFPC2 can be used to quickly explore the time domain of observations in the HST Data Archive.

## 1. Motivation

Software tools which provide *quick-look data analysis* with *moderate accuracy* (3–6 percent relative precision) could prove to be very powerful data mining tools for researchers using the U.S. National Virtual Observatory (NVO).

The NVO data server may also find quick-look analysis tools to be very useful from a practical operational perspective. While quick-look stellar photometry codes are excellent tools to create metadata about the contents of CCD image data in the NVO archive, they also can provide the user with *real-time analysis of NVO archival data*.

It is significantly *faster to transmit* to the NVO user *a quick-look color-magnitude diagram* (consisting of a few kilobytes of graphical data) *than it is* *to transmit the entire observational data set* which may consist of 10, 100, or more megabytes of data. By judiciously expending a few CPU seconds at the NVO data server, an astronomer using the NVO might well be able to determine whether a given set of observations is likely to meet their scientific needs.

Quick-look analysis tools thus could provide a better user experience for NVO researchers while simultaneously allowing the NVO data servers to perform their role more efficiently with better allocation of scarce computational resources and communication bandwidth.

Successful quick-look analysis tools must be fast. Such tools must provide useful information in just a few seconds in order to be capable of improving the user experience with the NVO archive.

## 2. QDPHOT

The MXTOOLS[1] package for IRAF has a fast stellar photometry task called QDPHOT (Quick & Dirty PHOTometry) which quickly produces good (about 5% relative precision) CCD stellar photometry from 2 CCD images of a star field. For example, QDPHOT takes a few seconds to analyze 2 Hubble Space Telescope WFPC2 frames containing thousands of stars in Local Group star clusters (Mighell 2000). Instrumental magnitudes produced by QDPHOT are converted to standard colors using the MXTOOLS task WFPC2COLOR.

## 3. QLWFPC2

I have recently implemented a parallel-processing version of the combination of the QDPHOT and WFPC2COLOR tasks using the MPICH[2] implementation of the Message Passing Interface (MPI) from the Argonne National Laboratory.

This new stand-alone multi-processing WFPC2 stellar photometry task is called QLWFPC2[3] (Quick Look WFPC2) and is designed to analyze two complete WFPC2 observations of Local Group star clusters in less than 5 seconds on a 5-node Beowulf cluster of Linux-based PCs with a Gigabit-Ethernet local network. QLWFPC2 is written in ANSI C and uses the CFITSIO[4] library (from HEASARC at NASA's Goddard Space Flight Center) to read FITS images from the HST Data Archive, and the Parameter Interface Library (PIL[5]) (from the INTEGRAL Science Data Center) for the IRAF parameter-file user interface.

---

[1] `http://www.noao.edu/staff/mighell/mxtools`

[2] `http://www-unix.mcs.anl.gov/mpi/mpich`

[3] `http://www.noao.edu/staff/mighell/qlwfpc2`

[4] `http://heasarc.gsfc.nasa.gov/docs/software/fitsio`

[5] `http://isdc.unige.ch/bin/std.cgi?Soft/isdc_releases_public#osa-2.0`

## 4. QLWFPC2 Performance

The current implementation of QLWFPC2 was tested on a Beowulf cluster composed of 5 single 1.8-GHz AMD Athalon CPUs with 3 GB total memory interconnected with a Gigabit-Ethernet local network and 120 GB of NFS-mounted disk and an additional 40 GB of local disk.

QLWFPC2 running on 4 processors takes about 2.4 seconds (see Figure 1) to analyze the WFPC2 archive data sets u37ga407r.c0.fits (filter: F555W; exposure: 300 s) and u37ga401r.c0.fits (filter: F814W; exposure: 300 s) of M54 which is the bright massive globular cluster near the center of the Sagittarius dwarf spheroidal galaxy. QLWFPC2 analyzed over 50,000 point source candidates and reported V, I, F555W and F814W photometry of 14,611 stars with signal-to-noise ratios of 8 or better.

The analysis of these HST observations of M54 lead to the serendipitous discovery of more than 50 new bright variable stars in the central region of M54 (Mighell & Schlaufman 2004). Most of the candidate variables stars are found on the PC1 images of the cluster center — a region where no variables have been reported by previous ground-based studies of variables in M54. This discovery is an example of how QLWFPC2 can be used to quickly explore the time domain of observations in the HST Data Archive.

## 5. Recommendations

- **Buy fast machines.** QLWFPC2 almost met the design goal of 5 seconds with a single CPU. Note that a very large number of machines operating at less than 1 GHz would not be able to meet the 5 second design goal.
- **Buy fast networks.** *Gigabit Ethernet is ideally suited for today's GHz-class CPUs and is now very affordable.* Old networks operating at Fast Ethernet speeds will be bandwidth-bound for tasks requiring large ($>1$ MB) messages. The test Beowulf cluster has a latency of 90 microseconds and a sustained bandwidth of 33 MB/s for large messages.
- **Buy fast disks.** The main disk of the test Beowulf cluster can read large FITS files at a respectable 30 MB/s with 7200 rpm disks. Nevertheless, reading two WFPC2 images still takes 0.6 seconds to read – which is a significant fraction of the measured total execution times.

**Acknowledgments.** This work is supported by a grant from the National Aeronautics and Space Administration (NASA), Interagency Order No. S-13811-G, which was awarded by the Applied Information Systems Research Program (AISRP) of NASA's Office of Space Science (NRA 01-OSS-01).

## References

Mighell, K. J. 2000, in ASP Conf. Ser., Vol. 216, ADASS IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 651

Mighell, K. J. , & Schlaufman, K. C. 2004 (in preparation).

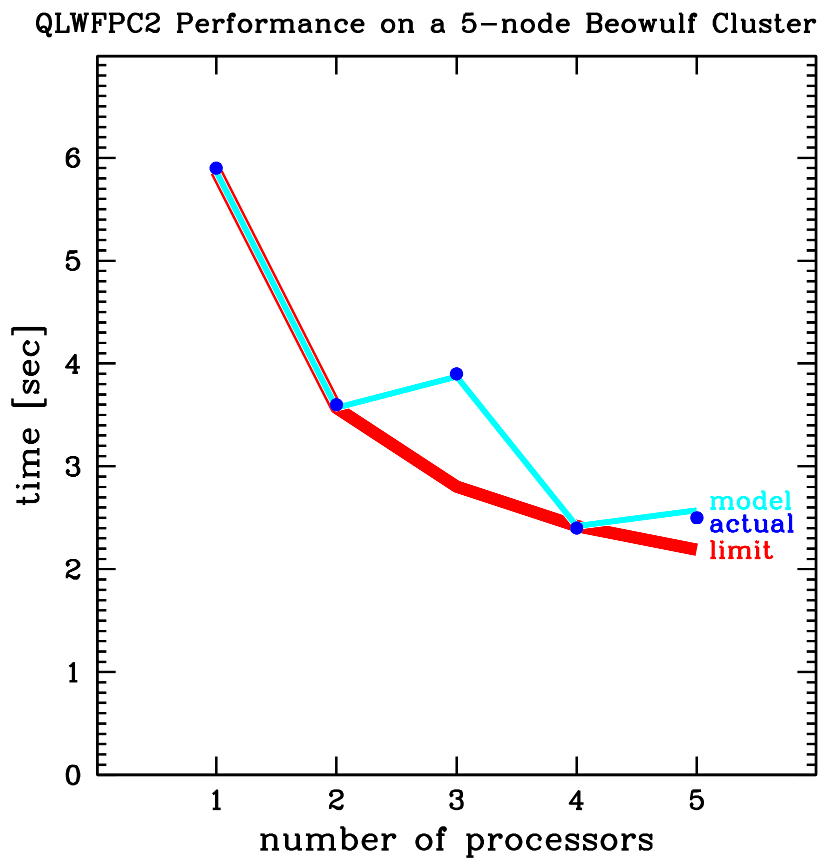


Figure 1. Typical QLWFPC2 performance results with two WFPC2 observations of a Local Group globular cluster running on a 5-node Beowulf cluster with 1.8 GHz CPUs and a Gigabit-Ethernet local network. The points show actual run times for between 1 and 5 processors; QLWFPC2 running on 4 processors takes about 2.4 seconds. The thin line shows a simple performance model based on measured cluster performance metrics (network bandwidth, disk drive bandwidth, and execution time of QLWFPC2 with a single CPU). The thick line shows the theoretical limit of performance. Note that the current version of the QLWFPC2 algorithm already meets the ideal performance values for 1, 2, and 4 processors. A single WFPC2 data set is about 10 Mbytes in size and is partitioned into four calibrated images from the PC1, WF2, WF3, and the WF4 cameras; the current QLWFPC2 analysis algorithm sends all of the image data from one WFPC2 camera to a single compute (slave) node for analysis — the increase in computation time for 3 (5) processors compared to 2 (4) processors reflects the underlying 4-fold partitioning of a single WFPC2 data set. Spreading the analysis of data from a WFPC2 camera to all compute nodes would improve the computation time for 3 and 5 (and more) processors but would not improve the results for 1, 2 and 4 processors which are already optimal.