Raw Data File Formats,

and the Digital and Analog Raw Data Streams

of the ABI PRISM™377 DNA Sequencer ©

A preliminary technical examination

by

Clark Tibbetts, Ph. D.
Professor of Microbiology

A5217 Medical Center North
Vanderbilt University
Nashville, TN 37232-2363

tibbetts@ctrvax.vanderbilt.edu

OUTLINE — PAGE

BACKGROUND

The ABI Prism 377 DNA Sequencing system is a significant upgrade of the familiar ABI 373A system. The flourescence detection component of the new system is the foundation for improved performance. The single photometer-filter wheel of the ABI 373A has been replaced by a CCD-based spectrometer. Serial scanning through each of four optical filters is functionally replaced by simultaneous monitoring of four virtual filter bins. The performance of the mechanical transport of the scanner across the platen of the sequencing gel is similar in each system, thus the ABI 377 should nominally realize a 4-fold acceleration of data acquisition. Improvements in the optical pathway of the new system increase the sensitivity of detection and quantitation of the fluorescent labels.

The architecture of the ABI 377 supports electrophoresis with 0.2 mm sequencing gels, at higher fields. Typical throughput of the ABI 373A and ABI 377, using 36 cm well to read gels, is summarized below:

	ABI 373A	ABI 377 (4X)
Samples per gel	36	36
Bases per sample	450	450
Typical accuracy	98 %	98 %
Scans per base	10	10
Filtered data channels	4	4
Pixels per scan	194	194
Scan Time - DAQ	1 sec	1 sec
Scan Time - Turnaround	0.5 sec	0.6 sec
Full Scan Cycle	6.0 sec	1.6 sec
Run time	10 hours	2.5 hours
Relative sample quantity	1.0	0.2

Both systems support disproportionately longer runs to enable extended read length, approaching the range of 600 to 800 bases per ladder.

A collaboration between ABI and the Washington University Genome Center has led to a ROM modification which doubles the lateral sampling rate of the ABI 373A scanner (388 pixels per scan). This modification supports discrimination of more sequencing ladders on each gel run, typically 48 to 60 ladders. This approaches a practical limit of about 4 lateral pixels per sample lane. Improved sample loading and lane tracking could support up to 96 samples per run (388/96 ≥ 4 pixels per lane). This “St. Louis ROM” modification has not yet (2/96) been made available to ABI 373A clients, although such an upgrade has been discussed for ABI 373A users in the genome research community. One pervasive agenda for this report is independent development of similar capacity for lading and analysis of more samples per gel with the new ABI 377 system. Such facility has been discussed by the manufacturer, but has not yet been made available to ABI 377 client laboratories.

RAW PHOTOMETRIC DATA PATHWAYS

Analog Data Pathway

Fluorescence elicited by laser illumination of the the sequencing gel follows an optical pathway through the scanner objective lens, reflection to the surface of a diffraction grating, and dispersion to the surface of a fixed CCD detector array.

Analog data acquisition from the CCD detector array involves recursion of three operations:

1) exposure of the cleared array to incident light sample (integration time),
2) binning of the vertical array of pixels, and
3) readout of the horizontal shift register array.

This cycle is repeated to sample each position (lateral pixel) across the scanned gel image, that is, 194 times during each 1.0 sec interval of traverse over the 7 inches between the left and right scanning limits. Assuming no dead time, the duration of each pixel read cycle on the ABI 377 is limited to 5.2 msec.

Mechanically, the 4X Run Mode of the ABI 377 involves similar mechanical transport of the CCD-detector as the ABI 373A filter/photometer. A full traverse of the gel by the ABI 373A scanner takes 1.5 sec, of which 0.5 sec is turn-around time. The full scan of four traverses is 6 sec. The photometer samples and integrates for 4.2 mseconds at each pixel interval of the scan, with 1 msec dead time between pixels.

Provided an appropriate timing system of triggers and gates to drive the CCD detector, raw data of the ABI 377 represents photometric measurements over defined wavelength intervals (horizontal pixel arrays) at each of 194 lateral positions per scan. The data appear as timed series of analog voltage outputs from the CCD detector, for A/D conversion, buffering, formatting and transfer from the sequencer to the workstation.

Digital Data Pathway

Buffered arrays of digitized scanning data and runtime data (gel voltage, current, watts, temperature, elapsed time, laser power, etc.) are formatted for transfer from the motherboard of the ABI 377 to the PowerMacintosh 7100 workstation. This transfer is mediated by serial data interface as RS232 at 57.6 kbaud, 8 data bits, 1 stop bit, no parity.

The 57.6 kbaud rate can support data transmission at about 7000 bytes per second. Each full scan of the gel generates 1552 bytes of photometric data per 1.6 second traverse: (194 pixels)*(4 data channels)*(2 bytes per datum). The serial interface thus could support up to seven times higher rates of data transfer to the PowerMac workstation, which **could** represent more spectral bins per scan, denser lateral sampling of the scanning interval, or combination of both.

There is a formal interactive communication protocol of control and data packet exchanges between the sequencer and the workstation. Each packet has specific structure with common features:

1) 5 byte length header -

The first five bytes of each packet are decimal ascii digits defining the (even) numbers of bytes which follow in the packet. This header is left packed with ascii space characters ('32') for packets from the workstation to the sequencer or zero characters ('48') from the sequencer to the workstation.

2) 2 bytes indicate packet type -

These are two ascii uppercase characters, appearing as one of a limited variety, namely **FR---**, **FW---**, **FE---**, or **FG---**.

3) Variable length body -

The body of the packets are comprised of the header-specified even number of bytes, less two for the two letter type indicator field. These bytes generally appear to represent short two-byte integer fields (motorola byte-word order).

Packets are transmitted from the workstation to the sequencer under control of the data acquisition software. Each packet is acknowledged by the sequencer as a response packet, typically of the same packet type, as information processed under control of the sequencer firmware. This one-for-one exchange of packets is maintained continuously before, during and after a sequencing gel run, as long as the ABI 377 instrument and the data acquisition program of the workstation are in operation.

Data File Structures

Photometric data from the sequencer is reformatted by the workstation and written to file. The initial record is the Gel File (filetype 'GEL0'; creator 'ABI1'), comprised of two major data blocks (although a third appears anticipated) and a resource fork. The general architecture of the ABI 377 gel file is remarkably similar to that generated by the ABI 373A. The first block of data is organized as fixed length records of 407 bytes. The first several of these are data-sparse headers which describe system locations of ancillary files and describe the data blocks of the gel file. These headers have single ascii uppercase characters at bytes 0 and 11 and data fields starting at byte 17. Typical gel files of the ABI 373A have 9 such header blocks, while those of the ABI 377 have 11.

These headers begin with mnemonic uppercase letters at byte 0 and byte 11 of the 407-byte record. The referenced information of each header records begins at byte 17, often representing only a single long integer with the rest of the header filled with '0' bytes. The nine header blocks of a recent ABI 373A sequencing run are abbreviated below:

Offset	Byte 0	Byte 11	Byte 17 & ff.	Comment
0	V	F	1.30	data acq'n software version
407	T	F	10 bytes	not translated
814	M	F	"INST FILE 93071256"	dye matrix file
1221	C	F	0	points to null
1628	N	F	2	number of blocks referenced
2035	S	F	3663	offset to primary data block
2442	L	F	11,334,136	length of primary data block
2849	S	F	11,337,799	offset to RGB bit map
3256	L	F	5,428,020	length of RGB bit map

The header entries of an ABI 377 Gel File are consistent with this format, although the data arguments starting at byte 17 are in some instances more verbose. The **N-F-** header of an ABI 377 file has the byte 17 argument of value 3. There are correspondingly 3 pairs of **S-F-** and **L-F-** headers in an ABI 377 Gel File. The first two pairs provide the same information as in the ABI 373A Gel Files, while the third pair appear with value zero arguments (pointers to nil).

The first **S-F-** and **L-F-** header pair point to the start and length of the primary data block, which continues to be organized as 407 byte records. The first 17 bytes of these data records serve as scan line headers, followed by 388 bytes of photometric data and ending with two bytes as a record checksum.

The 407 byte scan line records of the primary data block of the ABI 373A and the ABI 377 are organized as described below:

Byte(s)	Translation	
0	'>'	could be '<' or '^', start of new scan line record
1	'4'	
2	'0'	401 data bytes follow prompt character
3	'1'	
4 - 6	3 Hex ASCII digits for gel voltage (volts/10)	
7 - 8	2 Hex ASCII digits for gel current (mAmp)	
9 - 10	2 Hex ASCII digits for gel temp (°C)	
11	Decimal ASCII digit for current filter number (0, 1, 2, 3)	
12 - 16	5 Decimal ASCII digits for instrument's cumulative traverse counter 4 traverses per full scan, resets at 65536 to 0	
17 - 404	194 two-byte integers as single traverse photometry data	
405 - 406	C = 2 Hex ASCII digits for mod 256 checksum on bytes 0 - 404	

One apparent difference between the ABI 373A and ABI 377 Gel File data line headers is that the traverse numbers re-initiate at 00000 at the beginning of each file generated by the newer system.

The 17-byte scan line header is followed by an array of 194 two-byte short integers, recording the 194 photometric measurements across each traverse of the scanner window. These integers are recorded in Intel byte order (low byte, high byte), reminiscent of the association of the earlier ABI 370 model with an HP-Vectra workstation. This aspect of the ABI 373A data engine was unchanged, leaving the current Macintosh software to translate the primary photometric data to its native Motorola integer formats. The last two bytes of the scan line record are hex ASCII checksum digits, as the remainder from division of the sum of bytes 0 - 404 by 256.

The photometric data arrays are recorded in **boustrophedonic** fashion over the 194 lateral sampling positions: left to right across the detector window (from operator's view) when scanning with optical filters 0 and 2, and right to left when scanning with optical filters 1 and 3.

The data record format of the Gel File is an essentially verbatim transcript of the byte stream received by the Macintosh workstation from the ABI 373A sequencer. Each data packet begins with an unrecorded ascii '02' byte (STX), followed by the 405 bytes corresponding to the header and data fields. The checksum bytes are appended, followed by an unrecorded handshake return of ascii byte '06' (ACK) from the workstation to the sequencer. If the handshake byte is not returned, the synchronously timed serial advance of the filter wheel for the next scan is jeapardized and buffering of the scan data is soon compromised.

The bidirectional byte stream s of the serial interface for the ABI 377 is very different from that described for the earlier ABI 373A system. Incoming data packets are buffered and reformated by the PowerMacintosh workstation to archive the primary raw data record.

The pseudo-raster image of the gel can be rapidly superimposed with line arrays indicating the lane tracking courses of each sequencing ladder followed by the data acquisition software.The Gel File maintains records of sample lane tracking in the resource fork of the file, using a pair resources of types 'TRKC' and 'TRKP' in files from the ABI 373A. There is one such pair of resource data arrays for each DNA sample followed by the lane tracking software. These indicate the pixel stripe center positions of the sample ladders and the scan numbers at which the tracking software makes adjustments to follow drifting lane centers. Post-run editing of individual sample lane tracks exploits the Macintosh resource fork strategy well, eschewing prolonged file re-writing if the data fork itself were to be updated.

The second block of data in the gel file (referenced by the second pair of **S-F-** and **L-F-** header blocks) is a color pixel map for display of a pseudo-raster image of the sequencing gel. The offset and length of this block of data are provided in the second pair of S-F and L-F file headers. The primary arrays of photometry data are converted into a 24-bit RGB color map. Each scan line is represented as 194 four-byte integers with four additional delimiter bytes per scan line. The raster lines begin with the last recorded scan, decrementing towards the start of data collection. This facilitates the rapid updating of the gel image graphic display, drawn from top down although the time axis of the image increases from bottom to top.

The vector of raw data associated with each pixel (from the primary photometry records) is processed to represent only one of the four dye-specific colors - as red, green, blue and yellow, and the intensity of this pixel is represented on an 8-bit scale. The first of the four-byte pixel structure is 0. Byte 1, byte 2 and byte 3 encode red, green and blue pixels, in mutually exclusive fashion, respectively. Yellow pixels are represented by the same value in byte 2 and byte 3.

The low resolution of this pixel mapping suffices for review of the gel image and editing of the superimposed lane tracking profiles (see next section). The visual resolution and discrimination of individual oligomers and their 3'-terminal nucleotides is marginal. Extraction of raw data for alternative analysis and processing should target the larger primary data record of the gel file.

Post-run processing of the individually tracked samples produces the familiar trace data files (type 'ABI1'). These have essentially the same format whether generated from Gel Files of the ABI 373A or ABI 377 system.

Individually tracked DNA sequencing ladders are represented in 'ABI1' type sample data files with considerable anecdotal and quantitative detail. The ABI analytical software can display the raw as well as processed data streams from opened 'ABI1' sample data files. This directed our initial search for arrays of data fields encoding these trace data. Large, unencrypted, and uncompressed tandem arrays of **short** integers appear in the sample data files. These are easily recognized as extended regions over which every other byte has a small value, correlated with proximal, 2-byte spaced values (short integer values between 0 and 4096 have high-order bytes in the range of 1 to 16). Displays of the raw data channels or display traces are readily reconstructed from these arrays.

Other interspersed structures are recognized as arrays of ASCII characters, null-terminated c-strings or length-prefixed pascal strings. Locations and lengths of these data fields varied from file to file. Contextual translations of full files with a general byte dump program led to the consistent observation of a tandem array of 28-byte structures near the end of each sample data file. The first four bytes of these records are ASCII characters suggesting mnemonic relations to data structures of interest within the file (eg. ''DATA', 'MTRX', 'PBAS', SPAC'). These records provide a map to the idiosyncratic architecture of each data file. We soon recognized that **long** integers located at 18 and 26 bytes from the start of each file correspond to the <u>number</u> of tagged file records and the <u>offset</u> to the first of the 28-byte tagged file records, respectively. The tagged file records have a format consistent with a definition of a **c** data structure, shown below:

```
typedef struct {

        char[4] tag_name,           /* ASCII characters such as 'DATA' , 'PPOS' */

        long tag_number,  /* distinguish multiple tag entries, as '0001' */

        short data_type,             /* eg. 4 = integer, 7 = float, 10 = mm/dd/yy,
                                            11 = hh/mm/ss, 18 = pString  */

        short element_length,      /* referenced data element length in bytes */

        long number_of_elements, /* number of referenced data elements */

        long record_length,          /* length of referenced data array */

        long data_record,  /* the data, if 4 bytes or less, else the number of
                              bytes as offset from file start to the data */

        long cryptic_variable        /* often 0, often a volatile number with review
                                        and editing of data file with software */

} Tagged_Data_Record;
```

A representative file has a tandem array of 55 tagged data records, 28 bytes each, beginning at an offset of 152,583 bytes from the start of the file. Selected tagged record structures are presented in bold type, with spaces separating the translated values of their respective fields.

CMBF	**1**	**18**	**1**	**21**	**21**	**228**	**0**

This field refers to an external file description of the comb used for the sequencing gel. The data type **18**, length **1** indicates a pascal string type and the next two fields indicate the string is **21** bytes long. The string is located at byte offset **228**. The length byte at this location is 20, followed by the characters of the string "24-well sharks-tooth".

DATA	**1**	**4**	**2**	**8640**	**17280**	**405**	**0**
DATA	**2**	**4**	**2**	**8640**	**17280**	**17685**	**0**
DATA	**3**	**4**	**2**	**8640**	**17280**	**34965**	**0**
DATA	**4**	**4**	**2**	**8640**	**17280**	**52245**	**0**

These fields refer to the four tandem arrays of raw data traces, representing photometric data as short integers (type **4**, length **2**) recorded through optical filters 0, 1, 2 and 3, respectively. Each of these arrays consists of **8640** short integer values, **17280** bytes, located at offsets **405**, **17685**, **34965** and **52245**, respectively.

DATA	**5**	**4**	**2**	**270**	**540**	**69525**	**0**
DATA	**6**	**4**	**2**	**270**	**540**	**70065**	**0**
DATA	**7**	**4**	**2**	**270**	**540**	**70605**	**0**
DATA	**8**	**4**	**2**	**270**	**540**	**71145**	**0**

These fields refer to tandem arrays of run time data, representing measurements of gel voltage (volts/10), gel current, electrophoretic power and gel temperature.

DATA	**9**	**4**	**2**	**9524**	**19048**	**71701**	**5398796**
DATA	**10**	**4**	**2**	**9524**	**19048**	**90749**	**5398876**
DATA	**11**	**4**	**2**	**9524**	**19048**	**109797**	**5398844**
DATA	**12**	**4**	**2**	**9524**	**19048**	**128845**	**5394676**

These fields refer to the four tandem arrays of ABI software-processed display traces. The processed display time line does not superimpose on the raw data time line (scan numbers), since the display trace is generated by and algorithm which presents a more uniform separation between successive oligomer traces. For this file, the display trace array consists of **9524** short integer values, **19048** bytes, located at offsets **71701**, **90749**, **109797** and **128845**, respectively. The order of the 3' terminal nucleotides corresponding to these four display trace arrays is given by the following tagged record (Filter Wheel Order):

FWO_	**1**	**2**	**1**	**4**	**4**	**CAGT**	**0**

The data type **2**, length **1** indicates a simple array of ASCII characters. As only four bytes are specified, these are located as **CAGT** in the next to the last field of the record.

GELN	**1**	**18**	**1**	**9**	**9**	**249**	**0**

A 9-byte pascal string located at offset **249** names "Gel File" as the 'GEL0' file from which this sample's 'ABI1' data file was extracted.

GELP	**1**	**18**	**1**	**25**	**25**	**258**	**0**

A 25-byte pascal string located at offset **258** identifies the file pathway for relocation of the gel file (above) as "HARD DISK:373A Software:".

LANE	**1**	**4**	**1**	**2**	**2**	**3**	**0**

This sample's ladder was number **3**, as loaded onto the gel from left to right.

MTRX	**1**	**4**	**2**	**16**	**32**	**293**	**0**
MTRX	**2**	**4**	**2**	**16**	**32**	**325**	**0**

These tags point to arrays of 16 <u>signed</u> short integers, representing coefficients of 4x4 photometer-dye transformation matrices. The integer values appear to be scaled to decimal values on division by 1000. The values appear to be listed in row by row order, for transformation of the photometer data vector of filters 0, 1, 2, and 3 to the dye vector in the order C, A, G, T (from **FWO_**). Two related records (**MTXF 1, MTXF 2**) point to the name of an external matrix file, "297 MATRIX FILE". The tag numbers for these records may refer to matrices for the two sets of fluorophores, 0001 for the dye primer set and 0002 for the dye terminator set.

NAVG	**1**	**4**	**2**	**1**	**2**	**3**	**0**

This record mnemonically suggests "number averaged", and its value of 3 suggests that it refers to the lateral pixel averaging of photometer readings at the tracked pixel position(s) and the two flanking pixels at each scan. Some ROM versions of the *373A* support tracking without this lateral averaging, which would support analysis of denser arrays of narrower sample lanes across the gel.

NLNE	**1**	**4**	**2**	**1**	**2**	**24**	**0**

This appears to refer to the number of sample lanes (**24**) tracked on the original gel file.

PBAS	**1**	**2**	**1**	**774**	**774**	**149441**	**5394512**
PBAS	**2**	**2**	**1**	**774**	**774**	**151763**	**5394600**

These records are arrays of ASCII characters, corresponding to the ABI software's basecalls. Editing the called sequence introduces changes into the **PBAS 1** array and record, while **PBAS 2** points to the original, unedited array.

PDMF	**1**	**18**	**1**	**18**	**18**	**370**	**5394668**

This record points to a pascal string at offset **370**, naming the external primer-dye mobility correction file - "DyePrimer{-21m13}".

PLOC	**1**	**4**	**2**	**774**	**1548**	**147893**	**5398780**
PLOC	**2**	**4**	**2**	**774**	**1548**	**150215**	**5394508**

These records point to arrays of short integers, indicating the scan numbers of the ABI-processed display traces at which the edited and original basecalls are centered.

PPOS	**1**	**4**	**2**	**1**	**2**	**821**	**0**

This in-line data record indicates the scan number of the raw data stream at which the first ABI basecall is reported (linking the time lines of the raw and display trace data streams).

RUND	**1**	**10**	**4**	**1**	**4**	**5/13/1991**	**0**
RUND	**2**	**10**	**4**	**1**	**4**	**5/14/1991**	**0**

The start and stop dates of the gel run are indicated as in-line data: The first two bytes are a short integer for the year (1991). The third and fourth bytes indicate the month and day, respectively.

RUNT	**1**	**11**	**4**	**1**	**4**	**17:20:04**	**0**
RUNT	**2**	**11**	**4**	**1**	**4**	**7:18:03**	**0**

The start and stop times of the gel run are indicated as in-line data: The values of the first, second and third bytes correspond to a hh/mm/ss format.

S/N%	**1**	**4**	**2**	**4**	**8**	**152545**	**5394588**

This record points to four short integers at offset 152545, representing the scaling factors for the average signals recovered from the 'FWO_'-specified order of bases. In this case the values recorded are [C:161, A:325, G:72, T:45].

SMPL	**1**	**18**	**1**	**13**	**13**	**357**	**0**

This record points to a pascal string at offset 357, naming the sample as "sigma ss DNA".

SPAC	**1**	**7**	**4**	**1**	**4**	**11.1724**	**0**

The only **float** record encountered in the file is a four-byte in-line value of the average spacing between successive oligomer traces. The particular interval of the ladder over which this parameter is evaluated by ABI software has not been determined.

TRKC	**1**	**4**	**2**	**4**	**8**	**71693**	**0**
TRKP	**1**	**4**	**2**	**4**	**8**	**71685**	**0**

These records describe the lane tracking adjustments which determined the path of data excised from the gel image file on creating of this sample data file. The **'TRKC'** and **'TRKP'** records point to short integers, indicating the center pixel position and initial or adjustment scan numbers, respectively. For this particular sample, three adjustments were made from the initial setting: starting at 18 at scan 0, then at 17 from scan 3712, at 18 from scan 4448, and at 19 from scan 8570.

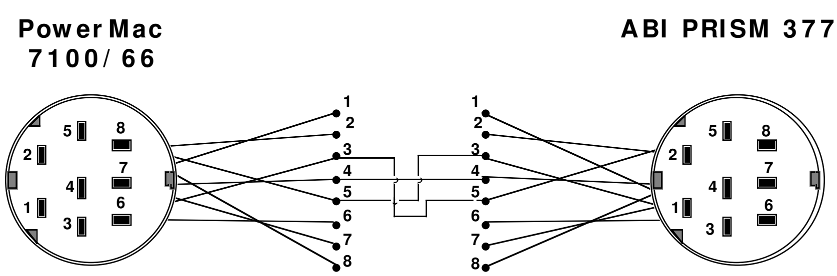
Based upon the descriptions of embedded data structures provided above, the raw data streams are readily recovered from the gel file or sample data file for experimental analysis.

ACCESS TO THE DIGITAL DATA PATHWAY

Exposed Serial Cable and Jumpers

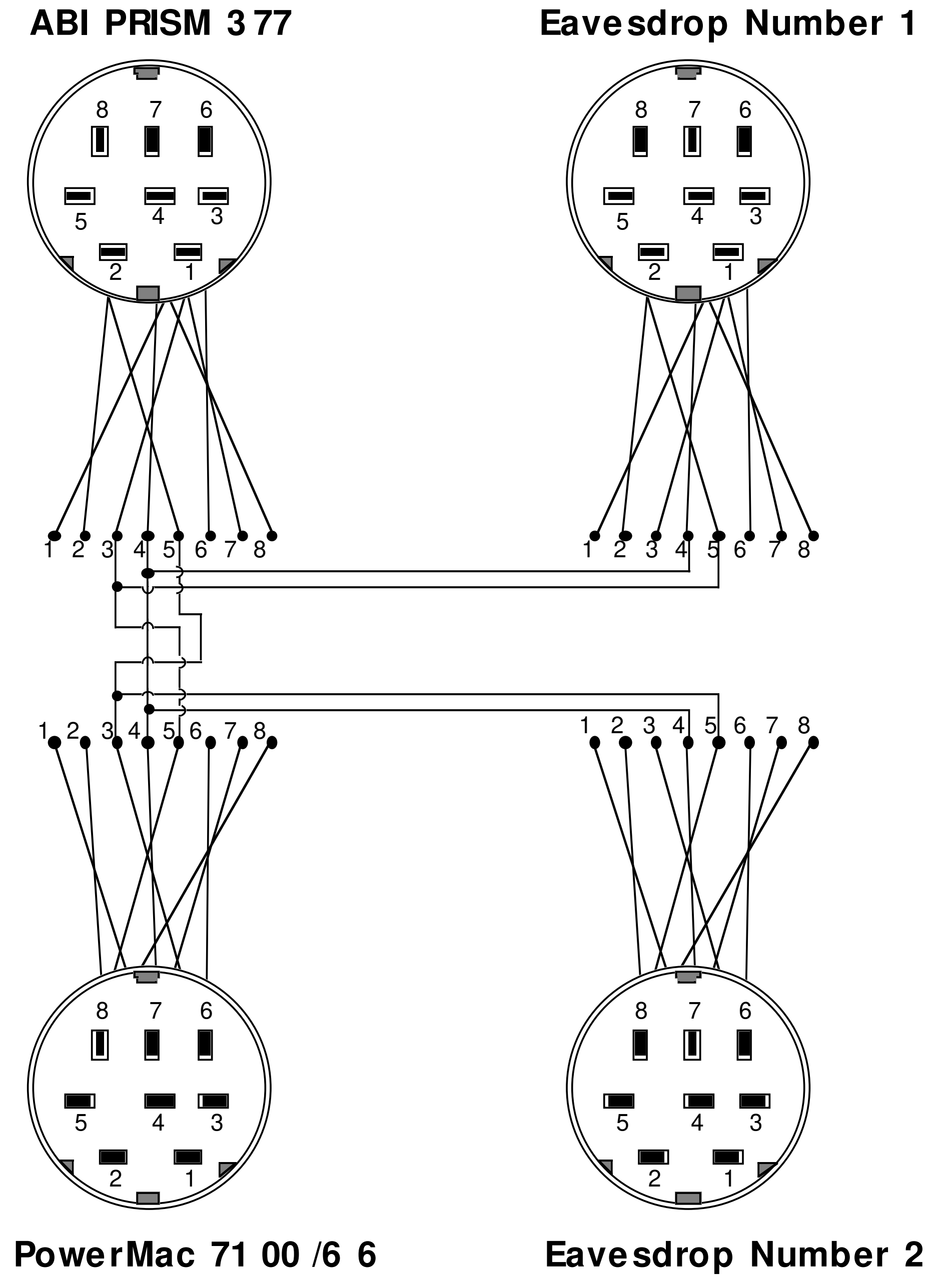
The serial cable linking the PowerMacintosh workstation with the ABI 377 is a standard cable with Mini 8 connectors, attached to the modem port of the workstation and the control/data port of the sequencer. The standard wiring of the cable crosses pins 1 and 2 (for hardware handshaking), pins 3 and 5 (for negative signal transfer), and pins 6 and 8 (for positive signal transfer). Pins 4 and 7 are respective signal grounds (note pin 7 not always connected to the Mac).

The nature of the serial interface was investigated by severing the cable and selectively joining component wires on a solderless jumper breadboard. Joining the wires connecting or crossing pins 3, 4 and 5, as shown in the diagram below, is sufficient to sustain native operation of the ABI 377 with the PowerMacintosh workstation's data acquisition software.



Bidirectional Eavesdropping Connections

The exposed serial cable was then tapped to provide two partial connections, each respectively connecting the transmit data pin 3 of the workstation or sequencer to the receive data pin of a eavesdropping Mini 8 connector. This is illustrated in the diagram below:



The eavesdropping connectors are connected to the modem and printer ports of a second workstation (Quadra 700). This computer is connected by ethernet to the local network, including the PowerMacintosh 7100 workstation for the sequencer.

Standard terminal emulation software (White Knight™) was used to monitor printable characters intercepted from the data streams during native operation of the ABI 377 sequencer system. Coherent communication was readily apparent with the serial port settings set for 57,600 baud, 8 data bits, 1 stop bit, no parity (half duplex). Direct streaming of the data to disk file provided lossless records of the data traffic monitored on either port.

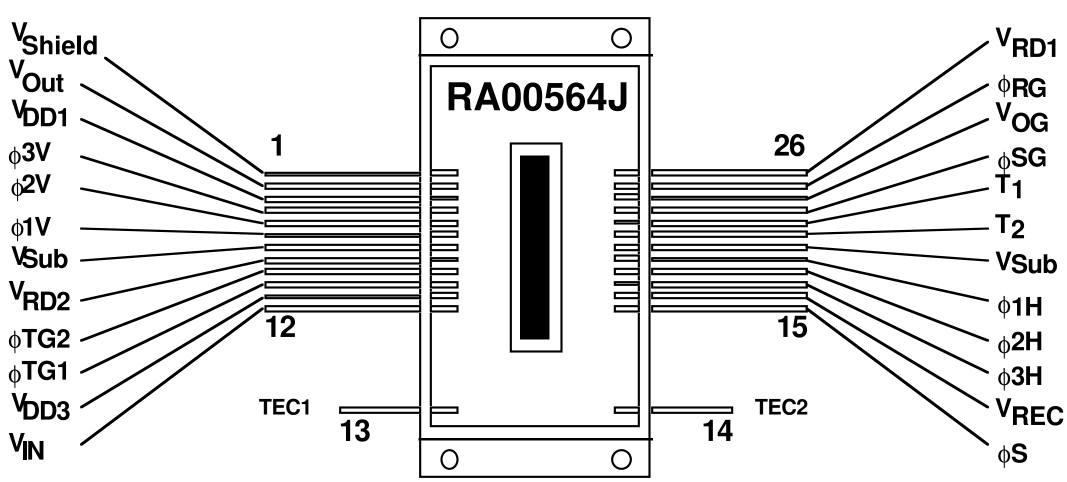
This experimental setup enables investigation of the communcation protocol between the workstation and the sequencer before, during and after sequencing runs. This is the process by which the data streams have been exposed and parsed, to eventually identify and characterize header, type and body structures of control and data packets described earlier in this report.

Sufficient information is now in hand to develop workstation interface emulation software for autonomous control of ABI 377 sequencer operations and data acquistion, mediated through its on-board firmware. Description of the communications protocols and packet structures is presented later in the **APPROACH TO AUTONOMOUS CONTROL** section.

ACCESS TO THE ANALOG DATA PATHWAY

CCD Camera

Descriptions of the new detector of the ABI 377, and the identification logo on the body of the CCD camera indicated that the detector was a product of EG&G RETICON. Their current red book provides a preliminary description of a full frame CCD image sensor (RA00564J) which appeared to match the properties of the sequencer's detector. Interviews with technical personnel of EG&G RETICON were constrained, appropriately reflecting that company's confidentiality aggreements as a subcontractor for development of the ABI 377. The specification of the RA00564J chip in the ABI 377 detector was confirmed, and there is possible interest in independent research and development of "evaluation boards" for commercial applications of this CCD chip. The retail cost of the chip itself is about $1460.



The pixel array of the chip provides 64 vertical axis sampling rows by 512 horizontal axis sampling columns. There are an additional 10 dead pixels in the horizontal shift register axis. Reading of the chip consists of the three-step cycle described earlier:

1) vertical shift register binning (64 cycles),
2) horizontal shift register steps with analog voltage amplification (522 cycles), and
3) exposure of CCD array to next imaging interval.

In the 4X run mode of the ABI 377, sampling 194 lateral positions per 1 second traverse of the sequencing gel constrains the time period for each full imaging cycle to 5.2 msec. If the read-image cycle can be reduced to 2.6 msec, the system would provide full speed scanning data at a resolution of 388 lateral pixels per traverse, realizing an objective for discrimination of more sample ladders per gel. Fluorescent signal sensitivity may not be compromised if the shortened cycle derives from accelerated offloading of the CCD chip.

The dispersed fluorescence and laser illumination from the detector's diffraction grating casts an image across the 512 + 10 pixel axis of the chip. Interview with technical personnel from ABI indicated that the resolution in this axis is approximately 0.5 nm per pixel. This is consistent with the definition of virtual filter bins embedded in the ABI 377 system's run modules. Installation of the instrument includes specification of the pixel column centering a 546 nm laser line. The four dyes used as primer labels (FAM, JOE, TAMRA and ROX) are characterized by differences of about 30 nm in the peaks of their respective fluorescence emission spectra. With calibration of the 546 nm line to pixel #209, the definition of the bins for virtual filter sets A and B are consistent with the 0.5 nm/pixel resolution of the chip.

Dye	Emission Max	Δ	Filter Set A	Δ
5-FAM	522 nm		152-175	
		32 nm		52 pix
JOE	554 nm		205-226	
		29 nm		59 pix
TAMRA	583 nm		264-285	
		24 nm		63 pix
ROX	607 nm		327-348	

(calibration 546 nm at pix # 209) **(average Δ ratio 0.48 nm / pix)**

Analog Data Cable

The RA00564J CCD chip of the ABI 377 detector is interfaced to the system through a D37 connector at the camera body. Within the camera itself there appears to be a conditioning board of proprietary design, fabricated in conjunction with the confidential ABI-EG&G development effort. The D37 connector extends by a flexible, shielded analog data cable to the mother board of the ABI 377 sequencer.

Some inferences may be drawn from the EG&G specifications of the RA00564J pinouts (below) and measurements made on specific D37 connectors at various operating states of the sequencer and its data acquisition interface.

Pinout Specifications for the EG&G RA00564J

Pin No.	Symbol	Function	Specifications
1	V_{Shield}	Video Shield	
2	V_{Out}	Video Output	
3	V_{DD1}	Drain Supply	21 volts DC
4	ϕ_{3V}	Vertical Phase 3	-11 to +7 volts
5	ϕ_{2V}	Vertical Phase 2	-11 to +7 volts
6	ϕ_{1V}	Vertical Phase 1	-11 to +7 volts
7	V_{Sub}	Substrate	0 volts DC
8	V_{RD2}	Reset Drain	
9	ϕ_{TG2}	Transfer Gate 2	
10	ϕ_{TG1}	Transfer Gate 1	
11	V_{DD3}	Drain Supply	
12	V_{IN}	Input Diode	
13	TEC1	Thermo-electric Cooler Supply 1	6.0 volts DC
14	TEC2	Thermo-electric Cooler Supply 2	6.0 volts DC
15	ϕ_S	Input Sampling Gate	
16	V_{REC}	Input Receiving Gate	
17	ϕ_{3H}	Horizontal Phase 3	-5 to +9 volts
18	ϕ_{2H}	Horizontal Phase 2	-11 to +3 volts
19	ϕ_{1H}	Horizontal Phase 1	-11 to +3 volts
20	V_{Sub}	Substrate	0 volts DC
21	T_2	Temp Diode	
22	T_1	Temp Diode	
23	ϕ_{SG}	Summing Gate	-11 to +3 volts
24	V_{OG}	Output Bias Gate	3 volts DC
25	ϕ_{RG}	Reset Gate	+5 to +15 volts
26	V_{RD1}	Reset Drain	

Two of the analog data cable's D37 connections (#10, #28) from the ABI motherboard do not pass through to the CCD camera. These are shunted to carry 22.5 volts DC to drive a cooling fan mounted at the back of the CCD camera body.

Several of the remaining pins of the cable appear to be pooled on the cable's terminal board, as indicated by low or nil resistance or common junction wiring. These sets are: (1, 7, 23), (2, 20), (3, 21), (4, 22), (8, 9, 19), (16, 34, 35), (17, 18, 36). The remaining pins appear as singular pass-through connections between the mother board and the camera. The following table summarizes measured voltages (relative to camera body ground) of the D37 connections at two operating states: workstation and sequencer on, with or without launch of data acquisition software.

Analog Data Cable D37 Connector Voltages

Pin	DAQ Off	DAQ On	Pin	DAQ Off	DAQ On
1	0.0	0.0	20	3.8	3.9
2	14.5	14.5	21	0.1	0.1
3	14.5	14.5	22	3.8	3.9
4	-14.5	-14.5	23	0.0	0.1
5	-14.5	-14.5	24	3.4	*******
6	21.4	21.4	25	3.7	*******
7	21.4	21.4	26	3.9	3.9
8	0.0	0.0	27	0.3	*******
9	0.0	0.0	28	1.8	1.8
10	-1.4	-1.4	29	4.8	4.7
11	0.0	0.0	30	2.8	2.6
12	-13.0	******	31	2.8	2.6
13	0.0	0.0	32	2.8	2.6
14	0.0	0.0	33	5.1	5.1
15	0.1	0.1	34	5.1	5.1
16	0.1	0.1	35	5.1	5.1
17	0.1	0.1	36	4.6	4.6
18	0.1	0.1	37	0.0	0.1
19	0.0	0.0			

********* Variation Patterns on Time Base of 10 msec**
Timing on pins 12, 24, 25, 27 discussed in the
APPROACH TO ENHANCED THROUGHPUTsection.

APPROACH TO AUTONOMOUS CONTROL

Serial Digital Communication Protocol

As described earlier, there is a continuous handshake exchange of control and data packets between the workstation and sequencer.

These packets have in common a length-specification header of 5 bytes, as ascii decimal digits. This header field is left padded, with ascii zero digits ('48') in packets from the sequencer, with ascii space characters ('32') from the workstation. This header field indicates the number of bytes in the packet which follow these first five bytes.

The sixth and seventh bytes of each packet are uppercase ascii letters; the first byte is generally '70' = 'F' and the second is typically 'D', 'W', 'R', 'G' or 'E'. These two characters of the packet headers are included in the length indicated by the numeric header field.

The remaining bytes of each packet (even number) appear to be parsed as as MM short integer fields.

The following exchange of packets was reconstructed from an edited 4X A sequencing module (no gel installed, electrophoresis set to 0 volts, no CCD data acquisition). A comparison with the native 4X A run module instruction set follows. The module was started from the manual control interface window, and the data acquisition interface does not set up or monitor scanning data from this mode. The eavesdropping cables were connected to printer and modem ports, each recording in stream-to-file mode.

This setup basically demonstrates the ability to edit and independently control the mechanical systems of the sequencer, and separation of the CCD-based data acquisition component from the mechanical transport of the scanner stage. This mode of operation does not produce data packets of the **10FG----->** ilk. The **----->** refers to more short integer fields, or in some cases also ascii character arrays, in return packets from the sequencer. These recorded arrays are enclosed as appendix materials. Differences in prologue and postlude handshaking numbers are minor, resulting from different times before startup of the program and manual initiation of the run process.

HSHK# PowerMac 7100 On, Launch Data Acquisition ABI 377 turned on, solid green state

1	2FD					00072FD	----->
2	10FR	123	22336	1	16	01092FR	----->
3	10FR	123	22336	1	16	00036FR	----->
4	10FR	123	22208	1	1	00036FR	----->
5	10FR	123	22208	1	1	00004FE	256
6	10FR	123	22208	2	1	00036FR	----->
7	10FR	123	22208	2	1	00036FR	----->
8	10FR	123	22208	4	1	00100FR	(Date, time, vers) ----->
9	10FR	123	22208	5	1	00036FR	----->
10	10FR	123	22208	4	1	00036FR	----->
11	10FR	123	22208	3	1	00100FR	----->
12	10FR	123	22208	4	1	00036FR	----->
13	10FR	123	22208	4	1	00100FR	----->
14	10FR	70	11256	4	3	00036FR	----->
15	10FR	70	11256	4	3	00100FR	----->
16	10FR	70	11256	8	1	00036FR	----->
17	10FR	70	11256	8	1	00100FR	----->
18	10FR	70	11256	3	1	00036FR	----->
19	10FR	70	11256	3	1	00100FR	----->
20	10FR	70	11256	1	3	00036FR	----->
21	10FR	70	11256	1	3	00100FR	----->
22	10FR	70	11256	7	1	00036FR	----->
23	10FR	70	11256	7	1	00100FR	----->
24	10FR	70	11256	1	3	00036FR	----->
25	10FR	70	11256	1	3	00100FR	----->
26	10FR	70	11256	7	1	00036FR	----->
27	10FR	70	11256	7	1	00100FR	----->
28	10FR	70	11256	1	3	00036FR	----->

(run initiated here, from manual control window)

35	630FW	(run instruction module)	00004FW	62
36	22FW	----->	00004FW	1
37	22FW	----->	00100FR	----->

(mechanical stage transport engaged, no data collection)

38	10FR	70	11256	7	1	00036FR	----->
39	10FR	70	11256	7	1	00036FR	----->
40	10FR	70	11256	3	1	00036FR	----->
41	10FR	70	11256	3	1	00036FR	----->
42	10FR	70	11256	2	1	00036FR	----->
43	10FR	70	11256	2	1	00100FR	----->
44	10FR	70	11256	1	3	00036FR	----->
45	10FR	70	11256	1	3	00100FR	----->
46	10FR	70	11256	7	1	00036FR	----->
47	10FR	70	11256	7	1	00100FR	----->
48	10FR	123	22208	4	1	00036FR	----->
49	10FR	123	22208	1	1	00100FR	----->
50	10FR	123	22208	4	1	00036FR	----->
51	10FR	123	22208	4	1	00100FR	----->
52	10FR	70	11256	1	3	00036FR	----->
53	10FR	70	11256	1	3	00100FR	----->
54	10FR	70	11256	7	1	00036FR	----->
55	10FR	70	11256	7	1	00100FR	----->
56	10FR	70	11256	1	3	00036FR	----->
57	10FR	70	11256	1	3	00100FR	----->
58	10FR	70	11256	7	1	00036FR	----->
59	10FR	70	11256	7	1	00100FR	----->
60	10FR	70	11256	1	3	00100FR	----->

Part of ongoing investigation involves survey of the dialog interchanges with execution of individual commands and command sets (edited modules), to provide the familiarity with the system required to program and alternative data acquisition interface for the system.

The **00036FR----->** packets are often data-sparse, as for example, those following the early setup exchanges have the following array of 17 short integer values:

00036FR 1 256 1 256 0 0 0 0 0 0 2 512 0 0 0 0 0

The five **00036FR----->** packets which follow transmission of run module instructions use fields to recapitulate the setup parameters, as:

00036FR 1 1 15 0 0 2 0 0 0 0 0 0 0 0 0 0 0 (First packet)

00036FR 1 0 0 0 0 209 546 1600 16396 10486 12 250 206 0 0 0 0 (CCD parameters)

00036FR 1 51 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 (Gel Temperature)

00036FR 1 400 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 (Laser power set)

00036FR 1 300 600 200 0 0 0 0 0 0 0 0 0 0 0 0 0 (Electrophoresis parameters)

(these packets taken from eavesdrop of a native 4x A setup)

After the setup of the run, and before the arrival of scanning data, these packets merely repeat the pattern:

00036FR 1 1 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0

The **00100FR----->** packets have 49 short integer fields. Once new run parameters have been written from the workstation, these are characterized as shown on the following page. Many field values appear to represent runtime updates of the system parameters.

00100FR	field #	value	property (?)
	1	3	
	2	1	
	3	291	read Gel kVolts/10
	4	239	read Gel mAmp*10
	5	69	read Gel Watts
	6	1	
	7	1	
	8	390	read Laser Watts*10
	9	26	read Gel Temperature °C
	10	1	
	11	0	
	12	1	
	13	32	?
	14	1	
	15	0	
	16	1	
	17	0	
	18	1	
	19	210	Set runtime (minutes)
	20	79	
	21	60	Scan speed? Inches/sec * 10 ?
	22	55	
	23	12579	Remaining runtime (seconds)
	24	12600	Run duration (seconds)
	25	65535	
	26	0	
	27	65535	
	28	0	
	29	0	
	30	0	
	31	0	
	32	0	
	33	0	
	34	1	
	35	400	Bytes per scan record chunk?
	36	194	Lateral pixel sampling
	37	0	
	38	0	
	39	8	Elapsed Volt*hours (!!!)
	40	0	
	41	0	
	42	0	
	43	0	
	44	0	
	45	0	
	46	0	
	47	0	
	48	0	
	49	0	

Edited Run Module and Command Structures

The run module instruction set of functions is the basis for initializing and controlling the electrophoresis and data acquisition processes. Investigations of the data packets related to single and multiple commands, executed from the manual control window provides some insight as to the management of the control process. The manual control window further provides a partial list of control functions and associated function numbers (and parameter fields). Additionally, one of the technical run modules accompanying our system (Module Burn 2 (12/22)), provided comment fields identifying specific function numbers.

Examination of the resource fork of the data acquisition software also provides an expanded list of control functions, with associated numbers related to position in the list. This is corroborated by strings listing arrays of control functions in the firmware.

One of the curious features of the aggregate list is the inclusion of functions (names) with no apparent relation to the operation of the ABI 377. These now appear to be an overlapping set of functions intended for the sample-handling and gel loading operations of the newly released ABI 310 single capillary system.

The format of the Resource Fork entries Func ID = 128 of ABI Prism™ Collection presents each instruction (function) as a fixed length, 60-byte list element. The first two bytes (hex) indicate the function number. The function name (ascii chars) begins at byte 13 (start counts at 0). The two bytes preceding the function names often carry a parameter value; if so, then the scale of the parameter is often listed using ascii characters at byte. Occasional functions appear to have multiple parameters. These parameter values may represent defaults, **check limits**, or perhaps updated values from the most recent run conditions.

This information, may present a significant pathway to introduce functional changes in the operation of the system. Resource Fork structures are readily edited and updatable, without requiring recompilation of the entire program. Values from the Resource Fork structures might be used to update parameter arrays for the 377 motherboard/firmware.

#	Command	Parameter	Value	(Apparent ABI 310 functions are inset)
0	Idle			
1	Electrophoresis On			
2	Electrophoresis Off			
3	Electrophoresis Volage	V	5000	
4	Electrophoresis Current	mA	60	
5	Electrophoresis Power	W	300	
6	Laser On			
7	Laser Off			
8	Laser Standby			
9	Laser Run			
10	Laser Power	mW*10	400	
	11	Home Autosampler X,Y		
	12	Go To Vial	#	96
	13	Sample Vial	#	96
	14	Home Autosampler		
	15	Autosampler Up		
	16	Autosampler Down		
	17	Home Syringe		
	18	Syringe Down	steps	0
	19	Syringe Up	steps	0
	20	Syringe Pump	minutes	10
	21	Syringe To	position	999
	22	Open Door		
	23	Close Door		
	24	Temperature Gain		
	25	Temperature Off		
	26	Temperature	°C * 10	1016
27	Start Data Collection	minutes	0	
28	Stop Data Collection	minutes	0	
29	Wait	seconds	32000	
30	CCD Offset Skip	pixels	500	
31	CCD Blank Read	pixels	500	
32	CCD 1st Filter Skip	pixels	500	
33	CCD 1st Filter Read	pixels	500	
34	CCD 2nd Filter Skip	pixels	500	
35	CCD 2nd Filter Read	pixels	500	
36	CCD 3rd Filter Skip	pixels	500	
37	CCD 3rd Filter Read	pixels	500	
38	CCD Remainder Skip	pixels	500	
39	CCD 4th Filter Read	pixels	500	
40	CCD Time between Reads	ms	1000	
41	CCD Read Integration Time	ms	1000	
42	CCD Vertical Shift CyclesScan		20	
43	CCD Horizontal Shift CycleScan		200	
44	CCD Vertical Skip Shifts		64	
45	CCD Vertical Read Shifts		64	
46	CCD Horizontal Skip Shifts		20	
47	CCD Horizontal Read Shifts		20	
48	CCD Points Per Scan	points	388	(possibly interesting as limit!!!)
49	TPU Tics	ms	0	
50	Message	message number	100	
	51	Open Buffer Valve		
	52	Close Buffer Valve		
	53	OUTER BEGIN LOOP		
	54	OUTER END LOOP		
	55	Inner Begin Loop		

#	Command	Parameter	Value	(Apparent ABI 310 functions are inset)
(continued)				
	56	Inner End Loop		
	57	Function Not Defined		
	58	Function Not Defined		
	59	Function Not Defined		
	60	Go To Autosampler Calibrate Pt	position	
	61	Move Autosampler Towards X+	steps	
	62	Move Autosampler Towards X-	steps	
	63	Move Autosampler Towards Y+	steps	
	64	Move Autosampler Towards Y-	steps	
	65	Calibrate Autosampler at Point	point	
66	Cold Boot			
67	Function Not Defined			
68	Function Not Defined			
69	Function Not Defined			
70	Function Not Defined			
71	Function Not Defined			
72	Function Not Defined			
73	Function Not Defined			
74	Function Not Defined			
75	CCD Gain	CCD Output	8	
76	CCD Offset	CCD Output	4096	
77	CCD Pixel Position	pixels	512	
78	Function Not Defined			
79	Function Not Defined			
80	CCD Scan Data On			
81	CCD Scan Data Off			
82	Stage Home			
83	Data Collect - No Scan	minutes	32000	
84	Imaging On	minutes	32000	
85	Temperature Control On	°C	60	
86	Temperature Off			
87	Temperature Idle			
88	Function Not Defined			
89	Function Not Defined			
90	Stage Left	0.001 in	8000	
91	Stage Right	0.001 in	8000	
92	Stage Period	ms	30000	
93	Stage Speed	0.1 in/sec	100	
94	Stage Acceleration	units	200	
95	Stage Channel Time	units	2000	
96	Skip Counts Before Collect	encoder counts	8000	
97	Function Not Defined			
98	Function Not Defined			
99	Function Not Defined			
100	Shutter Open			
101	Shutter Closed			
102	Laser Full Run			
103	Laser Tube Saver Time			
104	Function Not Defined			
105	Function Not Defined			
106	Function Not Defined			
107	Function Not Defined			
108	Function Not Defined			
109	Function Not Defined			

When run modules are sent from the workstation to the sequencer for execution, they are transmitted as ascii arrays of 10 bytes per instruction. Each 10-byte instruction is composed of 2 null-terminated C-strings, of lengths 3 and 5 bytes. The first of these represents an individual function, identified by number from the above list. The second field represents either zero, or a parameter value, presumably subject to compiled software or mechanical hardware limits.

The run module instructions are transmitted in a packet comprised of numeric and lettered header, four integer fields, and the n*10 byte instruction sets. The standard 4X A Run Module is thus represented as:

4X A Run Module Packet Header	fct	param	comment
610FW 70 4944 1 60	029	00000	wait 0 seconds
	029	00000	
	029	00000	
	029	00000	
	029	00000	
	102	00400	laser full on @ 40 mWatts
	010	00400	laser on @ 40 mWatts
	029	00000	
	029	00000	
	085	00051	set temperature to 51 °C
	029	00000	
	029	00000	(086 00000 here would shut off temperature)
	029	00000	
	005	00200	electrophoresis power limit 200 Watts
	004	00600	electrophoresis current limit 60 mA
	003	03000	electrophoresis voltage set @ 3000 V
	029	00000	(003 00000 sets to 0 V for no gel run)
	029	00000	(002 00000 hangs system looking for current)
	029	00000	
	031	00010	read 10 blank pixels from horizontal shift register
	030	00151	skip 151 pixels
	033	00024	read 24 pixels for dye # 1 (5-FAM)
	032	00029	skip 29 pixels
	035	00022	read 22 pixels for dye # 2 (JOE)
	034	00037	skip 37 pixels
	037	00022	read 22 pixels for dye # 3 (TAMRA)
	036	00041	skip 41 pixels
	039	00022	read 22 pixels for dye # 4 (ROX)
	038	00164	skip 164 pixels
	029	00000	
	029	00000	
	029	00000	
	076	00250	CCD offset counts (global baseline offset)
	075	00002	CCD gain (set at 1, 2, 4, or 8 times CCD count)
	042	00001	CCD V shift cycles scan
	043	00005	CCD H shift cycles scan
	044	00000	CCD V skip shifts
	045	00064	CCD V read shifts
	029	00000	
	029	00000	
	029	00000	
	048	00194	CCD points per scan
	095	00206	Stage channel time (units)
	090	00500	Stage left (inches/1000)
	091	07500	Stage right (inches/1000)
	096	01053	Skip counts before collect
	092	01600	Stage period (ms)
	093	00060	Stage speed (0.1 in/sec)
	094	00016	Stage acceleration (units)
	029	00000	
	029	00000	
	029	00000	
	029	00000	
	029	00000	
	080	00210	CCD Scan Data On (minutes)
	081	00000	CCD Scan Data Off (minutes)
	029	00000	
	029	00000	
	029	00000	
	029	00000	

Scanning Data Packets

Scanning data packets are formatted from the digitized data and buffered on the sequencer motherboard. Data solicitation packets from the workstation attempt to read buffered arrays corresponding to four complete traverses of the mechanical stage. This corresponds to 16 chunks of data, each representing the 194 lateral pixels for one of the four virtual filter bins. It appears that when there is insufficient data in the sequencer's output buffer, the workstation will grab lesser amounts, corresponding to 4, 8, or 12 such chunks. On subsequent scans the data solicitation packet from the workstation specifies less than 16 chunks, to bring the total including the preceding fetch up to a cluster of 16 data chunks.

Each data packet from the sequencer has an alpha-numeric header of type **06406FG##, 04806FG##, 03206FG##** or **01606FG##**. The **#** characters each represent a short integer header, values typically 4, 8, 12 or 16. The **FG** and the two **#**'s correspond to the 6 byte component of the numeric header field.

Following this header, each data chunk corresponds to 400 bytes. Of these, twelve represent six short integers as a chunk's subheader, followed by 388 bytes representing the 194 short integers of photometric data. The photometric data are presented in the virtual filter order 1, 2, 3, and 4 in each of the clustered sets of 4 chunks. Alternating clusters of four chunks present each filter's data in opposite order: 0->193 for left to right scans, and 193->0 for right to left scans. The byte orders of the short integer arrays are consistently MM-type, or HI-LO Motorola format.

We have not yet observed instances where the left-right scan order shifts with respect to the order of virtual filters, as appears to occasionally occur in the recording of data from the ABI 373A.

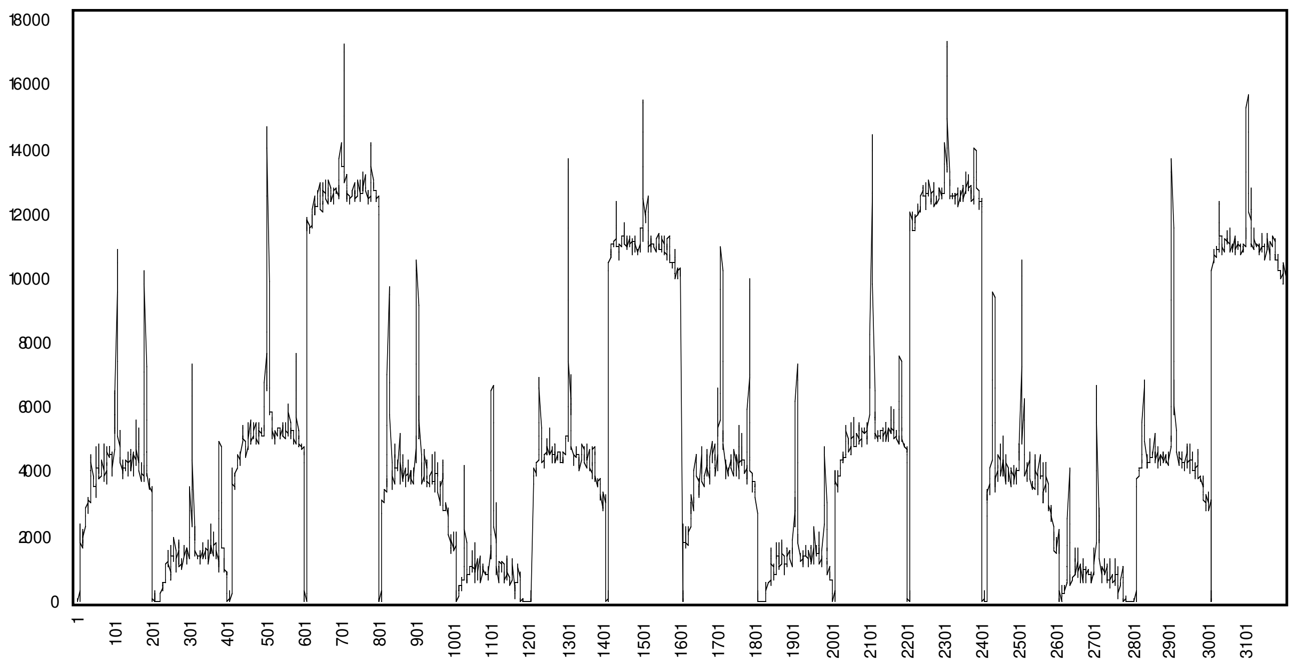
The six integer sub-header preceding each chunk of data has fields which include updates of the typical EPT variables: gel voltage (Volts/10), current (mAmps), power (Watts), temperature (°C). The last of the subheaders is a number associated with the numbering or labeling of the chunk (eg. 256, 257, 258, 259, etc.). These may mark each profile according to the filter number represented and the left-right order of the profile.

The following figure illustrates lossless recovery of scan data for packets of types **06406FG##** and **01606FG##.** These were recovered by eavesdropping on the serial link from the sequencer to the workstation during a Run 2X A module sequencing exercise. The mirror image pixel order is apparent in alternating clusters of four filter profiles, in the longer data packet structure. The CCD offset instruction (#76) of the 2X run module is typically set at 750, but for this particular region of the gel in place, the offset is excessive, and part of the second filter’s profile is truncated by the value-zero abscissa of the plot.

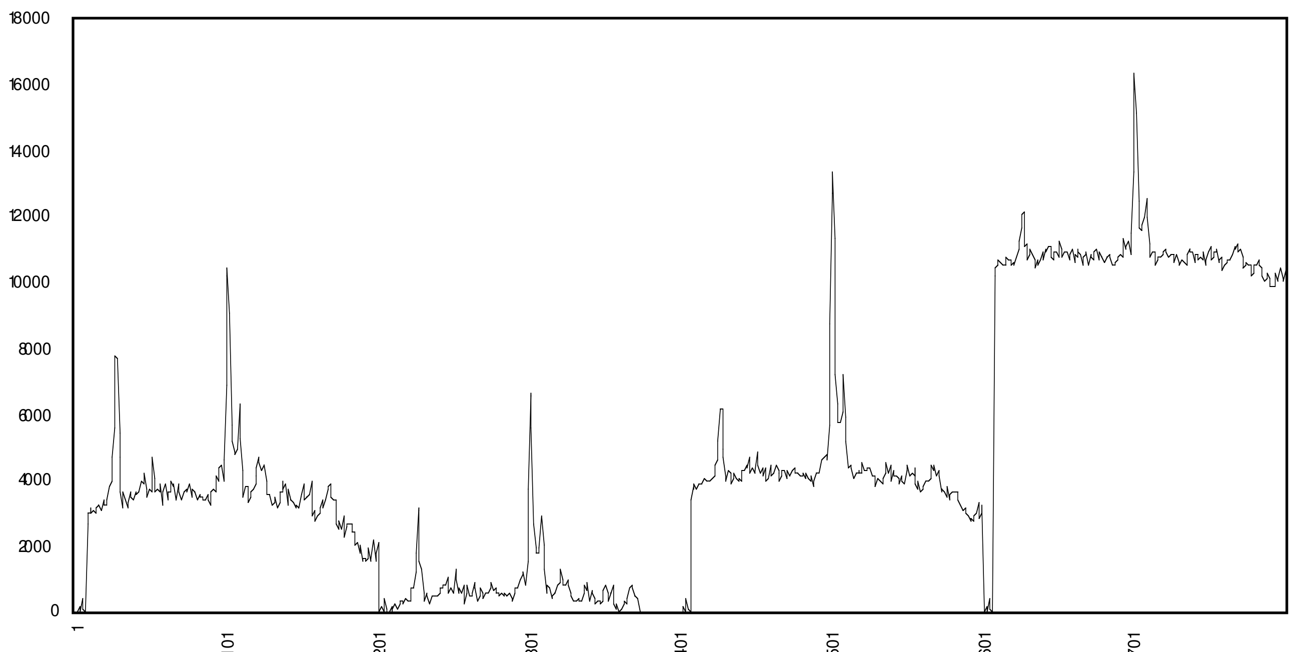
Scanning Data Packet Arrays

- 400 bytes per bit plane per scan, in blocks of 200 short integers (6 header and 194 data)

06406FG @ 8126 - 3200 short integer array of data packet



01606FG @ 66001 - 800 short integer array of data packet



Emulated Workstation Interface to Firmware

The Run Module File structure may be edited from within the Data Collection application, through the Manual Control Window. This approach uses two-column spreadsheet text editing to alter the called function-parameter steps, one step per row. Edited, saved run modules may be directly executed through selection in the Run Window. Manual control supports execution of the run modules, although without the scan display support.

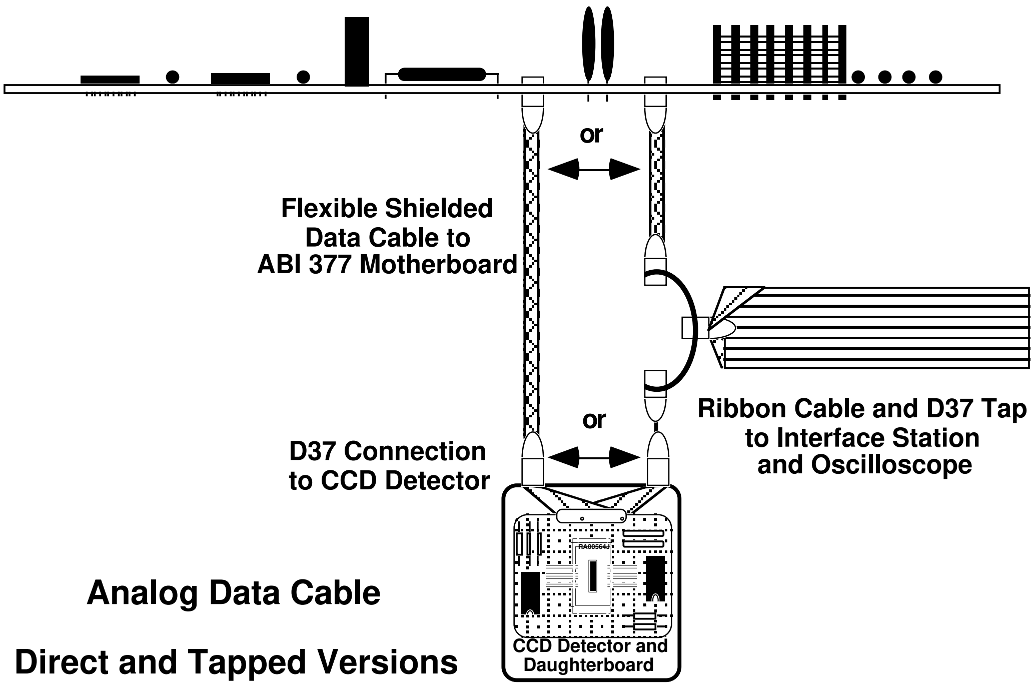
Alternatively, the Run Module may be edited directly in its binary encoded format. These files follow the typical ABI formats. Byte 18 provides a long integer for the number of tagged file records in the module. The location of the tagged file record is given as a long integer at byte 26. Run modules typically have one such record, offset to byte 128. Here a 28-byte long 'Fncs' record is readily decoded to indicate the offset address to the actual data (at byte 1024).The data here represents the encoded list of run module instructions and parameters.

Further specification of the order and structure of control and data packet handshaking should support development of an autonomous interface to the sequencer firmware. A concern remains for the monitoring of systems operations using this approach, particularly with respect to potential hazard conditions (door open, laser exposure, high voltage, buffer leaks). Not all of these alert situations have been mapped to the data fields of control packets returned to the workstation from the sequencer.

APPROACH TO ENHANCED THROUGHPUT

Analog Interface Timing

As described in a previous section, an eavedropping tap was placed at the connection of the shielded analog data cable to the body of the detector CCD camera. As with the digital data intercept, this arrangement provided opportunity to monitor individual lines between the ABI 377 motherboard and the detector without interfering with the native mode of data acquisition. Voltage measurements revealed that most of the lines of the cable were set and constant regardless of the mode of operation of the instrument - standby or scanning.



Analog Data Cable

Direct and Tapped Versions

Measurements of voltage on lines 12, 24, 25, and 27, however, revealed timing patterns suggestive of driver control of the detector or analog output from the detector. Discussions with technical staff of EG&G Reticon were necessarily limited, respecting the confidentialityof their contracted relationship with ABI. Our own research, however, confirmed that the ABI 377 detector included not only the RA0564J CCD chip, but a control interface board. The connections of the elaborate shielded cable between the camera and the ABI 377 motherboard were systematically investigated, and found to represent only analog data interchanges.

Nominally similar cameras are manufactured by EG&G Reticon. A specific example is 512x512 pixel array RA0512J chipas part of the MC4005 Scanning Matrix Camera. The ABI 377 detector camera simplifies the scanning arrangement compared with the MC4005 camera, as there are only 64 rows of the 512 horizontal array, and these can be fully binned for serial readout through the shift register and track and hold amplifier. The camera-internal daughterboards of these other EG&G systems typically interface to input lines for DC supplies, frame start and shutter and output analog video. The daughterboards program, in hardware, the responses to external timing and trigger signals, and support the amplifiers for clamp and hold signals destined for A/D conversion. These external functions would be readily supported by the motherboard's MC68332 multiprocessor. Information regarding the versatile capabilities of the Motorola MC68332 can be obtained on browsing through the Motorola or Robert Boys WWW home pages:

http://www.motorola.com or http://www.sentex.net/~rboys

The ABI 377 detector interface is channeled through an elaborately shielded data cable bearing D37 connectors. As mentioned earlier, lines 10 and 28 from the motherboard through this cable are shunted to provide the 22 volt source for an externally mounted fan on the CCD camera housing. The remaining 35 lines interface the motherboard to the camera daughterboard, and then to the 26-pin interface of the RA 0564J chip itself. Most of these lines provide conditioned potentials and currents to drive operations of the chip, including the thermo-electric cooling, current drains, and the clamp and hold circuits of the on-board chip amplifier.

Preliminary investigations suggest that lines 24, 25, and 27 provide the requisite external timing patterns to control the CCD camera's exposure and readout cycles. The voltage timing patterns on line 12 carry the held video outputs of the sampled horizontal virtual filter bins. These are described in the following section.

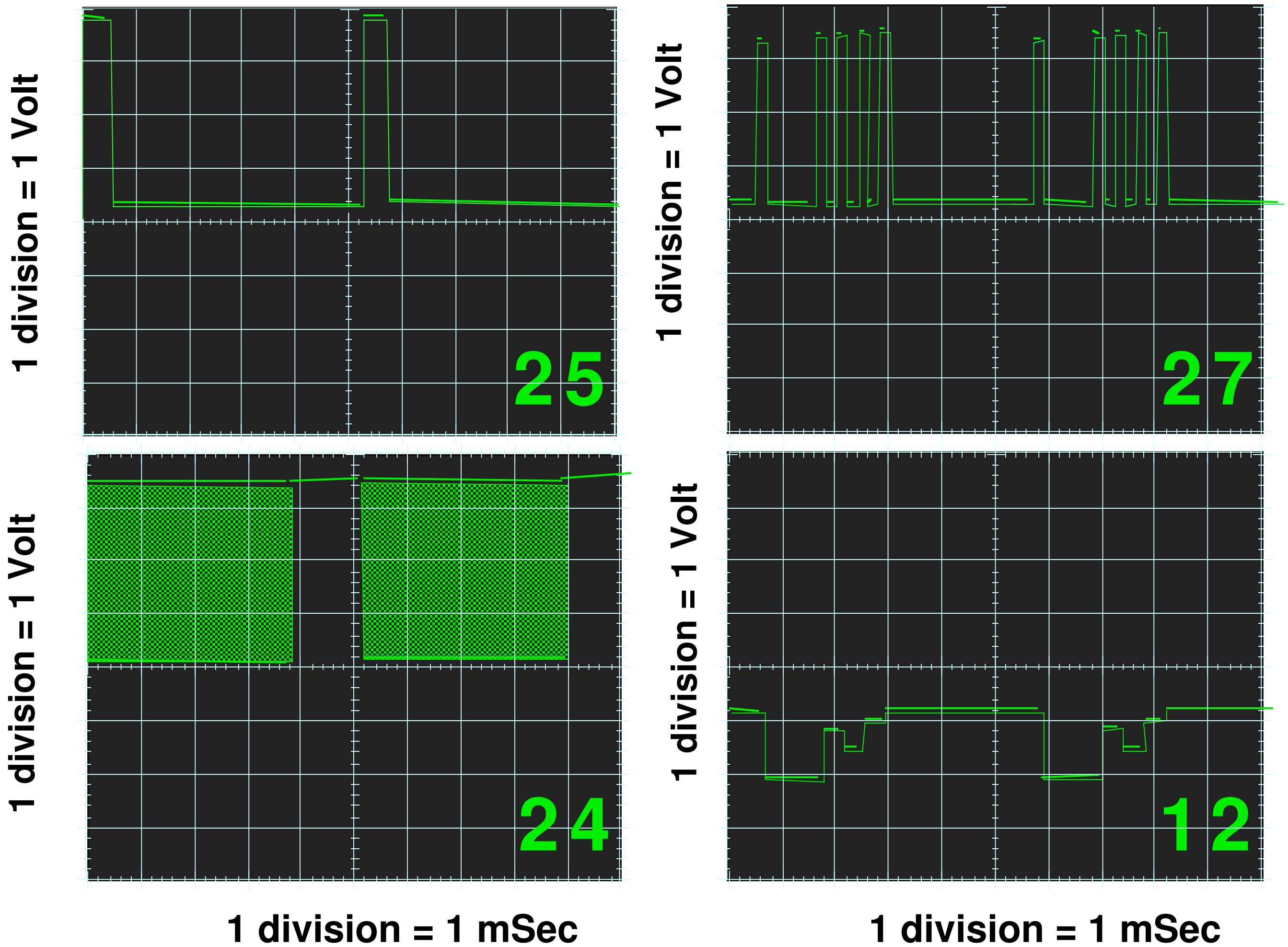
Read Cycle Sampling Control - Increased Lateral Resolution

The constant and time-variant voltages of the analog data cable lines (12, 24, 25, 27) have been investigated using the shunted lines as inputs to an oscilloscope. A fixed ground line was provided as connection to the metal body of the CCD camera.

The figure on the following page illustrates traces of these four lines of immediate interest, with a timebase of 1 msec/division and voltage scale of 1 volts/division. The tracings were recorded with the sequencer running under control of edited prototype run modules, in this specific case derived from the standard “Run 4X A” module. The instruction lines were altered to shut off the gel temperature circulator, hold gel voltage to zero, and reduce the run time to 20 minutes. A cassette-mounted acrylamide gel sandwich was installed on the instrument’s support structure, providing reflections and fluorescent signals at various positions as the laser scans across the detector window. Similar setups have been monitored with edited modules representing “Run 2X A” modules, or with alternatively defined arrays of five read-skip virtual filter cycles.

The scanner transport traverses back and forth across the detector window with a swath of about 7 inches. Data gathering is limited to about 6 inches of this interval, for collection of 194 pixels of data in 1 second (constant velocity specified for the transport on scanning in 4X mode is 6 inches per sec). This leads to a calculated 5.2 milliseconds per pixel readout during scanning, and this is indeed the duration of the repeating signal patterns observed in the traces below.

Native DAQ Timing from MC 68332



Note that the constraints of the native ABI firmware on the definition of the filter arrays have not yet been fully examined. It is clear that a baseline and four filter interval outputs are determined (5 A/D conversions) for each pixel's download from the imaging area of the CCD chip. The vertical binning step first pools all 64 rows of the imaging area, and this is followed by alternating amplification or drain of 10 bins, shifted out the (512 + 10) wells of the horizontal register. After a dead-time interval, the vertical-horizontal binning process re-initiates for the next pixel's readout.

The 5.2 msec cycle is expected for the system as it corresponds to scanning the 194 lateral positions of the detector window each 1.01 second. The remaining 0.6 seconds of each scan cycle likely represent acceleration, deceleration and dead turnaround periods for the mechanical stage. This is also consistent with the mechanical timing of the scanner of the earlier ABI 373A system.

The slower 2X scanning mode has the same features and durations of timing signals on the three input channels, but is followed by a longer dead time between readouts, with a total cycle of 13 msec. The 194 pixel readings per 2X traverse requires a total of 2.5 seconds. Since the overall scanner traverse cycle at 2X is 3.1 seconds, this again leaves 0.6 seconds of non-sampling turnaround time per traverse.

We first considered that the dead time between the binning-readout cycles might correspond to a (shuttered) period of exposure of the chip's imaging area, preceding each readout cycle. However this does not appear to be the case. Rather, the camera appears to operate in a simpler mode of unshuttered repeating readouts, leading to some (probably unimportant) image smearing from pixel to pixel. Each vertical binning operation pools accumulated charges into the horizontal well array, and this represents the exposure of the chip to incident light since the last such binning operation.

Ignoring the constraints of the ABI firmware, with respect to the formatting of digitized data streams from the camera and A/D conversion hardware, one could anticipate a modest increase in lateral scan rate under the nominal 4x conditions, by minimizing the dead time between readouts. From measurements as in the traces shown above, the ABI-native software-firmware and hardware controls the readout of each pixel in a time interval of about 3.9 milliseconds.

The ABI-firmware and collection software constrain analysis of the standard collection GelFiles to loadings with 36 or fewer sample lanes. We can propose a 33% increase in lateral scanning resolution by continuously re-initiating readings - to provide for example, 256 pixels per 6 inch traverse per second. This would readily permit the <u>practical</u> numbers of sample lanes per gel to be increased from 48 to 64 (assuming an average of at least 4 pixel stripes per lane).

The traces of native timing and data lines to the camera shown above were generated from the instrument by setting the external trigger of the oscilloscope to the rising edge of signal from line 25, a singular event of each readout cycle. The line 25 stays high for approximately 0.4 msec, then drops down for the duration of the 5.2 msec readout cycle. In 2X simulations (not shown here) the rise-and-fall timing of line 25 is the same (0.4 msec), but the duration of the full 2X cycle is 13 msec, as described above. During the period that line 25 is high, there are 64 pulses comprising the signal on line 24. While line 25 is high, it appears that line 24 provides the pulses to vertically bin out each of the 64 rows of the chips imaging area. The pulses on line 24 average approximately 160 kHz, although examination of the traces with an expanded time base demonstrates that these are asynchronous pulses, not simply a periodic waveform. This is consistent with counted pulsed driven from a circuit tied to a clock base of higher frequency. The ABI 377's motherboard's MC68332 CPU multiprocessor supports 1 Mhz and 5 Mhz clockbases, interestingly matched to EG&G's specifications for their RA0564J CCD chip. Either timebase could easily be counted into the pulse array of line 24 shown for the 4X mode readouts.

When line 25 falls back, the oscillations from line 24 continue (short interuption) for about 3.4 msec, corresponding to additional 522 pulses at an average of 160 Khz. This is what would be required to pass the entire horizontal array of wells out through either the chip's amplifier or current drain. Recall that the width of the imaging area of the chip is 512 pixels. There are 10 additional wells between the 512 horizontal wells and the amplifier/drain gate. Reading these after a clearout of the chip provides an opportunity to measure a baseline voltage from 10 wells that have received no binned charges.

After the 3.8 millisecond readout cycle, the line 24 stays high until the next readout cycle initiates. In 4X and 2X mode this dead period corresponds to about 1.3 msec and 9.2 msec, respectively. The operative model, is that the imaging area continuously accumulates charge from incident light through the entire readout and dead periods of each full pixel cycle. Thus scanning at 2X mode nominally offers increased sensitivity through longer image sampling per pixel, in a ratio of approximately 13 : 5. With simple assumptions relating signal and noise (square root of sample time), this suggests that detector sensitivity is about 60 % greater in the 2X mode than in the 4X scanning mode.

Virtual Filter Definition - Increased Spectral Bandwidth

Earlier in this report, the run module instructions were described for steps that specify the definition of the filter binning intervals of the detector. The rise and fall of the line 27 signal corresponds precisely to this programmed set of 4 reading intervals. From this it appears that line 27 acts as a gating signal, determining whether bins from the horizontal register are shunted into the current drains (when low) or into the chip's on-board amplifier (when high).

The programmed rationale of the ABI data acquisition software and firmware specifies ten intervals over the 522 pixels of the horizontal well array, alternately shunting each interval pair into the amplifier or drain. While line 27 is high or low, the oscillations of line 24 drive single pixel shifts per pulse.

The falling edge of the signal on line 27 apparently triggers clamp and hold output of the chip's amplifier, to provide an analog signal voltage on output line 12, corresponding to the most recently binned and amplified set of horizontal wells. A/D conversions take place on the ABI 377 motherboard, presumably within short lag intervals following the falling edge trigger of line 27 and appearance of the held voltage on line 12. At 4X mode operation, there must be 5 reads per pixel times 194 pixel per scanning second**. This corresponds to an average rate of 970 A/D samples per second, as the maximum operating speed for the commercially delivered, data acquisition system of the sequencer.** This rate is very small with respect to the 50,000 to 100,000 A/D samples per second which the MC68332 processor hardware should be able to support.

We are now undertaking a surrogate approach to data acquisition with the CCD-detector of the ABI 377. Our emerging strategy is to set the instrument into nominal 4X operating mode, except for the run module instructions programming the camera. External hardware will then be used to program the timing/counting and A/D conversions of the camera.

The control interface is under development, using a LabView shell, which coordinates with the National Instruments Nu-Bus boards to be used for the data acquisition.

We are currently using a second workstation (Quadra 700 or PowerMac 8100), with two data acquisition cards from National Instruments:

NB-MIO-16XL
NB-DMA-2800

These provide multiple analog and digital i/o channels, 16 bit A/D conversion (of line 12 inputs), with timbase selection up to 5 MHz for controlling the asynchronous waveforms required on analog output lines 24 - 27. The DMA board provides CPU-indpendent (non-interupt) direct streaming of digitized data to the workstation memory.

A separate CCD-detector and analog data cable have been purchased from ABI to support the development and evaluation of the software off-line from the instrument.

part # 603437	ASSY, CCD DETECTOR	$ 3,965.00
part #603637	PCA, CCD Termination Flex	$ 1,861.00

Approximately 3-fold acceleration of the line 24 pulse array should be manageable under a timebase of 1 or 5 Mhz. This would support lateral resolution at approximately 600 pixels per 1 second scan of the standard 6 inch window. Each readout of the chip would have to be performed on a cycle of about 1.7 milliseconds. Average pixel size under such conditions would be 0.01 inch or about 250 micron. For perspective in this discussion, the focused spot of laser illumination is said to be 100 microns in the plane of sequencing gel.

The National Instruments hardware described above can perform A/D conversions at a maximum rate of 18 µsec per sample. This sets an upper bound for sampling 94 bins off of output line 12 per readout cycle. It thus seems reasonable that we may be able to increase the spectral bandwidth of the detector system from 4 filters per pixel (5 reads) to 64 filters, allowing some additional time for asserting control through the programming of lines 24, 25 and 27. Some of the signal to noise loss from sampling over shorter intervals could be recovered from multiple A/D reads per bin, at the proportionate time-cost of numbers of bins read per pixel. However implemented, the prospects are very positive for driving the detector of the ABI 377 to higher lateral resolution and spectral bandwidth, with minimum alteration of the native hardware.

Coordination of the fluorescence measurements with the lateral position of the scanner is apparently determined in native operations by reference to an optical encoder that is mounted on the brushless d.c. motor that drives the transport stage. There are three optical switches (2 at left, 1 at right), mounted on the solid platform below the transport stage. The left inner switch specifies a home position for the left edge of a platen mounted below the optical stage. The outer switches appear to be extreme translation limit detectors; they aren’t closely approached during actual scanning. We will tap the terminals of the motor’s encoder, in order to investigate a subornation of its signals to coordinate our surrogate fluorescence measurements. Otherwise we may install an aligned pair of optical sensors to create autonomous triggers for the initiation of left and right traverses of the scanner stage.