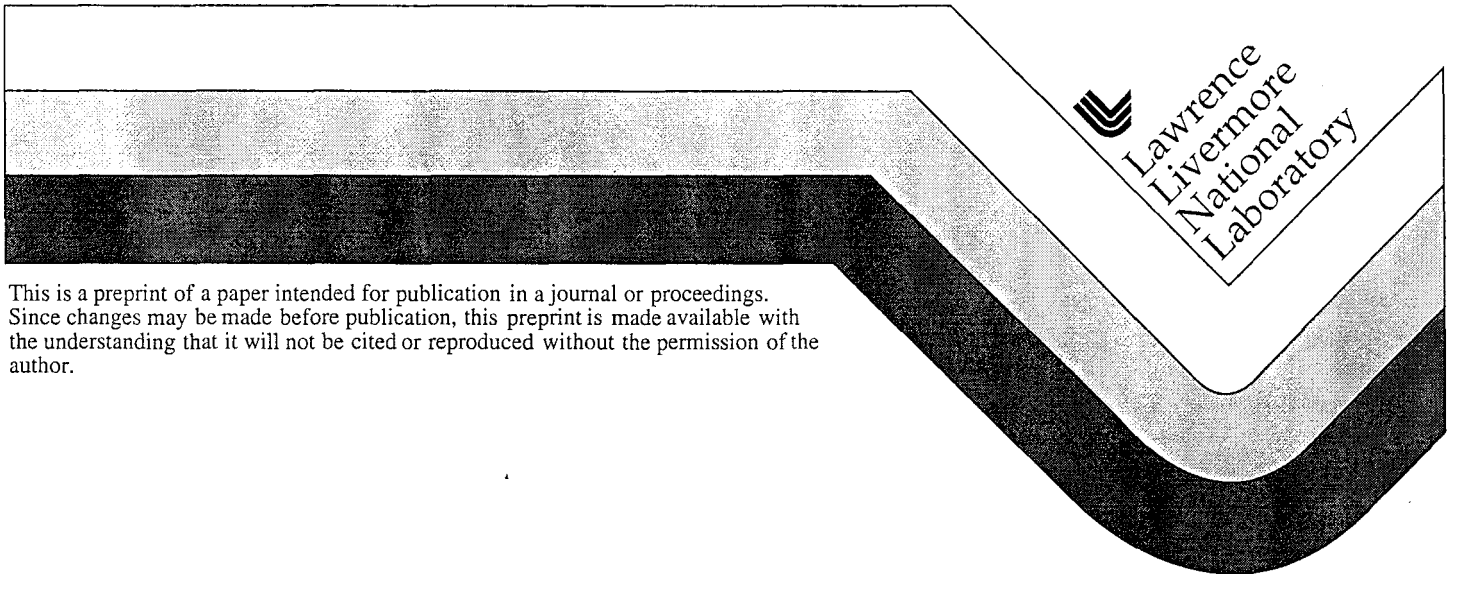
UCRL-JC-133132
PREPRINT

A Microchannel Electrophoresis DNA Sequencing System

J.P. Fitch, J.W. Balch
M. Bass, L.R. Brewer
A.C. Copeland, J.C. Davidson
L.M. Kegelmeyer, J.R. Kimbrough
R.S. Madabhushi, P. McCready
D. Nelson, R.L. Pastrone,
P.M. Richardson, S. P. Swierkowski,
L.A. Tarte, M. Vainer, T. Warth

This paper was prepared for submittal to the
33rd Annual Conference on Information Sciences and Systems
Baltimore, MD
March 17-19, 1999

January 1999



A Microchannel Electrophoresis DNA Sequencing System

J. P. Fitch, J. W. Balch, M. Bass[1], L. R. Brewer, A. C. Copeland, J. C. Davidson, L. M. Kegelmeyer, J. R. Kimbrough, R. S. Madabhushi, P. McCready, D. Nelson, R. L. Pastrone, P. M. Richardson, S. P. Swierkowski, L. A. Tarte, M. Vainer, T. Warth[2]
The Department of Energy Joint Genome Institute
Lawrence Livermore National Laboratory, MS L-452, Livermore, CA, 94550
[1]University of California, San Francisco and [2]University of California, Santa Barbara

SUMMARY OF PROPOSED REGULAR PAPER

In order to increase the DNA sequencing throughput of the Joint Genome Institute, we have developed a microchannel electrophoresis system. The critical new and unique elements of this system include 1) a process for the production of arrays of 96 and 384 microchannels on bonded glass substrates up to 14 x 58 cm and 2) new sieving media for high resolution and high speed separations. With custom fabrication apparatus, microchannels are etched in a borosilicate substrate, and then fusion bonded to a top substrate 1.1 mm thick that has access holes formed in it. SEM examination shows a typical microchannel to be 40 micrometers deep x 180 micrometers wide by 46 cm long. This technology offers significant advantages over discrete capillaries or conventional slab-gel approaches.High throughput DNA sequencing with over 550 base pairs resolution has been achieved in roughly half the time of conventional sequencers.

In February 1999, we begin a pre-production evaluation protocol for the microchannel and for three glass capillary electrophoresis systems (two from industry and one developed by Lawrence Berkeley National Laboratory for the Joint Genome Institute). In order to utilize these instruments for DNA production sequencing, we have been evaluating and implementing software to convert raw electropherograms into called DNA bases with an associated probability of error. Our original intent was to utilize the DNA base calling software known as Plan and Phred developed by the University of Washington. This software has been outstanding for our slab gel electrophoresis systems currently in the production facility. In our tests and evaluations of this software applied to microchannel data, we observed that the electropherograms are of a different statistical and underlying signal structure compared to slab gels. Even with substantial modifications to the software, base calling performance was not satisfactory for the microchannel data.

In this paper, we will present

- The microchannel DNA sequencing system and show the advantages compared to current slab gel and capillary systems.
- The signal processing modules needed including correction of multiple wavelength channels, signal averaging, non-uniform sampling, variable DNA mobility, and peak shape and spreading effects.
- A comparison of the DNA base signatures in the raw data of microchannels vs. slab gels including some simple modeling results. This will be propagated through the base calling software to show the impact on DNA sequencing.

Keywords: DNA sequencing, electrophoresis, signal processing

A Microchannel Electrophoresis DNA Sequencing System

J. P. Fitch, J. W. Balch, M. Bass[SF], L. R. Brewer, A. C. Copeland, J. C. Davidson, L. M. Kegelmeyer, J. R. Kimbrough, R. S. Madabhushi, P. McCready, D. Nelson, R. L. Pastrone, P. M. Richardson, S. P. Swierkowski, L. A. Tarte, M. Vainer, T. Warth[SB]

The Department of Energy Joint Genome Institute
Lawrence Livermore National Laboratory, MS L-452, Livermore, CA, 94550
[SF]University of California, San Francisco and [SB] University of California, Santa Barbara

ABSTRACT OF PROPOSED REGULAR PAPER

Last year, the Joint Genome Institute was the third leading sequencing group in the world with over 20 million bases of human DNA finished. The goal of the human genome project is to sequence the 3 billion bases of human DNA sequence by 2003. In order to increase the DNA sequencing throughput of the Joint Genome Institute, we have developed a microchannel electrophoresis system. One critical new and unique element of this system is a process for the production of arrays of 96 and 384 microchannels on bonded glass substrates up to 14 x 58 cm. In order to utilize these instruments for DNA production sequencing, we have been evaluating and implementing software to convert raw electropherograms into called DNA bases with an associated probability of error. Our original intent was to utilize the DNA base calling software known as Plan and Phred developed by the University of Washington. In our tests and evaluations of this software applied to microchannel data, we observed that the electropherograms are of a different statistical and underlying signal structure compared to slab gels. Even with substantial modifications to the software, base calling performance was not satisfactory for the microchannel data. In this paper, we will present 1) the microchannel DNA sequencing system and show the advantages compared to current slab gel and capillary systems. 2) The signal processing modules needed for DNA base calling including correction of multiple wavelength channels, signal averaging, non-uniform sampling, variable DNA mobility, and peak shape and spreading effects. 3) A comparison of the DNA base signatures in the raw data of microchannels vs. slab gels including some simple modeling results. This will be propagated through the base calling software to show the impact on DNA sequencing.

Keywords: DNA sequencing, electrophoresis, signal processing

1. INTRODUCTION

Electrophoresis may be defined as the differential drift velocity of electrically charged chemical ion or molecular species under the force of an electric field in a fluidic separation medium. Pioneering work in 1937 by Tiselius on electrophoretic protein separation in large tubes resulted in a Nobel Prize. Polyacrylamide and agarose sieving media was introduced to reduce resolution losses from diffusion and convection and are still used in the popular slab gel DNA sequencers. Large improvements in separation performance were introduced with very narrow (75 μm) bore fused silica capillaries in the 1980s. Early work with microchannels etched in flat glass substrates[2-4], versus drawn glass capillaries, indicated the potential of chemical separation done with the methodology of micromachining. Our microchannel arrays are fabricated in two borosilicate float glass substrates that are fusion bonded together into one microchannel plate. This plate may be cleaned or regenerated simply by pumping chemicals and sieving media into a single slot manifold port on the output end of the plate. The borosilicate float glass is rugged (both mechanical and thermal) and chemically resistant.

Microchannel electrophoresis arrays share many of the benefits of capillary arrays such as: short separation times, no lane tracking required, sieving media pumped directly into the array, and resolution of over 500 DNA bases in less than 3 hours of run time. Furthermore, our microchannel arrays have several additional positive attributes:

- The integrated array of microchannels in a single plate is very robust and dense.
- Only two components are fabricated and bonded to make a plate. Replicas can be made economically.
- Sieving media fill and replenishment as well as plate cleaning is done with a single O-ring sealed port.

- A planar optical window and precision placement of channels and fiducials enable better laser fluorescence detection.
- The channel is shape is wide and shallow—optimizing signal detection and sieving media heat dissipation.
- Designs can be changed with CAD tools and the fabrication method can handle diverse structures such as single ended loads, cross-channel loads, and fiducials or electrodes in the channels. Precise fiducials (1 μm) fabricated with the channels can also be used for alignment to the rest of the system.

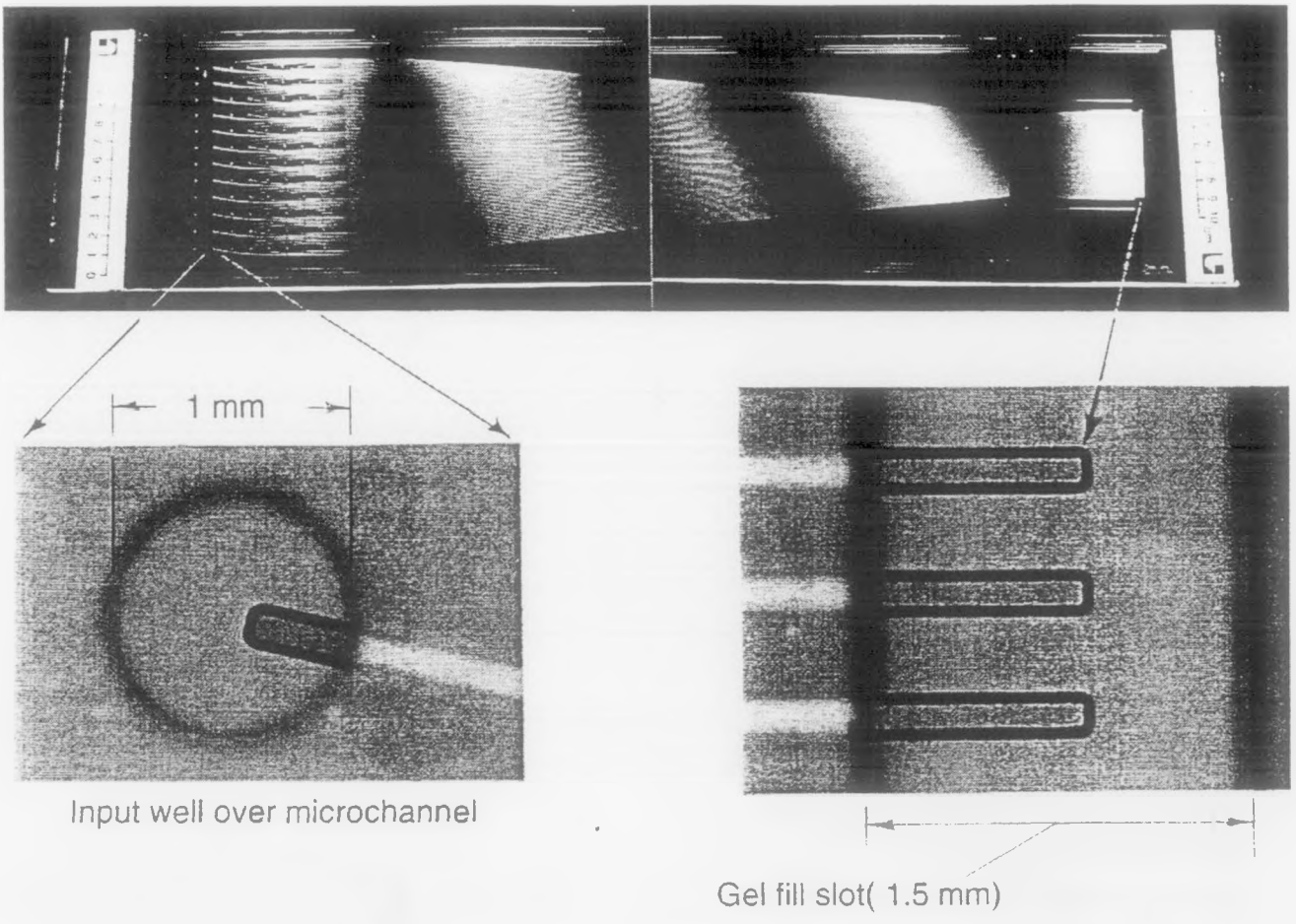


Figure 1. A 96-microchannel glass plate with microtiter input format on the left and the detection zone and fill slot on the right. The titer plate input port format makes each row of channels 9 mm different in length. The longest channels are 418 mm to the optical readout; the shortest ones are 63 mm shorter. The microchannels are about 45 x 170 μm in cross section.

2. RESULTS

Microchannel plates containing 96 channels (see Figure 1) have been used extensively over the past six months. One plate has been used more than 130 times; generating nearly 500,000 high quality base calls from various test samples. Depending on the purity of the samples we load into the system, we can use the plate 15 to 30 times before performance begins to degrade due to fouling of the channel walls. When this happens, the channels can be easily regenerated in situ by flushing with solvent followed by an acid etch and water rinse.

Samples are introduced into the channels through holes in the top plate. A small volume syringe is used to transfer the 100-500 nL required for analysis. We typically use commercially available gel loaders that contain 8 or 12 syringes for loading the plate. The sample input holes are arranged in an 8 x 12 array on 9 mm

centers corresponding to the industry standard microtiter plate. A 96-syringe device to load samples directly from a microtiter plate is a logical extension of our current technique. After loading the wells, DNA is then electrokinetically loaded onto the microchannels. Finally, the DNA drifts and separates under the influence of an applied electric field of 100 - 150 V/cm. Typical run times are 180 minutes, which allow us to collect data from over 500 peaks per lane.

A lane map for a 96-microchannel plate is shown in Figure 2. The arrival time of the primer peaks shows a sawtooth pattern because of the different drift lengths, for the different input rows. The built in fiducial pattern is shown on the sides of the figure as two constant lanes on the left and three constant lanes on the right; they are etched with the same pitch as the microchannels. A sample of a raw electropherogram from a lane somewhere in the middle of the plate is shown in Figures 3 and 4. One can clearly see the peak width and separation and good signal to noise ratio even before smoothing and filtering.

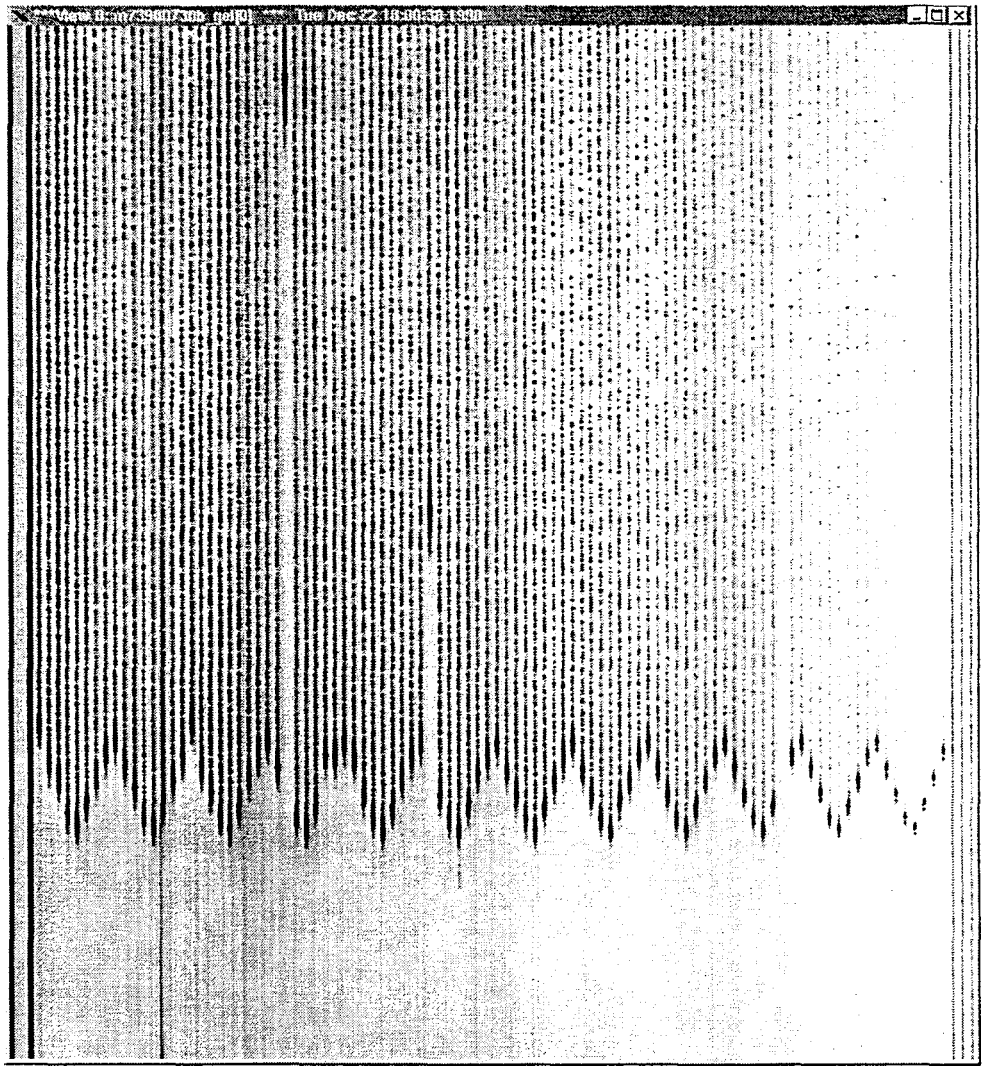


Figure 2. Lane map showing all 96 microchannel lanes for the blue dye that labels the DNA base C. Scanning across the plate every second, the displayed data is collected in 2.5 hours (time origin at bottom), with an applied voltage of 4.8 kV. The sample is an energy transfer dye labeled M13 sequencing standard. Note the uniformity of separation across the microchannels. Three lanes in this example show anomalous behavior due to accidental load variation but produced good data on subsequent runs.

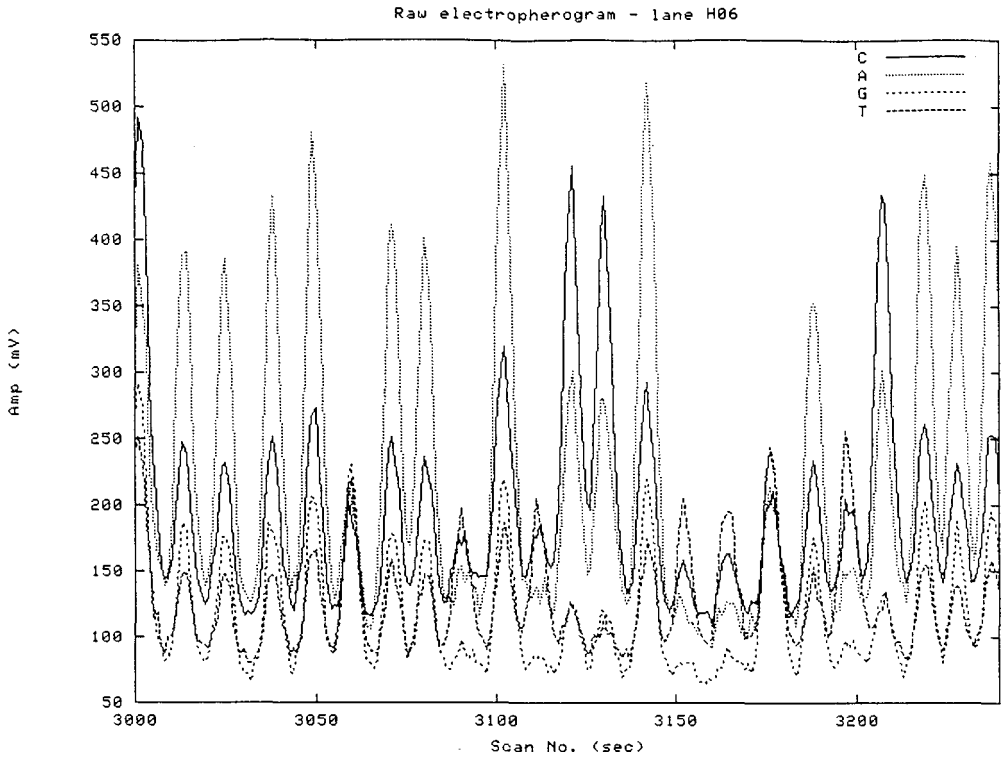


Figure 3. Raw data from lane H06 corresponding approximately to base numbers 50 through 75.

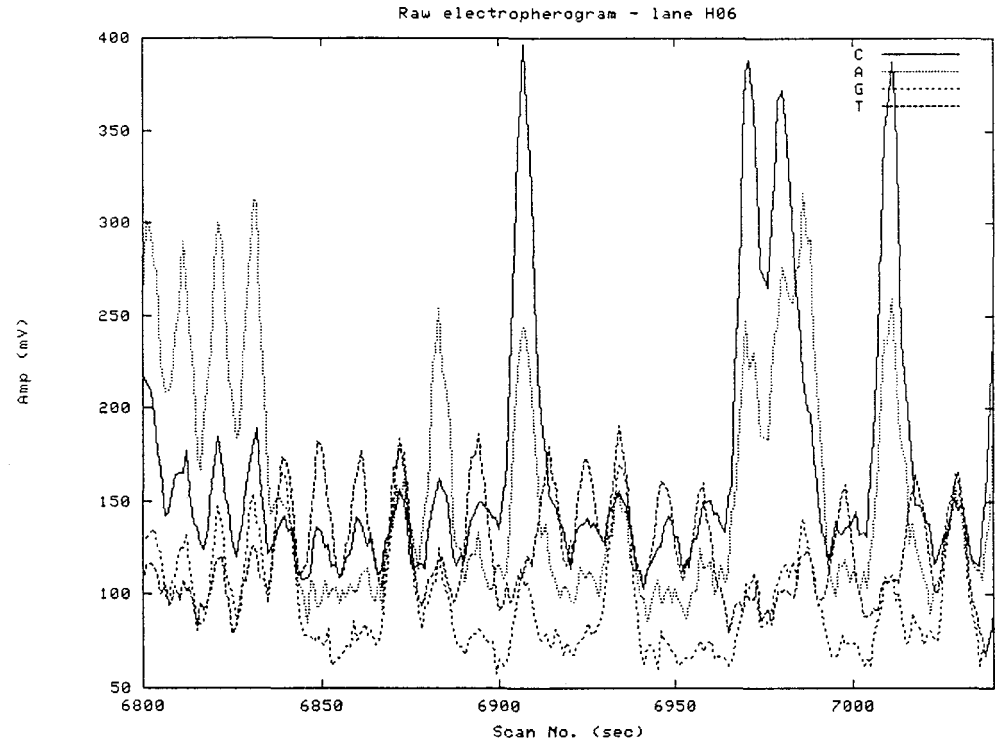


Figure 4. Raw data from lane H06 corresponding approximately to base numbers 400 to 425. This particular M13 standard was resolved to greater than 550 bases.

3. SIGNAL PROCESSING

Software modifications focused on the fundamental differences between conventional instruments and the microchann instrument:

- By replacing the baseline correction algorithms with a more adaptable and robust morphology-based algorithm, the noisy microchannel peaks retained their resolution, while losing most of the overlying high frequency noise. The morphological approach is non-linear and allows dynamic shape-based computations throughout the data trace.
- We have developed tools to generate dye-molecule mobility correction to account for the unique media/dye combination and channels of different length. We are now able to generate customized compensation tables for various types of lanes and obtain improved performance. These were applied separately to long and short lanes, with performance results shown in the chart below.
- Because the microchannel plate has lanes that vary in length, the electric field and thus velocity of the specimer vary as a function of position along each channel. As a result, the single look-up-table in plan/phred that correct: for non-uniform signal spacing along the length of a lane were not flexible enough to accommodate the range of lar lengths in the microchannel data, and we implemented a dynamic spacing correction for microchannel lanes.
- Lastly, we found that the plan/phred software does not trim the beginning of the signal adequately for microchannel data and so we trimmed the beginning of the sequences by about 100 bases before analysis.

The improvement realized by these software changes is illustrated in the performance chart in Figure 5.

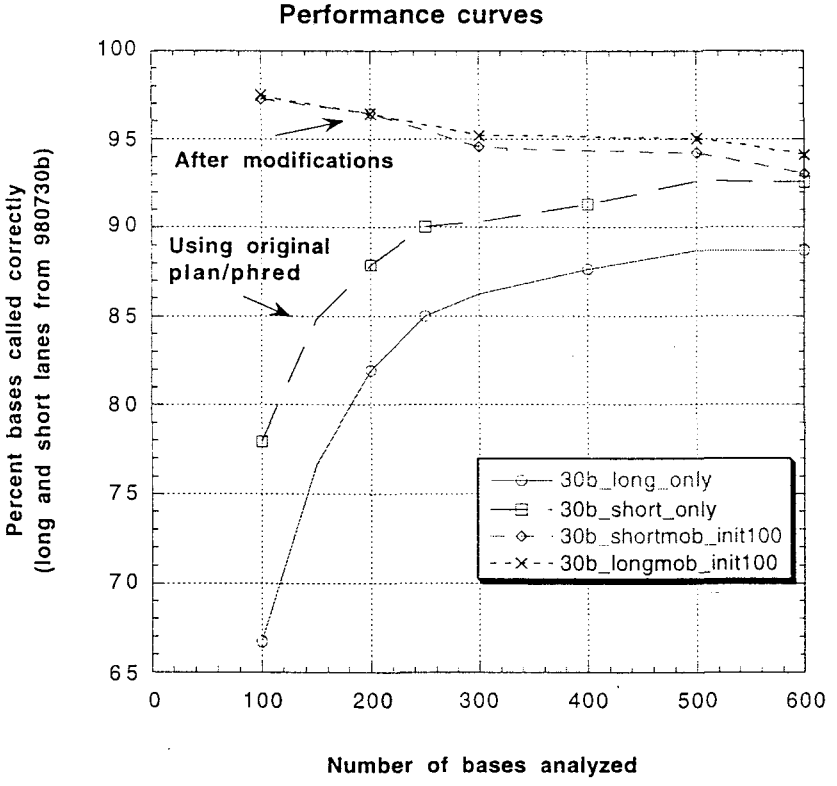


Figure 5. Performance improvement by modifying plan/phred software measured as an increase in the percent of correctly called bases from running known DNA samples in the microchannel sequencer. The 2 lower curves represent results from the original plan/phred software; the improved performance shown by the 2 upper curves, are the results of the software modifications described in this paper.

4. CONCLUSIONS

We have developed a microchannel-based electrophoresis system for high throughput DNA sequencing. Design and fabrication procedures have been demonstrated for high density microfluidic arrays in borosilicate glass up to 58 cm in size that are easy to reproduce, inexpensive, and physically robust. We have demonstrated DNA sequencing with up to 550 base resolution in a 96-microchannel plate-based system that has convenient in situ recycling and regeneration. Similar plates with 384 channels per plate have just been fabricated. Since the microchannel shape and size is very comparable to the 96-channel plate, these plates should provide comparable base-calling quality at four times the throughput. This type of high throughput system is needed for the JGI DNA sequencing goals.

ACKNOWLEDGMENTS

The authors acknowledge the support and funding provided by the NIH and DOE programs in human genomics as well as the LLNL LDRD project for DNA Base Calling.

REFERENCES

1. E. Carrilho, M. Ruiz-Martinez, J. Berka, I Smirnov, W. Goetzinger, A. Miller, D. Brady, B. Karger, *Anal. Chem.*, 68, pp. 3305-3313, 1996.
2. A. Manz et. al. , *J. Chromatogr.* 593, p. 253, 1992.
3. D. J. Harrison, et. al., *Anal. Chem.*, 64, p. 1926, 1992.
4. D. J. Harrison, et. al., "Micromachining a Miniaturized Capillary Electrophoresis-Based Chemical Analysis System on a Chip", *Science* 261, pp. 895-897, August 13, 1993.
5. BOROFLOAT™, Schott Corporation, Technical Glass Division, 3 Odell Plaza, Yonkers, N.Y. 10701.
6. Ultrasonic milling and mandrel by Bullen Ultrasonics, Inc., 4613 Camden Road, Eaton, Ohio 45320.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.