Ludmila Alaverdian[1]
Sinok Alaverdian[1]
Olga Bilenko[2]
Iouri Bogdanov[1]
Elena Filippova[*]
Dmitry Gavrilov
Boris Gorbovitski[2]
Michael Gouzman[1]
Georgy Gudkov[1]
Sergey Domratchev[1]
Olga Kosobokova[1]
Nadia Lifshitz[2]
Serge Luryi[1]
Victor Ruskovoloshin[1]
Andrew Stepoukhovitch[1]
Marina Tcherevishnick[1]
Georgy Tyshko[1]
Vera Gorfinkel[1]

[1]Department of Electrical
 and Computer Engineering,
 State University of New York,
 Stony Brook, NY, USA
[2]BioPhotonics Corporation,
 Stony Brook, NY, USA

A family of novel DNA sequencing instruments based on single-photon detection

We have developed a family of high-performance capillary DNA sequencing instruments based on a novel multicolor fluorescent detection technology. This technology is based on two technical innovations: the multilaser excitation of fluorescence of labeled DNA fragments and the "color-blind" single-photon detection of modulated fluorescence. Our machines employ modern digital and broadband techniques that are essential for achieving superior instrument performance. We discuss the design and testing results for several versions of the automated single lane DNA sequencers, as well as our approach to scaling up to multilane instruments.

1 Introduction

Capillary electrophoresis is widely used for the high-throughput DNA sequencing. Modern commercial sequencing machines employ a multicapillary format (ABI PRIZM-3100, CEQ-2000, ABI PRIZM-3700, Mega-BACE). The highest number of lanes (384) is offered by the MegaBACE-4000 sequencer. At the same time, several groups develop new high-throughput sequencing machines based on integrated-system technologies and prepare to overcome the 1000 channel barrier [1–10]. The U.C. Berkeley group led by Mathies [5] has developed a 1000 capillary instrument where the capillaries are positioned in grooves made in the outer surface of a metal cylinder. The rotating excitation (Ar-ion laser) and the detection objective are placed inside the cylinder. The excited fluorescence is detected by a confocal detection system, split into four channels by dichroic mirrors and measured by four photomultipliers. The University of Alberta group led by Dovichi [9] employs the off-capillary

sheath flow detection combined with a side illumination scheme. In the 16-channel rendition of this instrument, the optical detection is carried out with 16 fibers terminated with 16 GRIN lenses that focus the individual capillary images into fibers. The fibers are connected to 16 cooled ($-20°C$) avalanche photodiodes. Recently, a German group from the Max Planck Institute reported the development of a 96-lane instrument [10]. The 96 capillaries are subdivided into six arrays of sixteen. Scanning is avoided by transforming the radiation from an Ar-ion laser into a linear beam by guiding it through a line generator. The intensity of this beam is sufficient to uniformly illuminate all six arrays. A set of lenses and mirrors in conjunction with a CCD camera is used for detection.

Miniaturization of multichannel devices using microfabricated capillary arrays is another important direction in the development of DNA sequencers. A group at MIT led by Ehrlich [11–13] has developed several generations of microfabricated arrays on a glass substrate. Their devices have a rectilinear structure and employ an efficient cross-heart injection. Recently they reported a read length of 800 bp [13]. The Berkeley group [14, 15] has also developed several design modifications for microphoretic devices, starting from a rectilinear several-channel layout and recently moving to a higher density 96-channel radial layout [16] employing a four-color rotary confocal scanning system for fluorescent detection. The research group headed by Yeung at Ames Lab of Iowa State University [17] has developed integrated systems comprising a sequencing reactor, a purification column and a DNA separation carrier. Recently, they reported a new system with the minimal sample size of only 120 nL [18].

By their technical principle, the fluorescence detection systems used in current DNA sequencing instruments are very similar to those developed in the mid 1980s–1990s. They

Correspondence: Dr. Vera Gorfinkel, Department of Electrical and Computer Engineering, SUNY Stony Brook, Stony Brook, NY 11794-2350, USA
E-mail: vera@ece.sunysb.edu
Fax: +631-751-2457

* Present address: Brookhaven National Laboratory, P.O. Box 5000, Upton, NY, USA

excite fluorescence markers using a laser source and then capture fluorescence with an analog photo detector (either of photomultiplier tube (PMT) or CCD type). This signal is then digitized, transferred to a computer, and analyzed to determine the sequence. There are two major technical shortcomings common to all four-color machines: the relatively low detection sensitivity and the relatively narrow dynamic range, limited by the analog-to-digital conversion circuitry. Much higher detection sensitivity and much greater dynamic range are needed to substantially reduce the sequencing costs and improve the data quality.

A number of extremely sensitive fluorescence detection techniques are available based on registering single photons. Commonly, they are referred to as the photon correlation techniques. Until very recently single-photon counting techniques were mostly used for specialized scientific applications, such as the detection of single fluorescent molecules [19–23]. Though very sensitive, these techniques are not widely used in commercial DNA sequencing. The two primary obstacles preventing the wider use of these techniques are their complexity and the high cost per channel. A typical measurement setup for single-photon detection includes complicated devices such as fast time-to-analog converters, synchronization systems, costly analog-to-digital converters, and rather slow and expensive systems for computer recording of the detected data. Existing commercial photon counting devices are not so expensive (*e.g.*, the SR-400 Photon Counter from Stanford Research costs $ 5500, the PMT RS-232 Interface 7205 from REEVE Analytical costs $ 2500 and so on). The major shortcoming of these devices is their low data transfer rate, which is prohibitive for recording DNA sequencing runs. Thus, both an improvement in performance and a reduction of cost are needed in order that single-photon detection techniques could become of wide use in the large-scale DNA sequencing.

In this paper, we describe our development of a novel family of capillary DNA sequencing instruments carried out during the last five years at SUNY SB and Biophotonics Corporation and supported by the NIH (NHGRI, NCI). Our goal has been the implementation of capillary DNA sequencing instruments with ultrahigh sensitivity, large dynamic range, and a convenient modular architecture. The enhanced performance of our instruments is based on two technical innovations: the multilaser excitation of fluorescence of labeled DNA fragments and the color-blind detection based on the single-photon counting technique. We shall describe the design and the results of performance testing for several versions of ultrasensitive single-capillary DNA sequencing instruments employing single-photon detection. We shall also discuss our approach to extending the same operational principles to multilane instruments.

2 Materials and methods

2.1 Novel fluorescence detection technology

2.1.1 Fluorescence detection technique with multilaser modulated excitation

Fluorescence detection technique based on multilaser modulated excitation [24] permits the detection and the quantification of the relative concentration of fluorophores with different excitation spectra. The illumination source comprises simultaneous radiation of several lasers, each tuned to the preferred excitation wavelength of an individual fluorophore (Fig. 1a). Suppose we have N fluorophores and use N lasers for exciting fluorescence. Each laser is characterized by a well-defined wavelength λ_i and an output power $P_i(t)$ modulated with a distinguishable (*e.g.*, orthogonal) temporal code. Under the combined illumination by all lasers, the m-th fluorophore produces a fluorescence response F_m. Since the fluorophore excitation spectra may be significantly wider than the separation between the laser wavelengths, the response of any individual fluorophore may contain up to N distinct temporal components excited by those lasers whose spectra overlap with the excitation spectrum of the particular fluorophore. Amplitudes of these components will be proportional to the illumination power $P_i(t)$, the efficiency of excitation of a particular m-fluorophore by a particular i-laser, ε_{mi}, and the concentration of the fluorophore n_m, *viz.*

$$F_m(t) = \sum_{i=1}^{N} \varepsilon_{mi} \times P_i(t) \times n_m \tag{1}$$

The combined fluorescent response received by a single color-blind photodetector will be:

$$F(t) = \sum_{m=1}^{N} F_m(t) = \sum_{m=1}^{N} \left[\sum_{i=1}^{N} \varepsilon_{mi} \times P_i(t) \times n_m \right] =$$

$$= \sum_{m=1}^{N} \left[P_i(t) \times \sum_{m=1}^{N} \varepsilon_{mi} \times n_m \right] = \sum_{i=1}^{N} F_i(t) \tag{2}$$

where F_i is the fraction of the total fluorescence excited by the i-laser:

$$F_i(t) = P_i(t) \times \sum_{m=1}^{N} \varepsilon_{mi} \times n_m \tag{3}$$

The orthogonal components of $F_i(t)$ can be easily extracted from the total fluorescence response $F(t)$ measured by a color-blind photodetector. Writing Eq. (3) for all $1 \leq i \leq N$, we obtain a system of N linear equations with N variables $n_m (1 \leq m \leq N)$. This system can be solved, provided we know all elements of the excitation cross-talk matrix ε_{mi}. In the case when P_i are periodic functions of time (harmonic modulation), Eq. (3) can be rewritten and solved for the Fourier amplitudes.
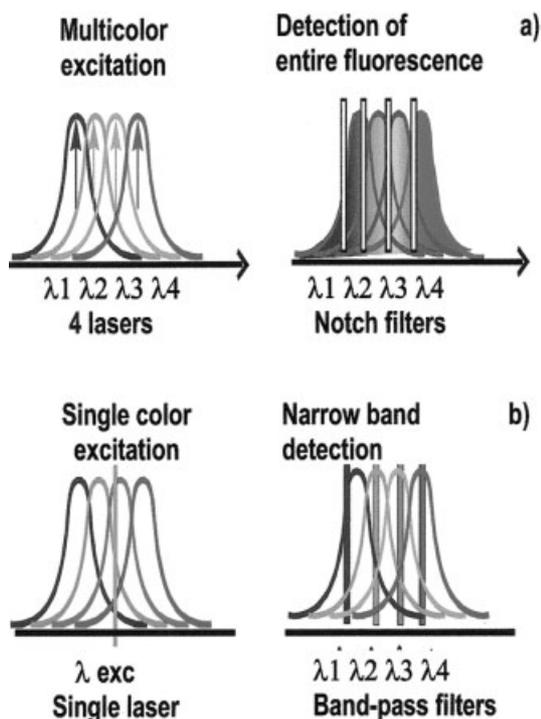
Nucleic acids

Figure 1. Comparison of multilaser excitation and multicolor detection schemes.

The multilaser excitation/color-blind detection scheme has several key advantages over the commonly used single laser excitation and color-resolved detection (see Fig. 1). Firstly, the multicolor illumination scheme allows optimization of the laser/fluorophore combination, which enhances the fluorescence emission by the fluorophore. Secondly, the color-blind photodetector in this scheme collects the entire fluorescence rather than just narrow spectral bands. Thus, our approach is more sensitive both because it generates larger fluorescence signal and captures more fluorescence. Still another advantage arises due to the modulated character of fluorescence signal. Color separation based on Fourier analysis filters out the nonmodulated noise, leaving only the modulated laser-induced signal. This has the effect of significantly improving the signal-to-noise ratio.

It should be emphasized that the advantages of the modulation multicolor technique are fully realized when there is minimum overlap between the excitation spectra of detected fluorophores. In our experience, using two-laser illumination in conjunction with two dyes whose spectra were separated by 70 nm, we could easily detect a 1% admixture of distinct DNA material marked with one dye on the background of 99% host material marked with the other dye.

2.1.2 Single-photon detection technique for multicolor modulated excitation

In contrast to an analog photodetector where the amplitude of the electric response signal is proportional to the incident light intensity, a single-photon detector generates a series of short electric pulses at a rate proportional to the intensity of the incident light. Therefore, as a response to amplitude-modulated photon flux a single-photon detector will produce a "frequency-modulated" train of single electric pulses. Consider a mixture of two fluorophores illuminated simultaneously by two lasers emitting at wavelengths λ_1 and λ_2 with the laser output power modulated at radio frequencies f_1 and f_2, respectively (modulated signals from two lasers and their combined signal are shown in Fig. 2a). One of the two fluorophores preferentially absorbs radiation from laser 1 modulated at frequency f_1, while the other preferentially absorbs radiation from laser 2 modulated at frequency f_2. Illuminated fluorescent molecules exist either in their
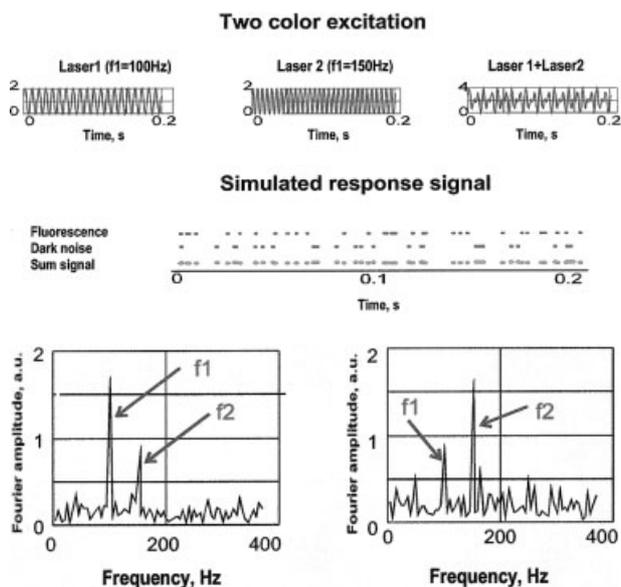

Figure 2. (a) Two-color modulated excitation and fluorescent response of two fluorophores with absorption coefficients depending on the excitation wavelength simulated by the Monte-Carlo technique. An excitation signal is produced by two lasers modulated at 100 Hz and 150 Hz (upper panel). Lower panel shows the simulated fluorescence response (blue trace) and dark noise (red trace). Average count rates of the fluorescent response and dark noise are chosen approximately equal. (b) Fourier spectra of the response signal simulated for two fluorophores with absorption coefficients depending on the excitation wavelength. Fluorophore 1 responds better to the signal of the laser 1 modulated at the frequency f_1 (left picture). Fluorophore 2 preferentially absorbs radiation from the laser 2 and its response at the frequency f_2 is higher (right picture).

ground or excited states. Excitation and fluorescence processes correspond to transitions between these states. The excitation of molecules occurs due to the absorption of photons emitted by the two lasers. The excitation rate is proportional to the laser output power. After excitation, fluorescent molecules stay in the excited state for some time and then spontaneously emit fluorescent photons.

To assess the theoretical sensitivity limit of our methodology, we used the Monte Carlo technique [25] to simulate the time intervals spent by a fluorescence molecule in each of the two states. Each transition from the excited state to the ground state causes emission of one fluorescent photon. As a result of the simulation, we obtain a train of single photons distributed in time according to the modulation of the laser output power. A simulated fluorescent response for detecting two fluorophores using two lasers is shown in Fig. 2a. While noisy photons are randomly distributed on the time axis, photons corresponding to the fluorescent response of fluorophores have a temporal distribution correlated with the incident modulation of the laser source. Fourier spectra of the fluorescent response of the traces obtained for two fluorophores with different absorption coefficients are shown in Fig. 2b. The difference in absorption causes a difference in the Fourier amplitudes of the fluorescent response of the two fluorophores and enables their identification as described in the preceding section. Results of our simulations show that for the noise level corresponding to the dark count of a typical single-photon PMT (100–300 c/s) receiving 50–100 fluorescence photons per identified color enables faithful identification of multiple fluorescent markers with a signal-to-noise ratio higher than 5 (see Fig. 2b).

2.2 Design and implementation

2.2.1 General design principles

In this section, we shall discuss the basic design concepts common to all our sequencers. All our instruments employ the single-photon detection of modulated fluorescence. Our sequencers consist of three main modules shown in Fig. 3. These modules are aligned at the assembly stage and then connected *via* standard optical fibers and electric cables. Thus, our instruments do not require any field adjustments. Such a design makes instruments compact, easy to configure, use, assemble, and repair.
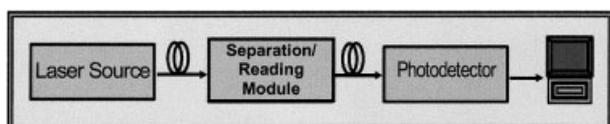
Figure 3. Architecture of Stony Brook sequencers
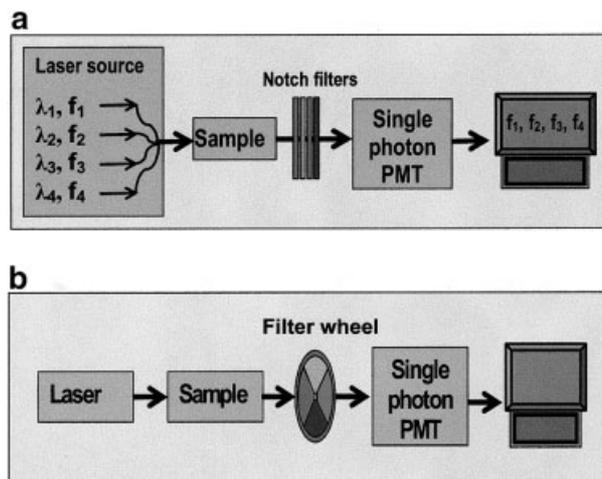
Figure 4. (a) Multilaser excitation scheme. Light from each of the four lasers is modulated with unique radio-frequency. Laser emission wavelengths are chosen to excite preferentially one dye of the four-dye set. The light from all four lasers is coupled in one optical fiber and delivered to the detection zone where it excites fluorescence of the labeled sample. The entire fluorescence is collected and impinged onto single photodetector through a set of four notch filters eliminating reflected and scattered laser light. The signal from the photodetector is asynchronously recorded by a PC, and fluorescence response from individual dyes is calculated using Fourier analysis. Single-laser excitation scheme. All four fluorescence markers in the sample are excited by a single laser. The fluorescence light goes through the rotating filter wheel with four narrow band-pass filters separating four fluorescence components and then received by a single photodetector. The recording is synchronized with the wheel revolutions. The FPGA counter integrates photon counts while the light is passing through each of the four filters, thus obtaining signal magnitude in four channels and transfers the count to the computer during each revolution of the wheel ($\sim$ 100 revolutions a second).

Our DNA sequencing instruments can be divided into two categories according to the method used for dye excitation and data acquisition (see Fig. 4). In both categories the fluorescence detection is carried out using one color-blind single-photon detector (PMT, Hamamatsu) in conjunction with a specially designed digital circuitry. Instruments of the first category (Fig. 4a) employ multilaser excitation with time-modulated laser intensities (see next section for details). In instruments belonging to the second category (Fig. 4b), all four dyes are excited by a single laser emitting a single continuous wave (CW) power. A rotating filter wheel with four narrow band-pass filters is used to separate the four different fluorescence components. Data recording is synchronized with rotations of the wheel. A specially designed electronic circuit integrates the data from the photodetector while the light is

passing through each of the four filters, thus obtaining the signal magnitude in all four channels. Comparison of the optical losses in the two device configurations in Fig. 4 shows that they are approximately equal. Indeed, the color separation and filtering optics in the multilaser configuration reduce the optical efficiency to ~20% (stack of four notch filters with 0.8 transparency and a 50% loss due to the use of only AC component for recognition of fluorescence markers $(0.8^4 \times 0.5 \cdot 0.2)$), which is very close to the 15% efficiency for the single laser excitation (reduction by a factor of 5 due to the filter wheel and the nonideal transparency of the color filter and the additional ~80% transparency of the laser rejection filter).

Figure 5 shows a general view of our single-channel desktop sequencer (Model SBS-2000). A labeled DNA sample undergoes separation in a single-capillary fiberized separation/reading module (left photo). The separated zones arriving at the detection window in the capillary are illuminated by a laser light delivered from a miniature fiberized laser source (right lower photo). The fluorescence from labeled DNA fragments is collected by a fiber receiver and delivered to the photodetection module (right lower

photo). After appropriate filtering the fluorescence is impinged onto a single-photon PMT. Electric pulses produced by the PMT are counted by a fast photon counter and transferred to a computer for recording and processing. During the separation process, sequencing traces are displayed in real time on the computer screen. After completion of the sequencing run, sequencing data processing can be performed either on the same computer or on any computer connected to the lab network. Computation time needed for the data processing for a typical sequencing run is about 3–5 min including the Fourier analysis by Fast Fourier Transform (FFT) technique.

2.2.2 Laser sources

We have implemented and compared several different illumination modules. Four-color modules are equipped with a set of four fiberized lasers whose modulated light is coupled to a single 62 μm fiber and delivered to the illumination window. Laser sources can be further classified by their spectral range. Our excitation module designed for ABI dyes comprises four lasers emitting at 488, 514,
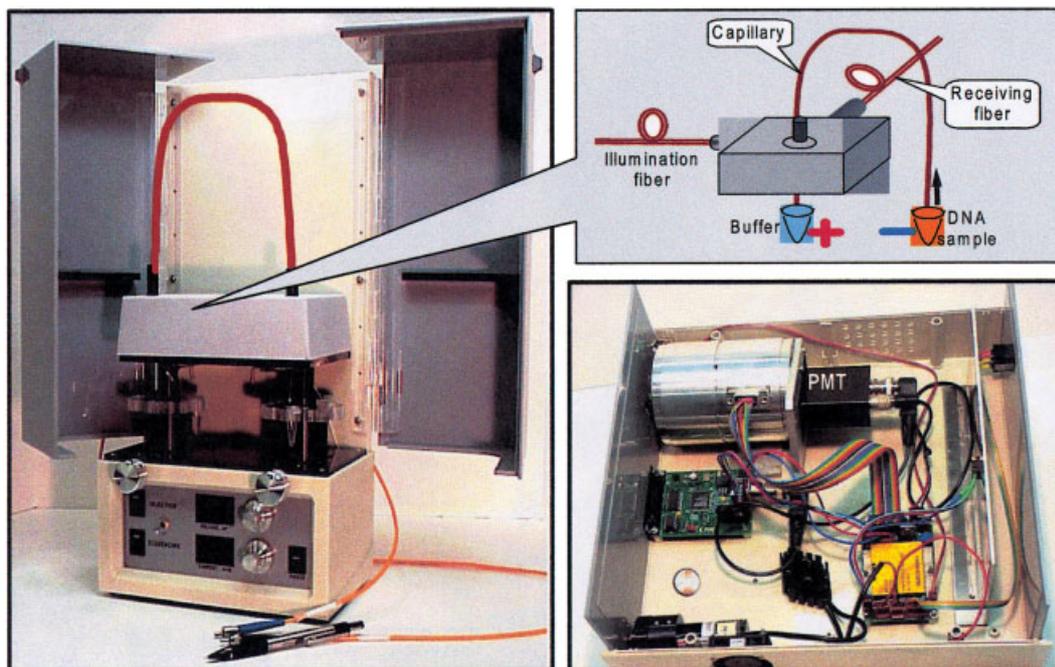

Figure 5. Stony Brook sequencer SBS-2000. Separation/reading module (left photo). Tube-loading carousels carry DNA sample tubes and running buffers for inlet and outlet tubes. The length of the capillary can vary from 30 to 60 cm. The electrokinetic injection is carried out at 0.5–3 kV, and then the running voltage (8–15 kV) is applied to the capillary. At the detection end the capillary is inserted into the reading head (see insert). Excitation light is delivered to the capillary *via* 62 μm illumination fiber from a miniature fiberized 532 nm 5 mW Nd-YAG laser built into illumination/photoreceiving module (right lower photo). Excited fluorescence is collected by a 200 μm fiber-receiver and delivered to the photon-isolated photoreceiving module where it is filtered, detected by a single-photon PMT, counted by a fast FPGA counter and transferred to a computer.

532 and 594 nm each, equipped with an electronic chopper modulating the lasers' output power in the range of 100–500 Hz. Our other module, designed for the Beckman-Coulter sequencing kit, is equipped with four miniature semiconductor laser diodes emitting in the 635–810 nm spectral range and modulated by current drivers in the range of 1–2 kHz. All lasers operate in the low-power regime below 500 μW per color.

The illumination module for our single-laser excitation instruments is equipped with only one fiberized laser working in the CW regime. Since our reading head has a fiberized input for the excitation light and uses achromatic optics, the laser wavelength can be selected and changed according to the user's need without any misalignment of the reading head. We have successfully tested instruments with 488, 514 and 532 nm lasers at the output power levels varied from 200 μW to 40 mW.

2.3 Separation/reading module

This module (see Fig. 5) comprises a miniature high-voltage supply (up to 15 kV) with a built-in voltmeter and a micro-ammeter, tube-changer carousels carrying tubes with DNA samples and running buffers for capillary inlet and outlet, a capillary holder for capillaries of ~325 μm OD and 50 μm–150 μm ID, and a fiberized reading head which can be used in conjunction with any of the above laser sources. The reading head design is shown in Fig. 6. The reading head is a metal fixture perforated with two orthogonal cylindrical bores that house both the detection and the excitation objective holders. Also attached to the fixture is a plastic capillary holder that rigidly fixes the position of the capillary window at the intersection of the excitation and the detection axes. The body of the fixture is traversed with several cuts in all three dimensions, with screw holes cut through the slits. These slits and screws are used for the initial alignment of the optical head. At the assembly stage the optical head is adjusted to maximize the strength of the fluorescent signal and minimize the background due to the reflected and scattered laser. Our mechanical design preserves optical alignment during capillary replacement. We found that the reading head does not get misaligned even after one year of instrument exploitation. The radiation from the laser source delivered *via* a fiber-illuminator and focused by a miniature grin lens built into the reading head pierces the capillary and illuminates a narrow (~50 μm) zone inside the capillary (Fig. 6, right). Labeled DNA fragments moving along the capillary fluoresce while passing the cross-section of the laser beam. Their fluorescence is collected under 100° body angle by a miniature built-in optical objective, coupled into a 200 μm fiber-receiver, and delivered to the photodetection module.
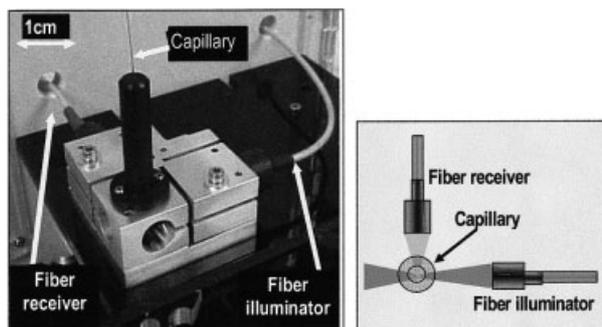

Figure 6. General view of the reading head (left) and cross section of the reading zone (right). The capillary with analyzed DNA sample is inserted into special precision fixture. Special mechanical design of the fixture preserves optical alignment during the capillary replacement. Radiation from a laser source delivered *via* 62 μm fiber-illuminator excites fluorescence of labeled DNA fragments migrating in the capillary. The excited fluorescence is collected by 200 μm fiber receiver and delivered to the photodetection module.

2.4 Photo-receiving modules

We have developed several types of photoreceiving modules corresponding to the variety of illumination modules. Each module consists of a fluorescence filtering system and a fluorescence detection system. Due to the extremely high sensitivity of single-photon detectors, special precaution was taken to isolate the photoreceiving modules from the ambient light. Each photoreceiving module comprises several compartments that are assembled in a "light-tight" fashion (see Fig. 7).

Design of the fluorescence filtering system depends on the type of the illumination module. For four-laser modules, the filtering is carried out with a set of four 10 nm 6 OD super-notch filters rejecting four laser wavelengths (Keiser Optics Inc.). For single-laser illumination modules, the fluorescent signal purified from the laser light with
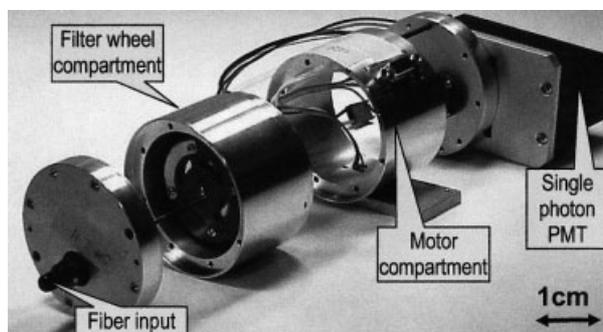

Figure 7. Fiberized photodetection module providing isolation from ambient light.

either notch or step filter undergoes time-division multiplexing, with a specially designed rotating filter wheel with four 10–20 nm, 3–4 OD band-pass filters (OMEGA Optics) corresponding to the emission maxima of the four used dyes.

Fluorescence detection system employs one single-photon detector. For the implementation of this detector we have used two types of single-photon PMTs (both available from Hamamatsu). The first type (Hamamatsu H7464, H6240) has a built-in pulse amplifier with the output pulse width of 35 and 70 ns, respectively. The second type (Hamamatsu R5600P) does not comprise an amplifier. For this device we have designed and built novel fast pulse amplifiers with the output pulse width of less than 2 ns. In order to handle such short pulses, we have designed and implemented fast digital counting circuits based on field-programming gate array (FPGA) technology. Our FPGA counters are able to count pulses of 2 ns and less pulse width without any pulse loss. Thus, maximum photon flux which can be detected by our photodetection system is as high as 100 million photons per second with pulse loss do not exceeding 20%, which can be easily compensated by two orders of magnitude higher than that of the best commercially available photon counters. From the FPGA counter the signal is transferred to the computer where it is recorded and processed.

Data transfer: We have developed several efficient circuits and software packages that enable different implementations of the detection and registration of single photons. For modulation-based methods we use an asynchronous data transfer protocol capable of transferring to the computer and recording the information about every detected photon at the time of its reception with the rate as high as 10^6 photons/s and the precision of 10^{-6} s. For laser sources with low modulation frequency (below 1 kHz) we employ an initial integration on the counter board, so as to transfer 32-bit signals rather than single pulses.

For single-laser excitation instruments with a rotating filter wheel, the recording is synchronized with the wheel revolutions. An FPGA-based photon counting circuit board is used to integrate the data from the single-photon counter while the light is passing through each of the four filters, thus obtaining the signal amplitudes in all four channels. These four values are transferred to the computer during each revolution of the wheel (up to 100 revolutions a second). Further processing of the fluorescent signal in the computer depends on the type of the illumination module used. For the four-laser illumination module, the color recognition is carried out by Fourier analysis. For the single-laser illumination module, the colors are resolved by time division multiplexing.

3 Results and discussion

3.1 Tests of sensitivity and dynamic range

3.1.1 Sensitivity tests

The high sensitivity of the single-photon PMT detectors is primarily due to their very high internal gain ($\sim 10^6$) combined with an extremely low dark count (~ 100 c/s at room temperature). In addition, since the electric output of a single-photon PMT can be processed by a digital circuitry, signal recording and processing steps do not add any noise to the detected signal. The only unavoidable noise in a single-photon detector is the noise associated with the stochastic nature of the measured photon fluxes. This is proportional to the square root of the measured signal. Therefore, the minimum noise of a single-photon detector is the noise associated with its dark count. In our photo receiving modules having dark count in the range of 100–300 c/s, this noise varies from 10 to ~ 20 c/s, which is negligible. Additional noise, which lowers the S/N ratio of the entire system compared to that of the photo receiver alone, is largely due to fluctuations of the laser illumination. This noise can be efficiently reduced by purifying the detected signal from the residual laser radiation and by reducing a background signal mostly caused by the fluorescence of the separation medium inside the capillary.

To quantify the sensitivity achieved in our instrument we prepared serial dilutions of Rhodamine-590 fluorescent dye in distilled water and pumped them through a 75 μm ID capillary inserted into the instrument's reading head. We started by pumping pure distilled water in order to establish a background level. Then we proceeded with the dye dilutions, steadily decreasing the concentration. Using a 532 nm 25 mW laser for excitation, we recorded dye fluorescence response in the same fashion as we do it routinely for DNA sequencing.

After the subtraction of a background level, the photon count corresponding to 10^{-10} M/L dilution was approximately 10^6 c/s. At this dilution, the concentration of fluorescence molecules is $n = 6 \times 10^{23} \times 10^{-10}$ molecule/L. The total number of molecules illuminated in the capillary by a 55 μm laser beam is $M = n \times V$, where V is the illuminated volume. In our experiments these values are $V \approx \pi/4 \times 55^2 \times 75$ μm^3 = 178 000 μm^3 and $M \approx 10\,000$ molecules. Thus, our instrument is able to detect about 10^6 photons per second from $\sim 10\,000$ fluorescence molecules, or about 100 photons per second per molecule.

In multiple experiments typical background level was about 50 000–100 000 c/s and the noise level N associated with the background was ~ 1000–1500 c/s which

is above the level predicted by the Poisson statistics. The additional noise is due to laser intensity fluctuations. Because the background level varied from one experiment to another, it was difficult to measure directly the fluorescence from, say, 100 dye molecules, the added fluorescence signal being too small compared to the variations of the background level. However, our instrument readily detected small groups of fluorescent molecules migrating in the capillary and producing 10 000 c/s peaks on the background of 100 000 c/s. The signal-to-noise ratio in these experiments was above 6 – in agreement with that expected from a group of 100 fluorophores each emitting at 100 c/s:

$$ S/N = \frac{100 \ (\text{molecules}) \times 100 \left(\dfrac{\text{counts}}{\text{sec} \ \times \ \text{molecule}} \right)}{1500 \dfrac{\text{counts}}{\text{sec}}} = 6.6 $$

To further decrease the minimum number of detectable molecules (say, down to 10) one needs to reduce the background level by at least 10-fold. This will require a better stabilization of the illumination power as well as more expensive notch and band-pass filters (with 6 OD on/off ratio instead of the 3 or 4 OD filters used in the present instrument).

3.1.2 Dynamic range tests and detection of mixed DNA samples

The dynamic range of single-photon counting systems is intrinsically very wide since the electric pulses produced by a PMT in response to single photons are very short (order of 1 ns). Therefore, the maximum count that can be measured by a PMT counter without a significant (less then 10%) count loss is of the order of 10^8 c/s. However, the commercially available PMTs with built-in amplifiers produce pulses of about 30–70 ns duration and therefore their linear range extends only up to $\sim 10^6$ c/s. Exceeding the PMT linear range may lead to an uneven distortion of signals from different channels and, thus, to a distortion of the color proportions in the cross-talk matrix. In order to increase the dynamic range of our detection system, we have designed and implemented a novel fast pulse amplifier which allows the registration of $\sim 10^8$ photons/s. We have also developed special software that increases the linear range of PMT detectors (and therefore the dynamic range) by $\sim 60\%$ (see Appendix 1). Defining dynamic range D of a photon counter as the ratio of its maximum count in the linear range to the minimum dark noise $D = S_{max}/N_{dark}$, we find that D of a single-photon counter depends on the integration time. For an integration time of 0.1 s, which is optimal for the processing of DNA sequencing data, the dynamic range of our single-photon

detector is $\sim 2.5 \times 10^6$. Compared to the one of the best cooled 16 bit CCD having minimum readout noise of ~ 3 bit per pixel per frame [26] at a rate of 10 frames/s, this is an improvement by more than 8 bits.

Results shown in Fig. 8 demonstrate the dynamic range capability of our instruments. Electropherograms in this figure were obtained using Promega GenePrint Power Plex 16 System samples. The samples contain 16 loci (fifteen STR and one Amelogenin) grouped into three groups each labeled with a different fluorescent marker. The green "JOE" labeled ladder (second panel) exhibits peaks with magnitude as low as 10^4 c/s. At the same time, the maximum peak amplitude on the Internal Lane Standard is greater than 1.6×10^6 c/s. Thus, the instrument allows the simultaneous accurate measurement of color signals, whose amplitudes differ by more than two orders of magnitude.

3.2 Test DNA sequencing runs

All test runs were carried out using the same single-capillary separation/detection module but different fiberized laser sources combined with the appropriate photoreceiving modules.

3.2.1 Four-laser excitation scheme

In order to test the detection technique based on the four-color modulated excitation we performed multiple sequencing runs using the Beckman-Coulter sequencing chemistry. A typical sequencing trace fragment for the Beckman-Coulter test sequence (CEQ DNA test sample, part 608070) is shown in Fig. 9 (raw data). For the fluorescence excitation we used a 4-color laser source comprising four semiconductor laser diodes. The output power of the four lasers (635, 675, 750 and 810 nm, each modulated by a distinguishable radio frequency in the range of 1–2 kHz) totaling about $\sim 400 \ \mu W$ was combined in one 62 μm fiber and directed into the capillary. A single-photon sensitive PMT was used for detection.

Run conditions: separation medium – Beckman CEQ separation polymer (CEQ TM Separation Gel-LPA I), 75 μm ID 50 cm long coated capillary from PolyMicro (see Appendix 2 for the capillary coating protocol), room temperature, 30 s injection at 3 kV, run voltage 8 kV. The obtained cross-talk matrix is: {A(1, 0.2, 0.1, 0.05), C(0.3, 1, 0.23, 0.05), G(0, 0, 1, 0.13), T(0, 0, 0.6, 1)}. Typically, we could distinguish peaks for DNA fragments as long as ~ 550–600 bases in a run time of ~ 2 h.
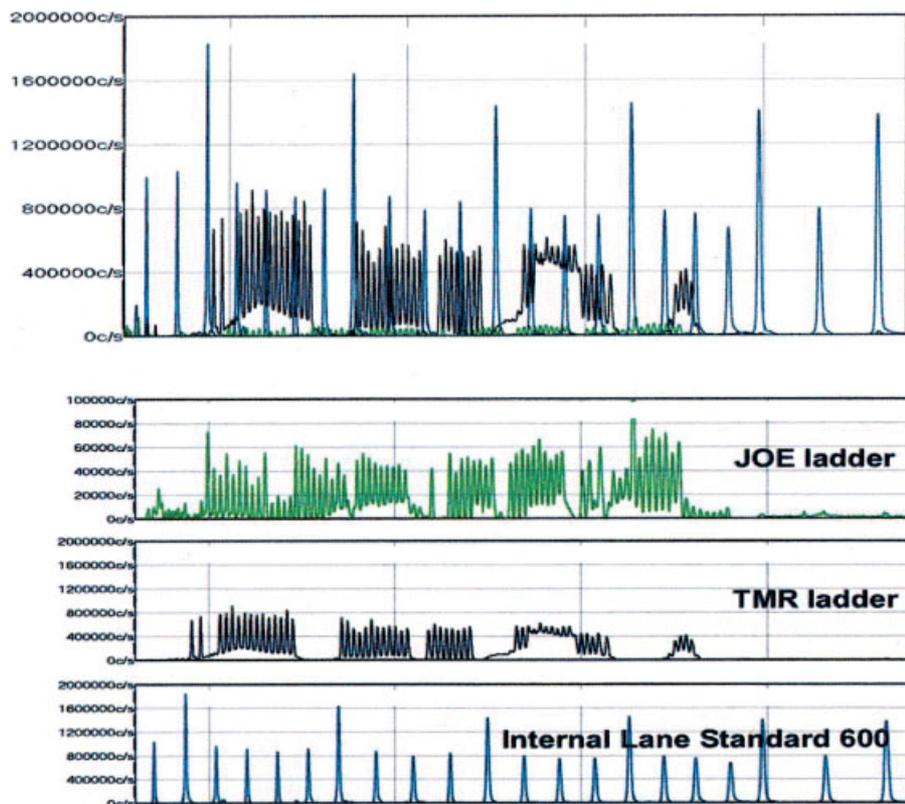
Figure 8. Allelic ladder mix (Promega, GenePrint PowerPlex 16). We used an allelic ladder mix to demonstrate the dynamic range of the device. The upper panel shows the electrophero-gram of the "JOE" and "TAMRA" ladders mixed with an Internal Lane Standard. Three lower panels show enlarged view of each ladder separately. Run conditions: separation medium POP-4; room temperature; excitation, 532 nm, 3 mW Nd-YAG laser; 3 s injection at 3 kV; run voltage, 15 kV, capillary length, 33 cm), run time, 30 min.
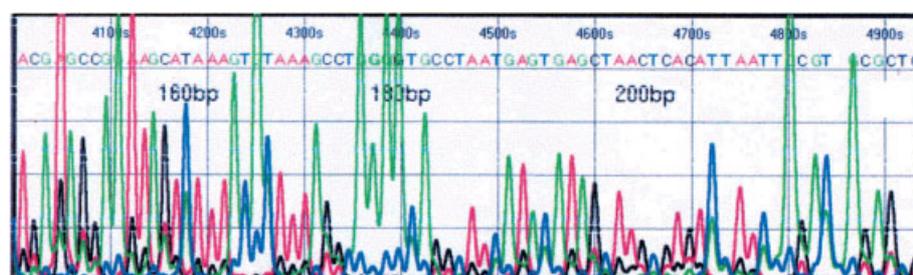

Figure 9. Fragment of Beck-man-Coulter test sequence, raw data (part 608070). All four fluorescence signals detected at each peak have been plotted in order to show a transfer matrix corresponding to the specific combination of lasers and filters installed in the instrument.

3.2.2 Single-laser excitation scheme

For testing the single-laser excitation scheme in conjunction with the single-photon counting detection we used various BigDye labeled DNA samples. In order to find the most efficient excitation wavelength we tested different single-laser sources. We found that for both dye sets the Nd-YAG laser ($\lambda = 532$ nm) provides more efficient excitation than Ar-ion lasers ($\lambda = 488$ nm and 514 nm) at the same output power. The obtained cross-talk matrix for BigDye labeled samples illuminated by Nd-YAG laser was: {A(0.22 0.83 0.58 1), G(0.07 0.44 1 0.36), T(0.38 1 0 005 0.1), C(1 0.3 0 0)}. A fragment of typical sequencing trace of DNA Controls/Standards Big DyeTMTerminator

Sequencing Standard (Part # 4304154) is shown in Fig. 10 (upper panel). The trace was acquired using a 20 mW Nd-YAG laser.

A read length typically obtained with BigDye labeled DNA samples in the Beckman polymer at room temperature, for 30 s/3 kV sample injection and 8 kV running voltage was as high as ~500–550 bases at 98.5% accuracy (using PHRED base calling software). Peaks could be distinguished for DNA fragments as long as ~650–700 bases in a run time of ~2.5 h. Sequencing experiments carried out in POP-5 polymer (ABI-3700 POP-5TM Performance Optimized Polymer, # 4313087) at room temperature in uncoated capillaries exhibited much shorter
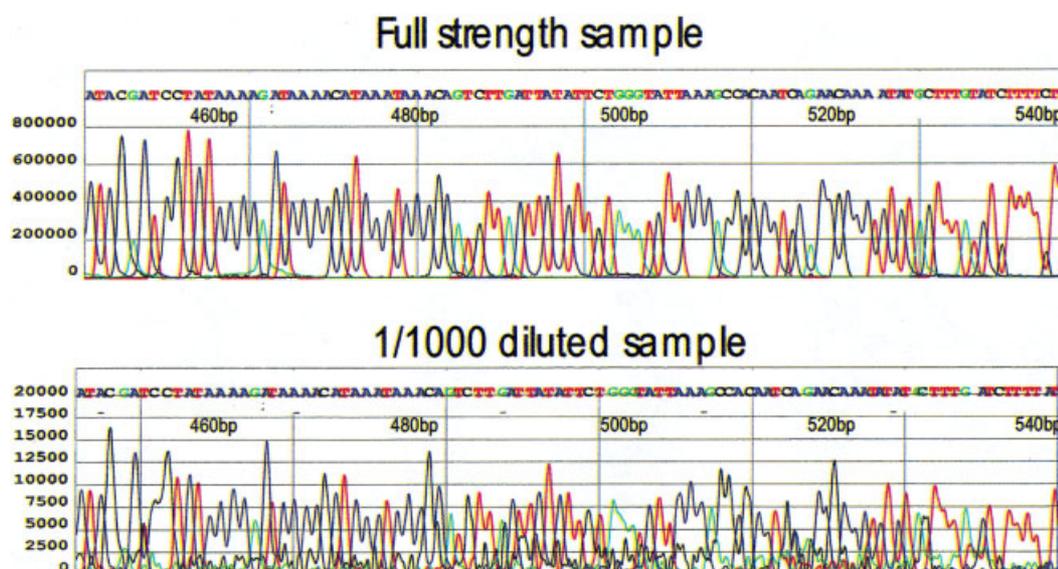
Figure 10. Fragments of sequencing runs of DNA Controls/Standards Big DyeTM Terminator Sequencing Standard (Part # 4304154). Full strength sample (upper panel) and 1:1000 sample diluted in water (lower panel). Run conditions: separation medium, Beckman CEQ separation polymer; 50 cm, 75 µm ID capillary from PolyMicro coated according to the protocol described in Appendix 2; room temperature; 30 s injection at 3 kV; run voltage, 8 kV; run time, 2.5 h.

read-length (~ 350–450 bp at 98.5% accuracy) mainly because of the reduced peak separation and faster deterioration of the peak height.

3.2.3 Sample dilution experiments

In order to determine the instrument sensitivity, we carried out a number of sequencing runs with various serially diluted DNA samples. We tested the ABI DNA sequence standards, as well as animal and plant DNA samples from the Cold Spring Harbor Laboratory (provided by Dr. W. R. McCombie) and clinical samples from the Memorial Sloan-Kettering Cancer Center (provided by Dr. N. A. Ellis).

Representative results of the experiments are shown in Fig. 10. Here we compare identical fragments of two sequencing runs of BigDye labeled DNA sequence standard (full strength sample and 1:1000 sample diluted in water). The tests were carried out on SBS-2000 equipped with a 20 mW, 532 nm Nd-YAG laser. Run conditions were identical for both the full strength and the diluted samples. In these runs the read length at 98.5% accuracy varied from 550 bases for the full strength samples to 510 bases for the 1/1000 diluted samples. Note, that the amplitudes of the peaks in the electropherograms do not scale with the dilution; while the content of the labeled material in the sample tubes differed by a factor of 1000, the amplitudes differed only by a factor of about sixty. We believe that this

is due to the enhancement of the injection efficiency caused by an increase of the electric field in the injection tube when highly resistive water is added. Our preliminary data indicate that the separation quality of diluted samples can be improved by optimizing the sample injection conditions. Moreover, we often observed significant improvement of the sequencing quality in highly diluted samples, probably because of the sample purification due to adding a large volume of distilled, deionized water.

In another example of enhanced sensitivity we used the dilution method in conjunction with primerwalking sequencing strategy to prove the capability of the SB sequencer of sequencing small genes using very low amounts of labeled DNA material. This work was done in collaboration with the Cold Spring Harbor Laboratory (Dr. W. R. McCombie's group). We reduced the amount of labeling material by a factor of 400 compared to the recommended in the ABI protocol and obtained a perfectly readable four-color sequence traces (to be published separately).

3.3 Extension to multilane sequencing instruments

Currently we are working on a 32-lane system based on the same detection technology and design principles as our single lane instruments (see Fig. 11a). Electrophoresis in the 32-channel instrument is carried out in a hybrid mul-
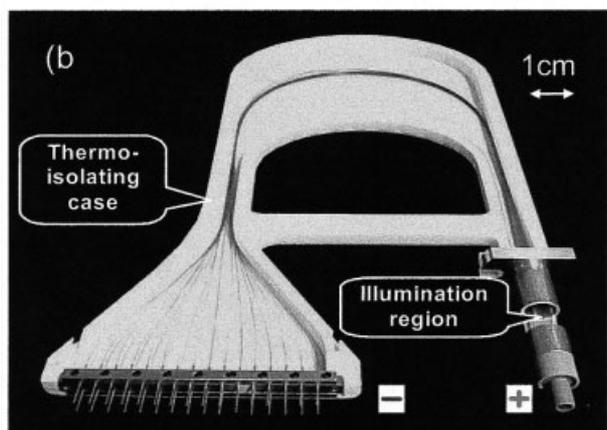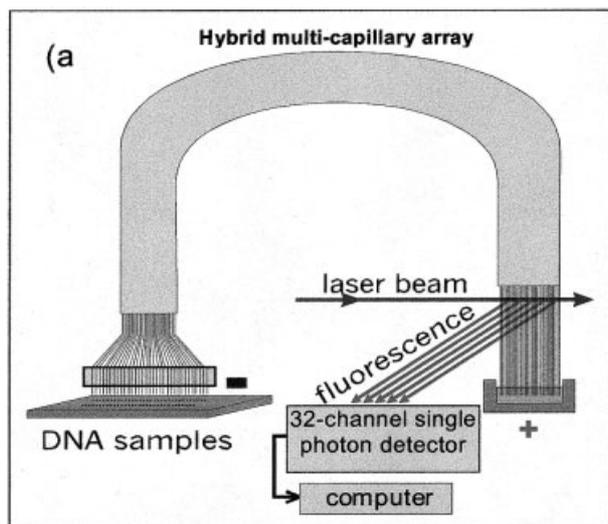
Figure 11. (a) Schematics of the 32-lane sequencing instrument and (b) photo of HMCA.

Figure 12. Quality factor, peak width and peak spacing curves for typical sequencing run.

ticapillary array (HMCA) (Fig. 11b) formed by 32 fused-silica capillaries (PolyMicro, 50–70 μm ID, 150 mm OD). The capillaries are assembled in a planar monolithic structure in the detection region and split apart in the DNA loading region. A special tool was developed for aligning the capillary array and turning it into a monolith. This tool removes the Kevlar coating from a segment of the capillary surface and positions and presses the capillaries to a V-grooved substrate to form a coplanar equidistant structure. This structure with the substrate is immersed into an optical compound with the refractive index equal to that of the capillary glass. The entire structure solidifies into a monolith. At the next stage the monolith is precisely oriented in a certain way in a metal holding fixture that is a part of the carrier. A special design of the holding fixture ensures its reproducible positioning in the optical reading head. The outlet end of the holding fixture is terminated by a polymer-loading HPLC type connector.
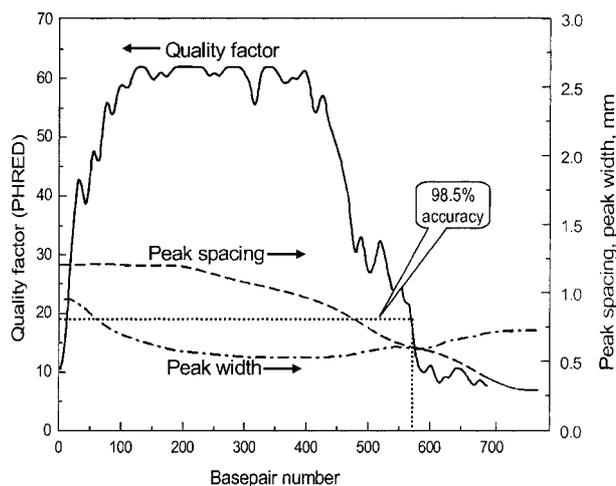
The inlet end of the HMCA is arranged into 8×4 or 16×2 arrays to conform to the standard 96- or 384-well plate. The HMCA is mounted on the special holder so that the split inlets of the individual capillaries are directed down to the 96- or 384-well tray with DNA samples. During the injection, the 32 capillaries are inserted into 32 separate loading wells on the DNA sample tray. During the sequencing run all 32 capillaries are immersed into a common reservoir with the running buffer.

At the detection end the laser beam delivered from the fiberized laser source illuminates the HMCA from the side. The monolithic nature of the HMCA detection zone ensures a practically uniform illumination of all capillary channels by a single mode laser. Fluorescence excited in all capillaries of the HMCA is delivered to a system of color filters and from there to a 32-channel single-photon sensitive PMT (Hamamatsu, H7260-P). Digital output from the PMT is sent to a 32-channel pulse amplifier and from there to a 32-channel fast FPGA photon counter (both of our home design) and further to a computer where the acquired data is displayed, recorded, and processed. The footprint of the instrument is $50 \times 10 \times 15$ inches. Our preliminary tests indicate that the 32-channel single-photon detection system has practically the same sensitivity and same dynamic range as those achieved in our single-lane instruments.

4 Concluding remarks

We have described the operational principles, as well as the design and performance of a family of novel DNA sequencing instruments based on two technical innovations in the field of excitation and detection of fluorescent

signals: the multilaser excitation of modulated fluorescence of labeled DNA fragments and the color-blind detection based on single-photon counting. Both innovations separately and independently contribute to a substantial sensitivity enhancement. Our sequencing instruments share a common modular architecture. Each instrument comprises several robust and compact modules connected with standard electronic cables and optical fibers. The modular architecture provides a tremendous flexibility in the design and the assembly of instruments for various applications, ranging vastly in their sensitivity, throughput, size and cost.

We have designed and implemented several configurations of single-capillary DNA sequencing instruments. Using BigDye sequencing chemistry and various DNA samples we have demonstrated two orders of magnitude improvements in both the sensitivity and the dynamic range, as compared to commercial instruments. Using the Beckman-Coulter polymer and BigDye sequencing chemistry we have achieved automatic read lengths exceeding 550 bp at 98.5% accuracy at room temperature. Visually, we distinguish peaks up to the fragment size of ~650–700 bases. Our detection technology and the instrument design are ideally suited to the implementation of automated high-throughput sequencing machines and systems. We believe that the high sensitivity and dynamic range of our instruments in conjunction with the emerging techniques for low-volume DNA sample preparation will lead to a very significant reduction of the sequencing cost. One of our important achievements is a completed compact affordable automated DNA sequencer that can be installed on a tabletop in a laboratory, connected to a PC and used by an individual researcher.

5 References

[1] Lu, X., Yeung, E. S., *Appl. Spectrosc.* 1995, *49*, 605–609.

[2] Wang, Y., Wallin, J. M., Ju, J., Sensabaugh, G. F., Mathies, R. A., *Electrophoresis* 1996, *17*, 1485–1490.

[3] Kheterpal, I., Scherer, J. R., Clark, S. M., Radhakrishnan, A., Ju, C. L., Ginther, G. F., Sensabaugh, R. A., Mathies, *Electrophoresis* 1996, *17*, 1852–1859.

[4] Wang, Y., Hung, S.-H., Linn, J. F., Steiner, G., Glazer, A. N., Sidransky, D., Mathies, R. A., *Electrophoresis* 1997, *18*, 1742–1749.

[5] Scherer, J. R., Kheterpal, I., Radhakrishnan, A., Ja, W. W., Mathies, R. A., *Electrophoresis* 1999, *20*, 1508–1517.

[6] Pang, H. M., Pavski, V., Yeung, E. S., *J. Biochem. Biophys. Methods* 1999, *41*, 121–132.

[7] Lu, S. X., Yeung, E. S., *J. Chromatogr. A* 1999, *853*, 359–369.

[8] Zhang, J. Z., Voss, K. O., Shaw, D. F., Roos, K. P., Lewis, D. F., Yan, J., Jiang, R., Ren, H., Hou, J. Y., Fang, J., Puyang, X., Ahmadzadeh, H., Dovichi, N. J., *Nucleic Acids Res.* 1999, *27*, E36–E43.

[9] John, H., Crabtree, S. J., Bay, D. F., Lewis, L. D., Coulson, G., Fitzpatrick, D. J., Harrison, S. L., Delinger, J. Z., Zhang, Dovichi, N. J., *Electrophoresis* 2000, *21*, 1329–1335.

[10] Behr, S., Matzig, M., Levin, A., Eickhoff, H., Heller, C., *Electrophoresis* 1999, *20*, 1492–1507.

[11] Schmaltzing, D., Adourian, A., Koutny, L., Ziaugra, L., Matsudaira, P., Ehrlich, D., *Anal. Chem.* 1998, *70*, 2303–2310.

[12] Salas-Solano, O., Schmaltzing, D., Koutny, L., Buonocore, A., Adourian, P., Matsudaira, P., Ehrlich, D., *Anal. Chem.* 2000, *72*, 3129–3137.

[13] Koutny, L., Schmaltzing, D., Salas-Solano, O., El-Difrawy, S., Adourian, A., Buonocore, S., McEwan, P., Matsudaira, P., Ehrlich, D., *Anal. Chem.* 2000, *72*, 3388–3391.

[14] Wolley, A. T., Mathies, R. A., *Anal. Chem.* 1995, *67*, 3676–3680.

[15] Liu, S., Shi, Y., Ja, W. W., Mathies, R. A., *Anal. Chem.* 1999, *71*, 566–573.

[16] Medintz, I. L., Paegel, B. M., Blazej, R. G., Emrich, C. A., Berti, L., Scherer, J. R., Mathies, R. A., *Electrophoresis* 2001, *22*, 3845–3856.

[17] Tan, H., Yeung, E. S., *Anal. Chem.* 1997, *69*, 664–674.

[18] He, Y., Zhang, Y. H., Yeung, E. S., *J. Chromatogr. A* 2001 *924,* 271–284.

[19] Dovichi, N. J., Martin, J. C., Jett, J. H., Trkula, M., Keller, R. A., *Anal. Chem.* 1984, *56*, 348–354.

[20] Chen, D. Y., Dovichi, N. J., *Anal. Chem.* 1996, *68*, 690–696.

[21] Soper, S. A., Davis, L. M., Shera, E. B., *J. Opt. Soc. Am. B* 1992, *9*, 1761–1769.

[22] Soper, S. A., Mattingly, Q. L., Vegunta, P., *Anal. Chem.* 1993, *6*, 740–747.

[23] Goodwin, P. M., Affleck, R. J., Ambros, W. P., Demos, J. N., Jett, J. H., Martin, J. C., Reha-Krantz, L. J., Semin, D. J., Schecker, J. A., Wu, M., Keller, R. A., *Exp. Techn. Phys.* 1995, *41*, 294–297.

[24] Gorfinkel, V. B., Luryi, S., *U.S. Patent 5,784,157* (issued July 1998).

[25] Rubenstein, R. Y., *Simulation and the Monte Carlo Method* J. Wiley and Sons, New York 1981.

[26] Spibey, C. A., Jackson, P., Herick, K., *Electrophoresis* 2001, *22*, 829–836.

[27] Bertolotti, M., in: Cummins, H. Z., Pike, E. R. (Eds.), *Photon Correlation and Light Beating Spectroscopy*, Plenum Press, New York 1974, pp. 41–74.

[28] Bédard, G., *Proc. Phys. Soc.* 1967, *90*, 131–141.

[29] Ewing, B., Hiller, L., Wendl, M. C., Green, P., *Genome Res.* 1998, *8*, 175–185.

[30] Ewing, B., Green, P., *Genome Res.* 1998, *8*,186–194.

[31] Savitzky, A., Golay, M. J. E., *Anal. Chem.* 1964, *36*, 8, 1627–1639.

[32] Yin, Z., Severin, J., Giddings, M., Huang, W., Westphall, M. S., Smith, L. M., *Electrophoresis* 1996, *17*, 1143–1150.

[33] Giddings, M. C., Severin, J., Westphall, M., Wu, J., Smith, L. M., *Genome Res.* 1998, *8*, 644–665.

Appendix 1

Linearity of a photon counter: The signal at the output of a single-photon counter (SPC) can be regarded as a sequence of short pulses. The photocount rate is defined as the number of pulses in this sequence during one second and is measured in counts-per-second (cps). In an ideal SPC with infinite temporal resolution, the count rate λ is proportional to the lightintensity I: $\lambda = \alpha I$. We assume that the number of pulses that arrive during a sampling period have a Poisson distribution.

For a nonideal SPC, the minimum time between two consecutive pulses that can be resolved is limited by the dead time τ_d of SPC. Statistical properties of the distribution of photocounts for a nonideal SPC are studied in [27, 28]. The limited temporal resolution leads to a reduction in the number of registered counts compared to the true number of photocounts that would be registered by an ideal SPC. For example, the registered count rate in a Hamamatsu H7467 PMT (which has a temporal resolution of $\tau_d = 70$ ns) shows about 10% deviation from the true count rate at a true count rate of $\lambda = 1.5 \times 10^6$ cps.

The fraction of lost counts increases with the increasing count rate. This results in a sublinear SPC response. Considering the application of SPC to DNA sequencing, even a small nonlinearity can produce significant distortions in the trace data processing. The true photocount rate is determined from the registered rate and the known temporal resolution of the device as

$$\lambda = \frac{\lambda_{reg}}{1 - \lambda_{reg}\tau_d}$$

This procedure is performed for each point of the recorded sequence separately.

Appendix 2

Capillary Coating Procedure: In all experiments we used fused silica capillaries from PolyMicro with 75 μm inner diameter. Below we describe our capillary coating protocol:

(i)   Rinse the capillary with methyl alcohol (5 mL).
(ii)   Rinse the capillary with distilled water (5 mL).
(iii)   Rinse the capillary with 1 M sodium hydroxide solution (5 mL) and leave the solution inside for 1 h.
(iv)   Rinse the capillary with distilled water (5 mL) and leave the solution inside for 1 h.
(v)   Mix 4 μL sodium 3-mercapto-1-propanesulfonate (MPS) and 1 mL of 6 M acetic acid. Fill the capillary with the solution and leave the solution inside for 1 h.
(vi)   Rinse the capillary with distilled water (5 mL).
(vii)   Prepare the coating solution: 4% acrylamide 0.1%, TEMED 0.2%, 95.7% APS water.
(viii) Fill the capillary with this solution immediately and leave the solution inside for 1 h.
(ix)   Rinse the capillary with distilled water (5 mL). Fill the capillary with Beckman-Coulter polymer.

Appendix 3

PHRED/PHRAP/CONSED is a widely used software package for processing and assembly of DNA sequencing traces. Algorithms used in PHRED for base-calling and quality values assignment are discussed in detail in [29] and [30]. Below we describe the preprocessing procedures used for preparing the SBS-2000 data in the form accepted by PHRED. The preprocessing includes noise filtering, peak smoothing, baseline removal, calculation of the Fourier amplitudes (for four-color excitation scheme), cross-talk filtering, mobility shift correction, data decimation and conversion to the SCF format. The preprocessing of data obtained from a typical sequencing run takes only several minutes including 1–2 min needed for the calculation of Fourier amplitudes by FFT.

The SBS-2000 produces 10–25 data samples per second resulting in 50–125 sampling points per peak of 5 s width. The recorded data set is first processed using a simple low-pass MA filter to reduce the noise due to random fluctuations in the single-photon counting data. The peak smoothing is accomplished by using a least-squares approximation of data by 3rd degree polynomials [31].

For baseline removal and cross-talk filtering we use the techniques described in [32]. The baseline is removed from each trace separately. Values of the piecewise-linear approximated baseline are subtracted from trace data. Cross-talk filtering follows baseline removal. The cross-talk matrix, calculated for each dye/machine combination, is stored in the configuration file. Options for either automatic or operator-supervised fine-tuning of the matrix for the processed data set are available.

The mobility shift correction used for SBS-2000 data processing is based on equalization of spacing between adjacent peaks. The method is similar to the algorithm described in [33] and uses nonlinear scaling of traces in the spatial domain to compensate for dye mobility differences. The algorithm requires neither prior calibration nor manual fitting of shift functions. First, the traces are linearly scaled in spatial domain to achieve an equal peak width at the beginning and at the end of the traces. Then, the peak detection algorithm is used to mark all well-

defined peaks. The mobility-shift correction procedure independently scales the traces trying to equalize the peak spacing while keeping the instant average peak spacing intact. The procedure scales the traces gradually in a number of steps starting from the middle of the run where the effect of mobility shift is minor and the resolution of sequencing is still good, and moves toward the beginning of the sequencing run where the mobility shift is the most severe. On each step, the estimate of instant peak is evaluated using a window of 11 peaks. A segment of shift function is chosen for each trace separately to minimize the average shift of peaks from their equidistant positions calculated over the window. Linear approximation of shift function segments is used. Applied to relatively small shifts on each step, the strategy of gradual spatial scaling allows compensation of a substantial mobility shift over the entire sequencing run. Further pro-

cessing steps include data decimation to the sampling rate oapproximately 10 data points per peak. The resulting file is converted to SCF v2.00 format.

Since in our sequencing experiments we used the ABI sequencing chemistry, we made base calling using the PHRED software trained for ABI-3700 machine. In order to check our data processing procedure we applied it to multiple sequencing runs with known sequences and found fairly good agreement with PHRED base-calling. We also compared the quality factor calculated by the PHRED software with the behavior of the peak width/ peak spacing curves. Typical graph is shown in Fig. 12. As can be seen from the graph, the base-calling quality rapidly decreases when the peak width and peak spacing become equal. The same behavior was consistently observed for all processed data files.