Dynamic analysis of capillary electrophoresis data using real-time neural networks

Boris Pokrić[a,*], Nigel M. Allinson[a], Edmund T. Bergström[b], David M. Goodall[b]

[a]UMIST, Department of Electrical Engineering and Electronics, P.O. Box 88, Manchester M60 1QD, UK
[b]Department of Chemistry, University of York, York YO10 5DD, UK

Abstract

A method for real-time processing of snapshot data obtained from a capillary electrophoresis instrument using laser-induced fluorescence detection is described. The detector is based on a charge coupled device imaging 25 mm of the capillary length. Output data from the camera consists of a series of snapshot images of the analyte bands moving across the capillary window. The snapshot images are first pre-processed in order to obtain spatial profiles of fluorescence intensity. Each profile from successive snapshot images is processed on-line using a neural network, allowing the real-time measurement of separation parameters such as the velocity and width for each peak. This approach enables dynamic information to be extracted during the separation and the reduction of storage requirements, since the profiles where there are no peaks need not be recorded. The signal processing is based on profile approximation using a radial basis neural network which provides accurate and sensitive peak determination for each camera snapshot. The high spatial resolution of the system and accurate fluorescence profile approximation leads to good peak shape description and therefore to improved deconvolution of overlapping peaks. Examples of real-time analysis of complex peak patterns in separations of sulfonated aluminium phthalocyanines are given. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Charge coupled device detection; Signal processing; Detection, electrophoresis; Laser-induced fluorescence detection; Neural networks; Aluminium phthalocyanine, sulfonated

1. Introduction

Over the last five to 10 years, capillary electrophoresis (CE) has proved to be a rapid and versatile separation technique for the analysis of water-soluble analytes [1,2]. It offers the benefits of high resolution, small sample volume, speed of analysis, automation and ease of method development. A free solution CE separation is based on the different electrophoretic mobilities of species due to differences in their charge-to-size ratios. In order to increase sensitivity, the single point detector zone length has been increased in commercial instruments up to 800 μm, and a detector cell is available with a path length of 1 mm [3]. However, both methods suffer from reduced spatial resolution. It is possible to maintain the high sensitivity without sacrificing spatial resolution by utilising charge coupled device (CCD) detector technology [4]. The CCD devices offer the benefits of low dark current, high quantum efficiency and good linear response with a high dynamic range [5], which makes them very suitable

for fluorescence detection [4,6–8]. Using this arrangement, the separation process can be monitored across a much wider capillary length (ranging from 1 to 70 mm), as reported using laser-induced fluorescence (LIF) detection [9–14].

The aim of this paper is to show that the spatial data can be processed in real-time, thus allowing the on-line measurement of sample mobility, concentration and peak width during an experimental run. The CE data processing is based on profile approximation using a radial basis neural network (RBNN). This type of neural network has been applied to a wide variety of problems such as image processing [15], speech recognition [16], adaptive equalisation [17] and medical diagnosis [18]. The strength of the RBNN is that it uses a single internal layer of locally-tuned radial basis functions and linear learning rules that ensures very fast convergence, which is crucial in real-time applications [19,20].

2. Capillary electrophoresis with CCD detectors

The CCD-based LIF system has been designed [21] so that the beam from a 0.95 mW He–Ne laser is focused by crossed cylindrical lenses on to a capillary length of ~25 mm. Fluorescence is collected orthogonal to the incident beam by an Astrocam 4201-TE/A CCD camera system. The camera is based on the EEV 05-20 CCD chip which has 770×1152 pixels (each is 22.5 μm square) with the readout noise of 6 electrons/pixel rms at 160 kHz readout rate. The dark current can be further reduced to 0.17 electrons/pixels/s by cooling the device to ~−40°C and operating the CCD in a multi-phase pinned mode [21]. A bandpass interference filter (Omega 685DF10) with 10 nm passband centred at $\lambda=685$ nm is positioned between the lenses.

The photo-induced charge in adjacent photosites is "binned" prior to read-out in order to increase the CCD frame rate [22], making an effective pixel dimension of 45 μm square. This binning, together with a slight magnification in the optics, result in a limiting spatial resolution along the capillary length of 50 μm. The pixel intensity values are digitised using a two-channel 16-bit A/D converter operating at 160 kHz. It takes 1.2 s to read the whole frame.

The system can operate in two different modes, time-delayed integration (TDI) and snapshot mode. The TDI mode is more sensitive [4,22], and produces much smaller output data files [21]. The system's limit of detection (LOD) in this mode is $7 \cdot 10^{-11}$ *M* for a 9-nl injection of Rhodamin 700. However, the additional information available using the snapshot mode can be of value at the expense of reduced sensitivity (~50-times higher LOD for Rhodamin 700). The true velocity of an analyte as it crosses the detection zone can be measured [4,21]. This gives an accurate measure of mobility which is free from effects such as sample stacking and transient isotachophoresis (ITP) conditions that can occur in the early stages of analysis [23]. The work presented here concentrates on the methods for real-time processing of the snapshot data so that the additional information available in this mode can be obtained on-line. Furthermore, the data storage requirement problem [4,21,24] can be greatly reduced.

The output from the CCD in the snapshot mode is a 16-bit resolution image which is pre-processed to obtain spatial profiles of fluorescence intensity. An example of a fluorescence profile obtained in LIF mode during the separation of Tinolux BBS is shown in Fig. 1. Tinolux BBS is a mixture of *n*-sulfonated aluminium phthalocyanine ($AlPcS_n$). Sulfonated aluminium phthalocyanines are water soluble fluorescent dyes that absorb strongly at ~675 nm and emit at ~680 nm. The substitution of the ring with sulfonate groups results in a product mixture of the various isomers of the mono-, di-, tri- and tetrasulfonated forms ($AlPcS_{1-4}$) [25,26]. The chemical structure of $AlPcS_2$ is shown in Fig. 1a, and Fig. 1b shows the time instance when $AlPcS_3$ is in the detection zone. The snapshots are taken with exposure time of 60 ms so that the peak broadening occurs for analyte velocities larger than 0.83 mm s^{-1}. The peak broadening effect can be reduced by decreasing the exposure time at the expense of reduced sensitivity, or binning more pixels at the expense of lower separation efficiency. However, our studies have shown that peak broadening effects can be greatly reduced using inverse filtering techniques [27–29] so maintaining sensitivity and separation efficiency. These methods are based on image restoration techniques when there are known degradation parameters (i.e., known distance of horizontal motion within the detection zone). Such peak broadening
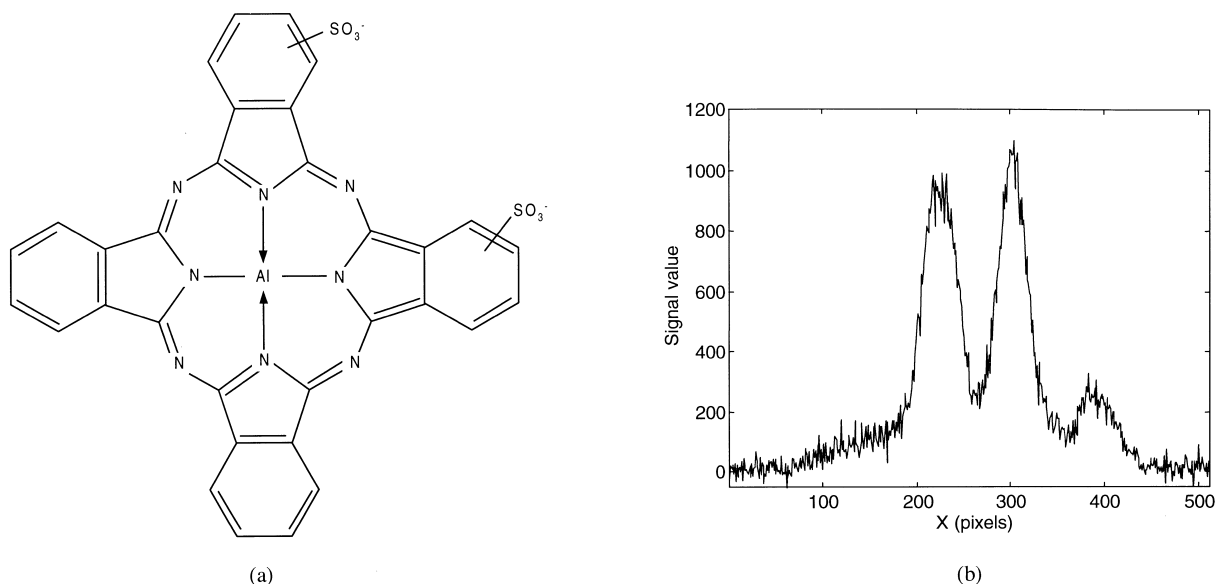
(a)   (b)

Fig. 1. (a) Disulfonated aluminium phthalocyanine (AlPcS$_2$) structure. Tinolux BBS is a AlPcS$_n$ mixture and (b) fluorescence profile of AlPcS$_3$ contained in Tinolux BBS. Separation conditions: injection 5 nl, 10 μM Tinolux BBS, capillary, 550 mm (length from capillary injection end to the end of detection zone 370 mm)×75 μm, buffer 20 mM 3-cyclohexylamino-1-propanesulfonic acid (CAPS), pH 10.5, separation voltage 20 kV, LIF detection at 680 nm.

effects were not observable for the experiments described in this paper, and so are not considered.

3. Electropherogram construction using spatial profiles

Using the spatial profiles (acquired with a frame time of τ), a conventional electropherogram can be constructed as shown in Fig. 2a. The pixel at position 0 equates to a single point detector at the end of the detection zone furthest from the capillary injection end (see Fig. 2b). Each pixel value, $p_i(t)$, in the acquired profile is projected onto the electropherogram time axis, t_E, by calculating the time, $t_i(t)$, necessary for the analyte at pixel i to reach pixel 0 position. Therefore, the time position on the electropherogram, t_{Ei}, for each pixel value at the position i can be calculated as:

$$t_{Ei} = t + t_i(t) \quad \text{for } i = [0, N-1] \tag{1}$$

where t is the experiment time (i.e., time instant when the snapshot was acquired); $t_i(t)$ is the time necessary for the analyte at the pixel i to reach pixel

0; N is the total number of pixels across the capillary window.

The time $t_i(t)$ can be calculated as (see Fig. 2b):

$$t_i(t) = \frac{il_d t}{l_c n + l_d (N - 1 - i)} \quad \text{for } i = [0, N-1] \tag{2}$$

where l_c is the capillary length from injection end to the beginning of the detection zone; and l_d is the length of the detection zone.

Fig. 2a shows how boundary pixels at the positions 0 and $n-1$ are projected for different time instances (i.e., t_0, $t_0 + \tau$ and $t_0 + 2\tau$) onto the electropherogram time axis. Pixels at positions $p^1(t_0)$, $p^1(t_0 + \tau)$ and $p^1(t_0 + 2\tau)$ should be projected onto the same point of the electropherogram, where the superscript 1 indicates the peak which appeared first in the detection zone. Similarly, points $p^2(t_0 + \tau)$ and $p^2(t_0 + 2\tau)$ characterise the second peak. The electropherogram of Tinolux BBS constructed using this method is shown in Fig. 3a. Fig. 3b shows the electropherogram of the same separation constructed using smoothed spatial profiles demonstrating the improvement in the electropherogram signal-to-noise
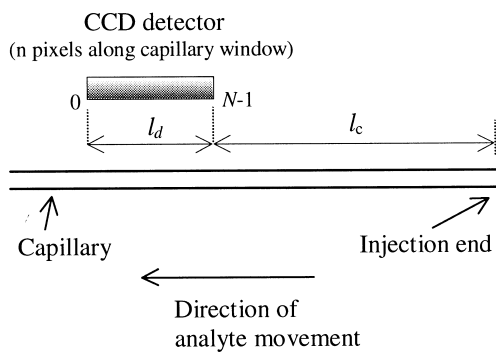
Snapshots

$y(t_0)$ $\qquad\qquad$ $y(t_0+\tau)$ $\qquad\qquad$ $y(t_0+2\tau)$

$t^1(t_0)$ $\qquad\qquad$ $t^2(t_0+\tau)$ $\qquad\qquad$ $t^2(t_0+2\tau)$

$t^1(t_0+\tau)$ $\qquad\qquad$ $t^1(t_0+2\tau)$

0 $\qquad$ $p^1(t_0)$ $\qquad$ $n-1$ $\qquad$ $p^1(t_0+\tau)$ $\qquad$ $p^2(t_0+\tau)$ $\qquad$ $p^1(t_0+2\tau)$ $\qquad$ $p^2(t_0+2\tau)$

Electropherogram

t_E

τ

$t_0+t^1(t_0)$

$t_0+\tau+t^1(t_0+\tau)$

$t_0+2\tau+t^1(t_0+2\tau)$

$t_0+\tau+t^2(t_0+\tau)$

$t_0+2\tau+t^2(t_0+2\tau)$

(a)

CCD detector
(n pixels along capillary window)

0 $\qquad$ $N-1$

l_d $\qquad\qquad$ l_c

Capillary $\qquad\qquad$ Injection end
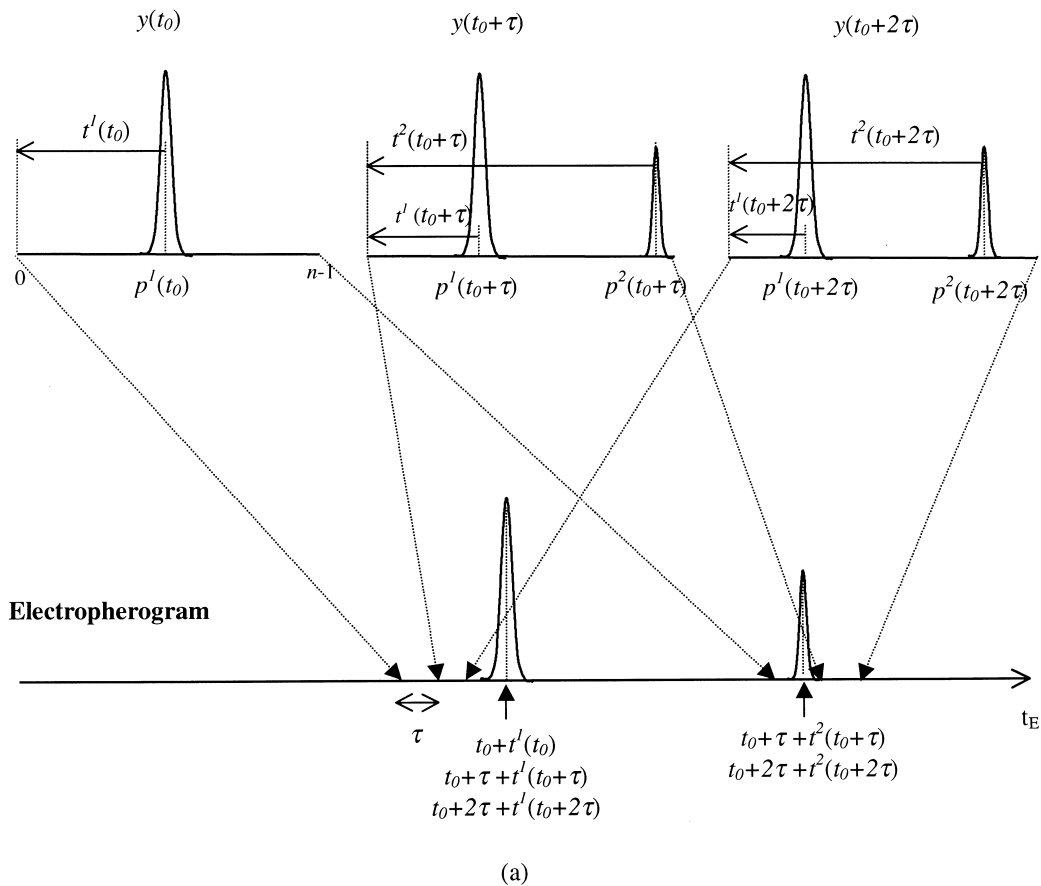
Direction of
analyte movement

(b)

Fig. 2. (a) Constructing an electropherogram using pixel 0 (pixel furthest from the injection end). The pixel value is projected to the electropherogram time axis by calculating the time necessary for the analyte at that pixel to reach pixel 0. All pixel values are projected using this methodology. (b) Position and dimensions of CCD detector with respect to the injection end.
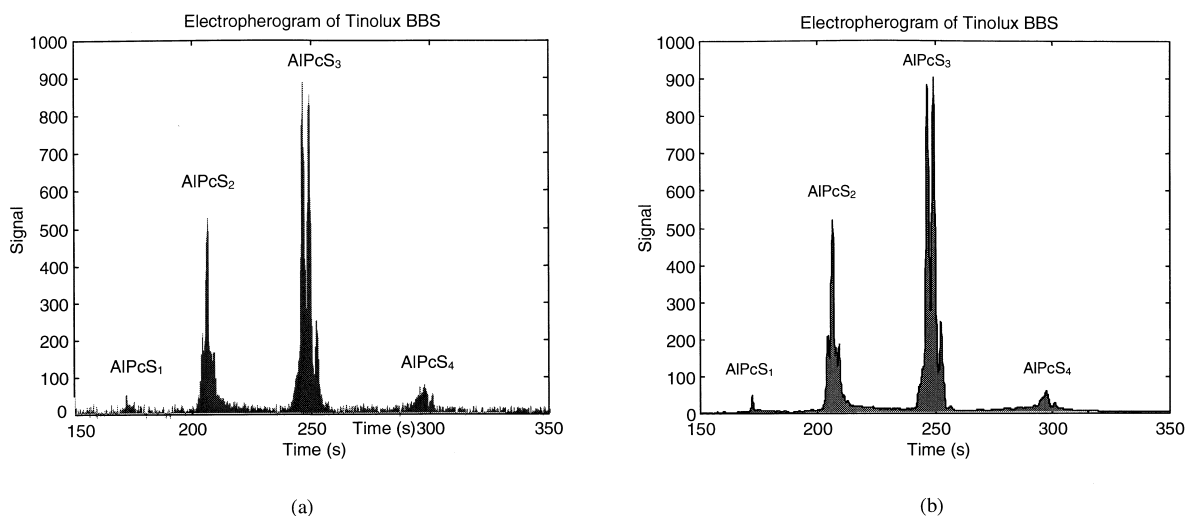
Fig. 3. (a) Electropherogram constructed using the pixel furthest from the capillary injection end. Separation conditions: injection 5 nl, 10 μM Tinolux BBS, capillary, 550 mm (length from capillary injection end to the end of detection zone 370 mm)×75 μm, buffer 20 mM CAPS, pH 10.5, separation voltage 20 kV, detection at 680 nm. (b) Electropherogram constructed using the processed snapshot data.

ratio (S/N). The smoothed profiles are obtained by approximating the raw data points using the RBNN described in later sections.

4. Dynamic processing of CE data

Constructing a conventional electropherogram inevitably causes some loss of dynamic information which is present in the successive profiles. In order to use the full potential of the spatial data, it would be advantageous to process every successive profile, so extracting all of the available information. This would allow the on-line calculation of mobility, concentration and peak width. Furthermore, by continuous peak tracking, it should be possible to discriminate between analyte and artefacts peaks. The signal processing methodology for fluorescence profiles comprises the following steps: (i) S/N ratio enhancement using profile averaging, (ii) profile approximation, (iii) peak identification and (iv) velocity and peak width calculation.

The S/N ratio enhancement results from averaging the time-shifted successive profiles in order to decrease superimposed uncorrelated noise present in the data. The method for real-time CE peak approximation is based on a supervised constructive RBNN

which uses a variable set of non-linear functions to represent the experimental profiles. Once the profile is approximated, the peak parameters (amplitudes, heights and standard deviations) can be identified and subsequently, velocity and width for each peak calculated. Each of these procedures has been tested on a wide variety of LIF data and typical results are presented in the following sections.

4.1. S/N ratio enhancement using profile averaging

Fluorescence profiles contain superimposed uncorrelated noise dominated by detector readout and electronic system noise with a negligible shot-noise component. In contrast, absorbance profile signals with their very high background illumination are shot-noise limited. The real benefit of using all the available spatial data is that the S/N ratio can be greatly improved through averaging, as shown in Fig. 4, by integrating the signal as it passes through the capillary window. The signal at the time instance t_0, $y_a(t_0)$, is obtained by adding the current snapshot $y(t_0)$ to the previously integrated signal $y_a(t_0 - \tau)$, where τ is the frame time between successive camera snapshots. Before these signals are added together, all pixel values in $y_a(t_0 - \tau)$ are shifted in the
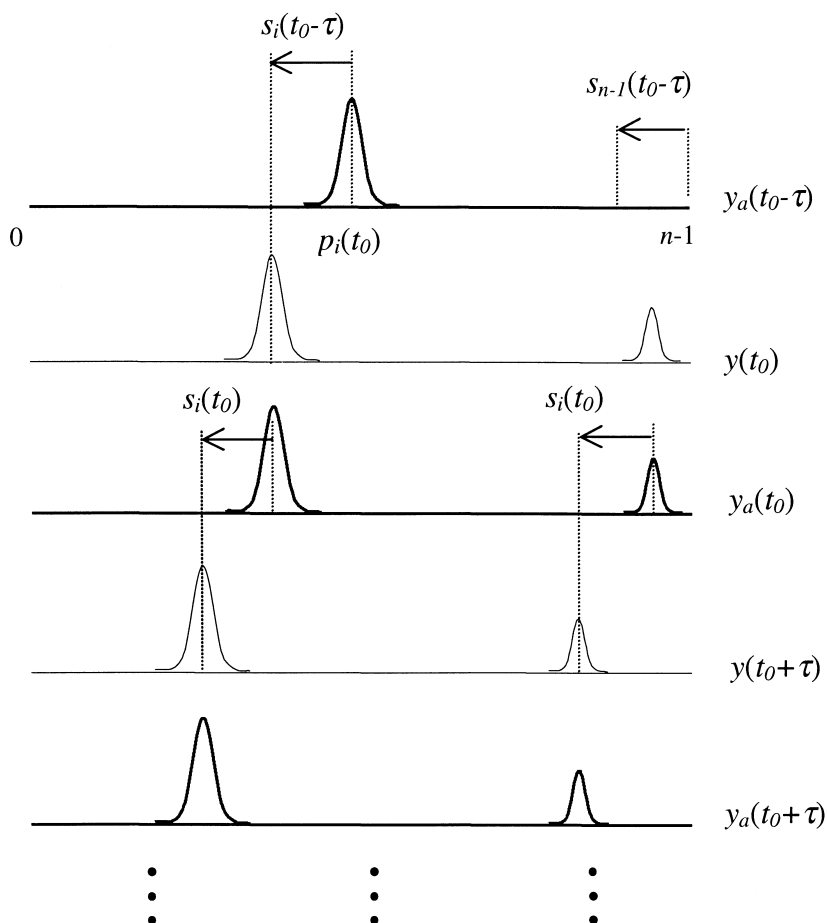
Fig. 4. S/N ratio enhancement is achieved by snapshot averaging. Since the analyte velocity at all time instances is known, the signal is integrated by shifting all of the pixel values in the previously integrated snapshot $y_a(t_0 - \tau)$ and adding it to the current snapshot $y(t_0)$. This results in a new integrated signal $y_a(t_0)$ whose pixel values are shifted and added to the new snapshot $y(t_0 + \tau)$.

direction of analyte movement. For every pixel value at the position i, a shift value $s_i(t_0 - \tau)$ is calculated using the CCD and capillary dimensions (see Fig. 2b):

$$s_i(t_0 - \tau) = \frac{n - 1 - i + n\dfrac{l_c}{l_d}}{t_0 - \tau}\tau \quad [\text{pixels}] \qquad (3)$$

Therefore, the integrated signal $y_a(t_0)$ is given by:

$$y_a(t_0) = p_i(t_0) + p_k(t_0 - \tau) \quad \text{where } k = i - s_i(t_0 - \tau)$$
$$\text{for } i = [0, N - 1]$$
$$\text{and } k \geq 0 \qquad (4)$$

The averaged (normalised) signal, $\bar{y}_a(t_0)$, is calculated using an index array $N_i(t_0)$ which maintains a count of how many pixel values are integrated at a particular position i:

$$\bar{y}_a(t_0) = \frac{p_i(t_0)}{N_i(t_0)} \quad \text{for } i = [0, N - 1] \qquad (5)$$

Through averaging, the S/N ratio is increased by a factor of $\sqrt{N_i(t)}$. The maximum possible improvement in S/N ratio within the detection zone is inversely proportional to the analyte observed speed in the detection zone. Fig. 5 shows an example of AlPcS$_1$ in the detection zone with and without signal integration. For this example the analyte velocity is
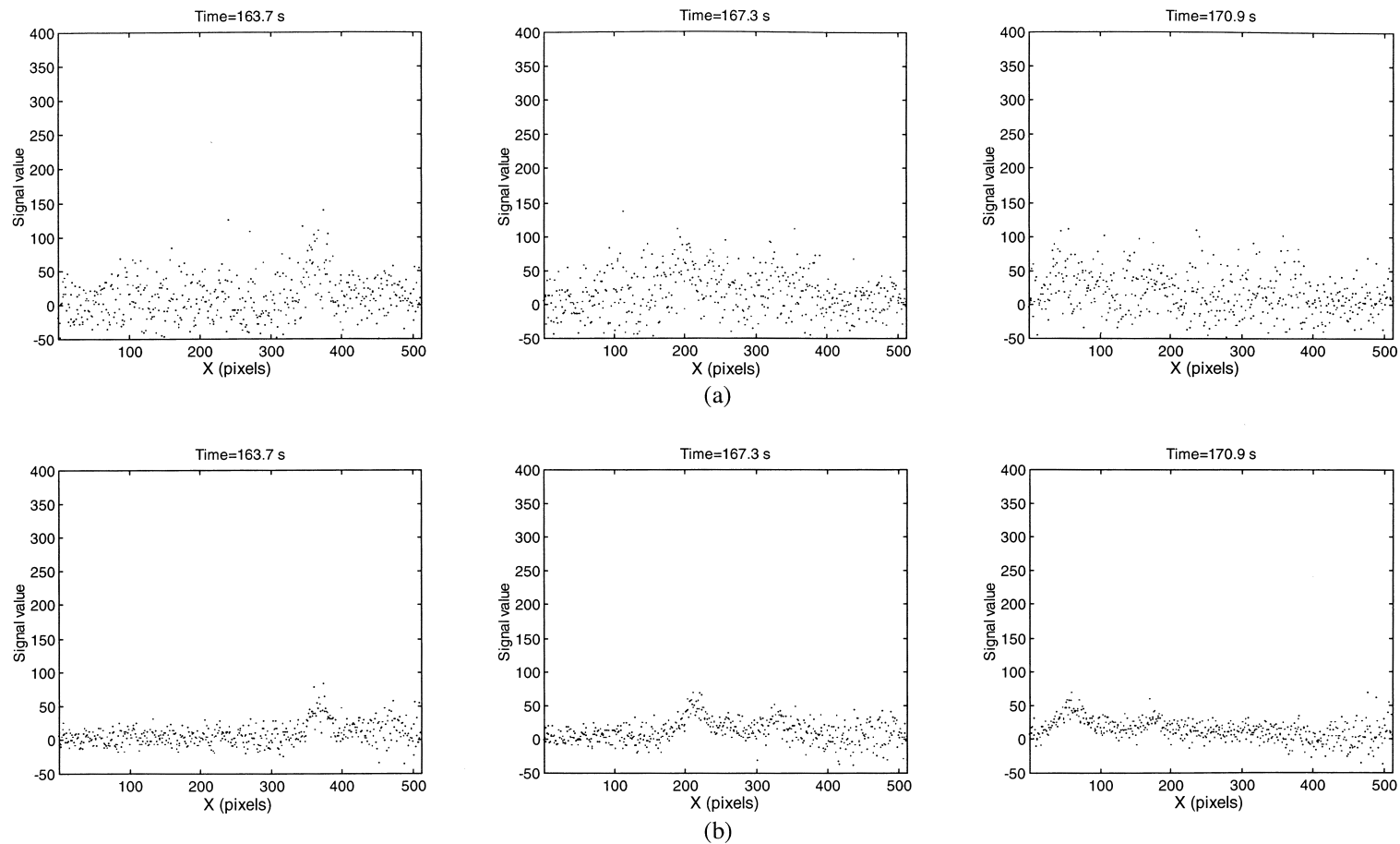
Fig. 5. (a) Snapshots of AlPcS$_1$ without using signal averaging and (b) averaged snapshots of AlPcS$_1$. Separation conditions: injection 5 nl, 10 μM Tinolux BBS, capillary, 550 mm (length from capillary injection end to the end of detection zone 370 mm)×75 μm, buffer 20 mM CAPS, pH 10.5, separation voltage 20 kV, LIF detection at 680 nm.

3.16 mm s^{-1} and the camera frame rate is 0.83 Hz. This enabled nine snapshots containing the fully described peak to be acquired [i.e., $N_{512}(t)=9$], resulting in a S/N ratio improvement of 3 and therefore three-times lower LOD. Similarly, the S/N ratio improvement for AlPcS$_4$ in the far left hand region of the detection window with a lower analyte velocity (1.25 mm s^{-1}) gave an S/N ratio improvement of $\sqrt{15}$. The LOD for a 9-nl injection of Rhodamin 700 is reduced from $3.5 \cdot 10^{-9}$ M to $1.0 \cdot 10^{-9}$ M using this S/N ratio enhancement method.

The software signal integration offers useful S/N ratio (and therefore LOD) improvements, but a loss in separation efficiency occurs if the calculated shift values, s_i, do not match the analyte velocity. This problem is also present in the TDI mode [4], since the shift rate is determined using only the geometry of the CE system (i.e., size of the observation zone, length of the capillary and the experiment time). In the snapshot mode, the peaks can be monitored as they cross the observation zone, so that the true velocity can be measured prior to the calculation of shift values. This eliminates the problem of reduced separation efficiency [4], but increases the computational effort, since the spatial profiles need to be approximated twice.

4.2. Real-time neural networks for signal approximation

The real-time fluorescence signal approximation is carried out using the RBNN architecture shown in Fig. 6. This type of neural architecture is known as a multilayer feedforward network, where the network consists of an input layer of source nodes, one output layer and one or more hidden layers of neurons [30]. In this case there is only one hidden layer of neurons, so the RBNN can be considered as a three-layer feedforward network.

The RBNN approximates a set of K input data points $d_k = \text{f}(x_k)$ by a weighted sum of L, usually non-linear, functions φ termed radial-basis functions. This can be represented by the following equation [17]:

$$d_k \approx \sum_l \omega_i \varphi(\|x_k - c_l\|) \quad \text{for } k = \{1,....K\} \text{ and } l$$
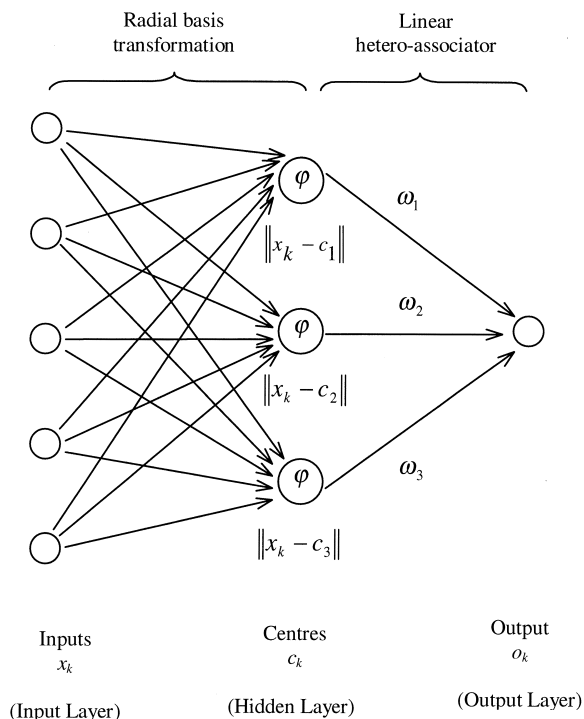$$= \{1,....L\} \qquad (6)$$

Fig. 6. Architecture of a typical radial basis neural network (RBNN).

where c_l represent the set of centres of these functions and $\|.\|$ denotes a norm that is usually taken to be the Euclidean distance. Several choices are possible for the basis functions including inverse multiquadratics, splines and Gaussians. In the following examples, simple Gaussian functions are employed; namely

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-c)^2}{2\sigma^2}\right) \qquad (7)$$

where σ^2 is the variance of the Gaussian distribution and c is its centre. The principle of RBNN operation can be summarised as a complex non-linear mapping from the input pattern space to the output pattern space in a two-step process. In the first step, the non-linear mapping from the input layer to the hidden layer is performed. This is followed by the second step where the hidden layer is linearly transformed to the output layer via a set of learnt weights ω_l [30]. The calculation of the weights ω_l is based on the well known pseudoinverse method [30]:

$$\begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_l \end{bmatrix} = \begin{bmatrix} \varphi_{11} & \varphi_{12} & ... & \varphi_1 K \\ \varphi_{21} & \varphi_{22} & ... & \varphi 2K \\ \vdots & \vdots & \vdots & \vdots \\ \varphi L1 & \varphi_{L2} & ... & \varphi_{LK} \end{bmatrix}^{-1} \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_K \end{bmatrix} \quad (8)$$

where $\varphi_{ji} = \varphi(\|x_i - c_j\|)$.

The crucial step in function approximation is the choice of the radial basis function centre values (c_l) and their variance (σ^2). Since the profile approximation and subsequent processing needs to be performed in real-time, any non-linear optimisation during the RBNN training should be avoided. Therefore, the variance of all the basis functions is set to the same specified value, and this value is changed only if required as discussed below [31]. The calculation of centre values is performed using a supervised learning rule allowing the radial basis functions to be added in a constructive fashion [30,31]. This is achieved using orthogonal regularised least squares forward selection [30]. In this way, the radial basis functions are placed in those positions of the input space where potential peaks are present. The initial radial basis function variance can be calculated using the kurtosis measure of a peak in the detection region [32]. An alternative method for setting the initial variance is to estimate the peak width using the

known injection plug length and longitudinal diffusion coefficient. However, the latter requires some a priori knowledge about the analyte. Further adjustments of the radial basis variance are made according to the RBNN convergence performance and related cost function [32]. Fig. 7a shows an example of a RBNN function approximation performed during the separation of Tinolux BBS at the instance when the AlPcS$_2$ analyte was passing through the detection zone. Using the supervised learning procedure for kernel centre positioning, 12 radial basis functions with σ pixels are sufficient to approximate the data with an error smaller than the amplitude of the superimposed system noise, η. The approximation error of the profile acquired at the instance t, $e(t)$, is given by:

$$e(t) = \frac{1}{K} \sum_{i=1}^{K} \left[\sqrt{N(t)} \cdot |F(x_i) - d(x_i)| - \sqrt{F(x_i)} \right] \quad (9)$$

where $F(x)$ is the response of the RBNN to a given input value x; and $N_i(t)$ is the index array used for averaged signal normalisation (see Eq. (5)).

The term $\sqrt{F(x_i)}$ takes into account the additional multiplicative shot noise when a peak is present. The noise amplitude, η, can be estimated from the
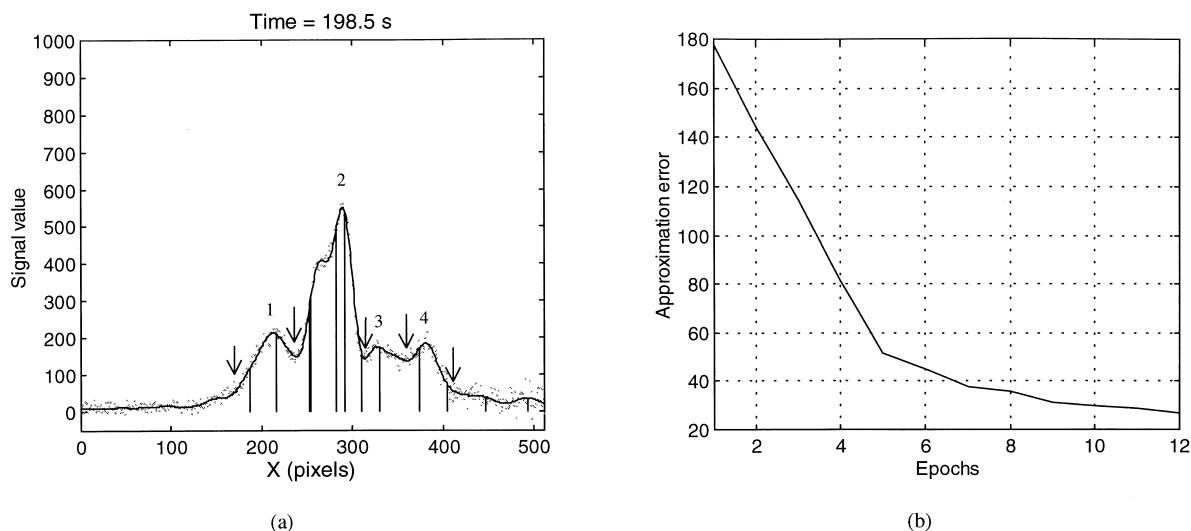

Fig. 7. (a) RBNN approximation of AlPcS$_2$ using 12 basis functions positioned at centres marked by vertical lines. Arrowheads represent the pixel boundary lines (see text for details). The identified peaks are marked with numbers 1, 2, 3 and 4. (b) Change of the fitting error during approximation procedure. Separation conditions: injection 5 nl, 10 μ*M* Tinolux BBS, capillary, 550 mm (length from capillary injection end to the end of detection zone 370 mm)×75 μm, buffer 20 m*M* CAPS, pH 10.5, separation voltage 20 kV, LIF detection at 680 nm.

profiles when no peaks are present. Fig. 7b shows the change in the approximation error during 12 training epochs, where one epoch refers to the period needed to calculate one kernel position using the full input data set.

Once the data has been approximated, it is necessary to identify meaningful peak positions in these profiles. This is achieved by using the first and second derivatives of the RBNN response in order to locate local minima and maxima, and points of maximum curvature in order to geometrically separate peaks. The peak boundary positions are marked with arrows in Fig. 7a; also shown are the four identified peak positions. The peak standard deviations are calculated using the RBNN response between the boundary positions. Using

$$F(x_k) = A_i \exp\left(-\left(\frac{x_k - p_i}{\sigma}\right)^2\right) \tag{10}$$

the estimated peak standard deviation, σ_a, is calculated by averaging values of the RBNN response standard deviations between the two specified boundary points, b_{left} and b_{right}:

$$\sigma_a = \frac{1}{(b_{\text{right}} - b_{\text{left}})} \sum_{k=b_{\text{left}}}^{b_{\text{right}}} \left| \frac{x_k - p_i}{\sqrt{\ln \frac{A_i}{F(x_k)}}} \right| \tag{11}$$

where p_i are the kernel positions; b_{left} and b_{right} are the left and right boundaries of peak p_i, respectively; and A_i is the amplitude of the peak.

An example of this approach is shown in Fig. 8b for the AlPcS$_3$ profile, which has been approximated by three Gaussian kernels. The calculation of standard deviations is valid only if the peak has an appropriate Gaussian shape. If the peak cannot be approximated by a Gaussian function and providing that experimental conditions suggest that there is no peak distortion (i.e., small sample concentration and known analyte properties), the detected peak may represent two or more unseparated components. This situation is shown in Fig. 9, where the highest peak is decomposed into two peaks by using a constrained non-linear minimisation procedure between the peak boundary points and the previously identified peaks. The minimum peak variance can be estimated from the upper bound to the radius of any ion in solution $(10^{-10}$ m) [21] and using the relationship for diffusional variance
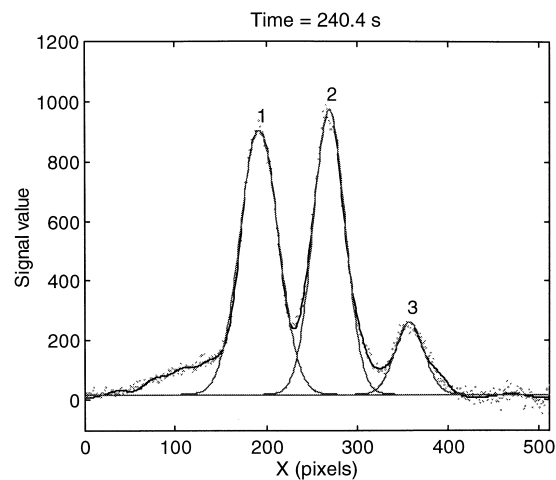
$$\sigma^2 = 2Dt \tag{12}$$

where the diffusion coefficient D relates to the hydrodynamic radius r as

$$D = \frac{kT}{6\pi\eta r} \tag{13}$$
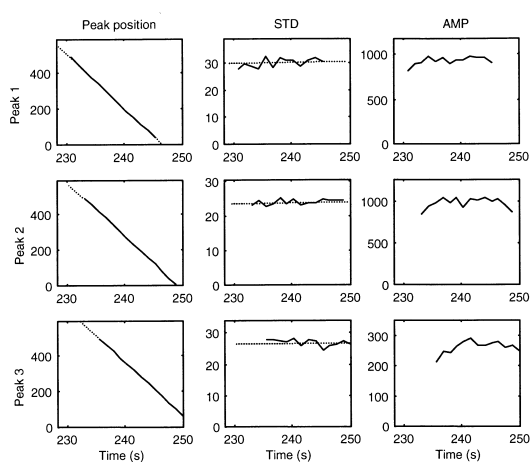
Fig. 8. (a) On-line AlPcS$_3$ peak tracing results showing peak position, standard deviation (STD) and amplitude (AMP) for the three main peaks. Dashed lines in the peak position and standard deviation graphs provide estimates for velocity and diffusion coefficients, respectively. (b) Example of averaged snapshot profile with RBNN approximation and identified peaks. Separation conditions as in Fig. 8.
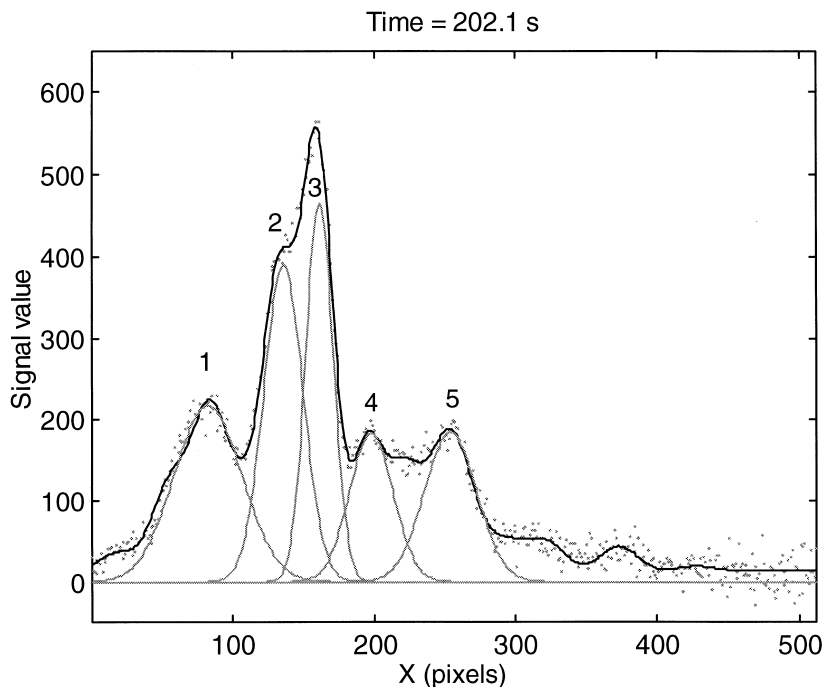
Fig. 9. Significant peak detection in AlPcS$_2$ LIF profile. Peaks numbered 1, 4 and 5 are detected using the same methodology as in Fig. 7. Since the highest peak is not symmetrical, it is decomposed into two peaks numbered 2 and 3. Separation conditions as in Fig. 8.

where k is the Boltzmann constant, T the temperature and η the viscosity.

4.3. Mobility and diffusion coefficient calculation

In conventional single point illumination CE systems, the mobility calculation of a particular ion is based on only one peak position measurement. Using a CCD detector, the mobility can be calculated by tracking peaks across the observed capillary region. Since the number of identified peaks in successive snapshots might not be the same (i.e., effect of noise, artefacts, etc.), it is necessary to track only valid peaks (i.e., peaks representing true analyte components). Artefacts can be recognised using a velocity histogram [33] and maximum allowed amplitude change constraint. The velocity histogram method provides an effective way to discriminate between analytes and artefacts, since the artefacts usually appear in only one or two consecutive frames. Once the valid peaks are identified, their velocities and mobilities can be calculated. The observed velocity, v_{obs}, for the ith peak at time t_c is found using a linear regression fit through mini-

mising the mean squared error for all available observed peak positions, $p_i(t)$. Namely,

$$\bar{v}_{obs}(t_c) = \frac{\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k p_i(t_k) - \sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}p_i(t_k)}{\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k^2 - \left(\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k\right)^2}$$

(14)

This linear regression includes a chi-square merit function which takes into account the different errors for the measured peak positions due to signal averaging. Greater weight is given to peak positions where $N_i(t)$ is greater.

Fig. 8 show the results of the on-line peak tracing of AlPcS$_3$ during the separation of Tinolux BBS. Each peak is processed so that the peak position, peak amplitude and peak standard deviation (related to the mobility, concentration and diffusion coefficient, respectively) are calculated for each frame. The slopes of the dashed lines in peak position graphs give $v_{obs} = 1.54$ mm s^{-1} for all three AlPcS$_3$ peaks. The velocities of AlPcS$_1$, AlPcS$_2$ and AlPcS$_4$ at the end of the detection zone are 2.16 mm s^{-1},

1.77 mm s^{-1} and 1.25 mm s^{-1}, respectively. The corresponding 95% confidence intervals for the four AlPcS$_2$ are ± 0.018 mm s^{-1}, ± 0.014 mm s^{-1}, ± 0.012 mm s^{-1} and ± 0.009 mm s^{-1}, respectively.

The consistency of observed velocity measurements is shown in Fig. 10. This test was performed on a dilution series of AlPcS$_2$, first at a concentration of 500 nM and followed by 160, 50, and 16 nM. The velocity measurements at different concentrations are marked with small circles with indicated 95% confidence intervals. As it can be seen, the velocity measurements are in good agreement, taking into consideration the 30-fold concentration change. The mean velocity is 1.773 mm s^{-1} with the standard deviation of 0.003 mm s^{-1}. The mean velocity calculated from the electropherogram is 1.764 mm s^{-1} with the standard deviation of 0.017 mm s^{-1}. The reason for the much greater standard deviation in the velocity measurement using the electropherogram is the approximately 1 s timing uncertainty after the high voltage supply is switched. This is observed as a slight shift of the peak in the electropherograms (see Fig. 10a).

The measurement of diffusion coefficients using peak heights for selective determination of proteins has been reported elsewhere [34]. A method for the determination of diffusion coefficients using real-time peak width measurements is presented in this paper based on our previous work [21]. Eq. (12) shows how the variance changes with time due to diffusion. In the absence of any other time-dependent variance contributions, it is possible to calculate the diffusion coefficient for the ith peak, D_i, using the following relationship, and the standard deviations, $\sigma_a(t)$, previously calculated from Eq. (11),

$$D_i = \frac{\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k\sigma_{ai}^2(t_k) - \sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}\sigma_{ai}^2(t_k)}{2\cdot\left(\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k^2 - \left(\sum_{k=1}^{c}\sqrt{N_{p_i}(t_k)}t_k\right)^2\right)}$$

(15)

where c is the number of measurements.

The slopes of the dashed lines in the standard deviation graphs in Fig. 8a relate to the diffusion coefficients D. The standard deviation of 30 pixels relates to the peak width at full-width half-maximum (FWHM) of 3.53 mm. However, the observation time for which the detected compounds were in the active capillary region is too small for a reliable calculation of the diffusion coefficients. For example, the expected diffusion coefficient of AlPcS$_2$ of $3.5\cdot 10^{-10}$ m^2 s^{-1} [21] would give a standard deviation change of only $+2$ pixels within the 15 s observation time. Therefore, more reliable diffusion coefficient
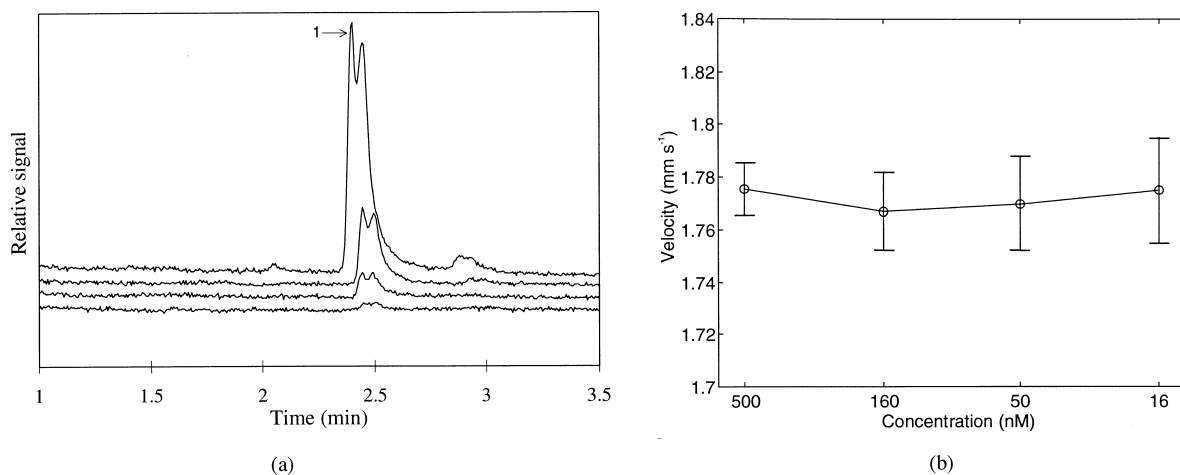

Fig. 10. (a) Electropherograms of dilution series of AlPcS$_2$. There are four electropherograms for the concentrations of 16 nM, 50 nM, 160 nM and 500 nM. Separation conditions: injection 2.5 nl AlPcS$_2$, capillary, 450 mm (length from capillary injection end to the end of detection zone 260 mm)$\times$75 μm, buffer 5 mM CAPS, pH 10.5, separation voltage 15 kV, detection at 680 nm. (b) Velocity of the peak marked 1 calculated using the real-time profile processing showing the consistency of mobility measurements for different concentrations. Also shown is the 95% confidence interval for calculated velocities (± 0.010 mm s^{-1}, ± 0.015 mm s^{-1}, ± 0.018 mm s^{-1}, ± 0.020 mm s^{-1}).

measurements would require flow speed control, so that peak velocities can be adjusted to give a longer observation time. Furthermore, since the S/N ratio enhancement would attenuate any change in the peak shape, the following procedure for diffusion coefficient measurement is proposed. Once the desired peak is detected, the flow is stopped so that the peak is in the middle of the detection zone. The operating mode of the camera is changed to the 4×1 binning (four pixels binning perpendicular to the capillary) with a longer exposure time (typically 10 s) [21]. The longer exposure increases the S/N ratio without sacrificing resolution because the analyte is no longer moving. The peak widths are calculated from every captured camera image and the diffusion coefficient is determined using Eq. (15). The number of required measurements can be determined on-line depending on the desired confidence interval [35]. Our initial studies showed that the measurement of peak widths can provide accurate determination of diffusion coefficients [21] and using the method for real-time profile processing presented here, this process can potentially be fully automated.

5. Real-time issues

The profile analysis task is ideally performed within the frame acquisition time of the CCD, which in the LIF mode is 1.2 s. After the initial image processing, the obtained profiles are transferred to a digital signal processing platform, where subsequent processing tasks are performed. The processing platform consists of five Analog Devices Super Harvard Architecture Computer (SHARC) processors with shared memory and communication busses, so allowing efficient parallel processing. When the profile processing for the current camera frame is complete, the results are displayed by the instruments' host personal computer, the feedback signal for flow control is generated and the processing platform waits for the next available profile.

6. Conclusions

In this paper a technique based on RBNN for the on-line analysis of fluorescence profiles obtained from a CCD-based CE system has been presented. The real-time profile processing provides dynamic information during the separation which allows rapid and automated measurement of peak velocities and widths. The discrimination between analytes and artefacts is improved and storage requirements can be greatly reduced, since the profiles where no peaks are present need not to be recorded. The real-time processing method has been successfully tested for sulfonated aluminium phthalocyanine and other samples. Currently, the method is being extended to real-time analysis of multi-wavelength absorbance profiles.

Acknowledgements

References

[1] P. Camilleri, Capillary Electrophoresis – Theory and Practice, CRC Press, Boca Raton, FL, 1993.
[2] R. Kuhn, S. Hoffsette, Capillary Electrophoresis – Principles and Practice, Springer-Verlag, New York, 1993.
[3] G. Ross, P. Kaltenbach, D. Heiger, Todays Chem. Work 6 (1997) 31.
[4] J.V. Sweedler, J.B. Sheer, H.A. Fishman, R.N. Zare, R.H. Scheller, Anal. Chem. 63 (1991) 496.
[5] W.B. Burke, Image Acquisition – Handbook of Machine Vision Engineering, Chapman and Hall, London, 1996.
[6] R.B. Bilhorn, P.M. Epperson, J.V. Sweedler, M.B. Denton, Appl. Spectros. 41 (1987) 1125.
[7] J.V. Sweedler, R.D. Jalkian, M.B. Denton, Appl. Spectros. 43 (1989) 953.
[8] R.B. Bilhorn, J.V. Sweedler, P.M. Epperson, M.B. Denton, Appl. Spectros. 41 (1987) 1114.
[9] J. Johansson, T. Johansson, S. Nilsson, Biomed. Optoelect. Clin. Chem. Biotechnol. 2 (1993) 2629.
[10] J. Johansson, D.T. Witte, M. Larsson, S. Nilsson, Anal. Chem. 68 (1996) 2766.
[11] S. Lillard, E. Yeung, M. Mccloskey, A. Micheal, Adv. Fluores. Sens. Acta 2980 (1997) 133.
[12] S. Xiong, J. Li, J. Cheng, Anal. Chim. Acta 322 (1996) 187.
[13] J. Preisler, S. Yeung, Anal. Chem. 68 (1996) 2885.

[14] T. Timperman, K. Khatib, J.V. Sweedler, Anal. Chem. 67 (1995) 139.

[15] A. Saha, J. Christian, D.S. Tang, C.L. Wu, Adv. Neural Information Proc. Systems 3 (1991) 482.

[16] K. Ng, R.P. Lippmann, Adv. Neural Information Proc. Systems 3 (1991) 970.

[17] S. Chen, B. Mulgrew, S. McLaughlin, Proceedings of the 3rd IEEE International Conference on Communications, Chicago, IL, 1992, p. 1267.

[18] D. Lowe, A.R. Webb, Network 1 (1990) 299.

[19] D.S. Broomhead, D. Lowe, Complex Systems 2 (1988) 321.

[20] J. Moody, C.J. Darken, Neural Comput. 1 (1989) 281.

[21] D.M. Goodall, E.T. Bergström, Anal. Chem., submitted for publication.

[22] P.M. Epperson, M.B. Denton, Anal. Chem. 61 (1989) 1513.

[23] R.L. Chein, D.S. Burgi, J. Chromatogr. 559 (1991) 141.

[24] K.E. Oldenburg, X. Xi, J.V. Sweedler, J. Chromatogr. A 788 (1997) 173.

[25] A.J.G. Mank, C. Gooijer, H. Lingeman, N.H. Velthorst, U.A.T. Brinkman, Anal. Chim. Acta 290 (1994) 103.

[26] M. Ambroz, A. Beeby, A.J. MacRobert, M.S.C. Simpson, R. Stevenson, D. Phillips, Photochem. Photobiol. 9 (1991) 87.

[27] H. Tang, L.W. Cahill, Proceedings of the IEEE International Symposium on Circuits and Systems, Singapore, June 1991, p. 520.

[28] A. Rosenfeld, A.C. Kak, Digital Picture Processing, Academic Press, London, 1976.

[29] M. Sonka, V. Hlavac, R. Boyle, Image Processing, Analysis and Machine Vision, Chapman and Hall, London, 1994.

[30] S. Haykin, Neural Networks – A Comprehensive Foundation, Macmillan, New York, 1994.

[31] H. Abdi, J. Biol. Systems 2 (1994) 247.

[32] B. Pokrić, N.M. Allinson, D.M. Goodall, E.T. Bergström, IEE Proceedings: Vision, Image and Signal Processing, submitted for publication.

[33] U. Tomomasa, M. Iguchi, Z. Morita, Optical Diagnostics Fluid Thermal Flow 683 (1993) 2005.

[34] J. Wu, J. Pawliszyn, Anal. Chem. 64 (1992) 2934.

[35] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recepies in C, Cambridge University Press, Cambridge, 1992.