



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Multi-Snapshot Imaging for Chromatographic Peak Analysis

Citation for published version:

Hopgood, J, Connelly, M, McHoull, B & Troy, D 2018, 'Multi-Snapshot Imaging for Chromatographic Peak Analysis', *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 119-129.
<https://doi.org/10.1109/TBME.2018.2826144>

Digital Object Identifier (DOI):

[10.1109/TBME.2018.2826144](https://doi.org/10.1109/TBME.2018.2826144)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

IEEE Transactions on Biomedical Engineering

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Multi-Snapshot Imaging for Chromatographic Peak Analysis

James R. Hopgood, *Member, IEEE*, Matthew Connelly, Barry McHoull, and Darren Troy

Abstract—Objective: Snapshot imaging has several advantages in automated gel electrophoresis compared with the finish-line method in capillary electrophoresis; this comes at the expense of resolution. A novel signal processing algorithm is proposed enabling a multi-snapshot imaging (MSI) modality whose objective is to substantially improve resolution. MSI takes multiple-captures in time as macromolecules are electrophoresed. Peaks from latter snapshots have high resolution but low signal-to-noise ratio (SNR), while earlier snapshots have low resolution but high SNR. **Methods:** Signals at different capture-times are related by a scale-in-separation, shift-in-separation, and amplitude gain. The proposed method realigns the multiple captures using least-squares and fuses them. The algorithm accounts for the partial waveforms observed as the chromatic peaks exit the sensor’s field-of-view. **Results:** MSI improves resolution by approximately 10% on average per minute of additional electrophoresis. **Conclusions:** Comprehensive analysis of the resolution are quantified on several datasets demonstrate the effectiveness of MSI. **Significance:** MSI can double the resolution compared with traditional snapshot imaging over a typical set of captures.

Index Terms—Signal processing & modelling, chromatography, snapshot imaging, finish-line, parameter estimation, fusion.

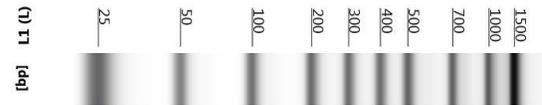
I. INTRODUCTION

ELECTROPHORESIS is a fundamental and ubiquitous technique in the separation sciences for analysing individual macromolecules in biological samples, such as DNA, RNA, and proteins [1], [2]. It has important uses in forensics, molecular biology, biochemistry, and genetics. Specific applications include protein analysis [3] and genetic fingerprinting [4]. While next generation sequencing facilitates the precise order of nucleotides within a DNA sequence, gel electrophoresis provides an economically viable quality control stage, and is generally flexible, quick, parallelisable, reliable, and cheap.

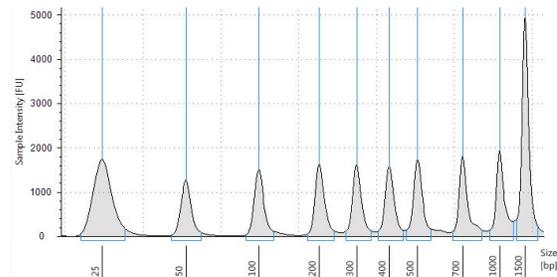
Fundamental to the performance of any separation technique is the notion of resolution, which measures the ability of the system to separate macromolecules. Considerable research effort has been expended on improving resolution as detection of a single additional band can have a substantial impact on the evaluation of the assay results, for example in DNA fingerprinting or purity analyses [1], [2], [5], [6].

There are two common detection methods for measuring the separation of macromolecules in electrophoresis, namely the finish-line method and snapshot optical imaging [1], [7]. In the finish-line method, molecules are detected after electrophoresis for a constant distance using a single point detector

James Hopgood (james.hopgood@ed.ac.uk) is with the School of Engineering, Institute of Digital Communications, University of Edinburgh (UoE), UK. This work funded by Royal Academy of Engineering’s Industrial Secondment Scheme, and EPSRC Impact Acceleration Account (UoE). Matthew Connelly, Barry McHoull, and Darren Troy are with Agilent Technologies, Edinburgh.



(a) Image of typical electrophoresis trace.



(b) A 1D electrophoresis signal.

Fig. 1. Traditional snapshot of an electrophoresis trace of a standard DNA ladder, as (a) an image and (b) a 1-D signal. The standard electrophoresis process is run for sufficiently time that the signal fills the sensor’s entire FOV.

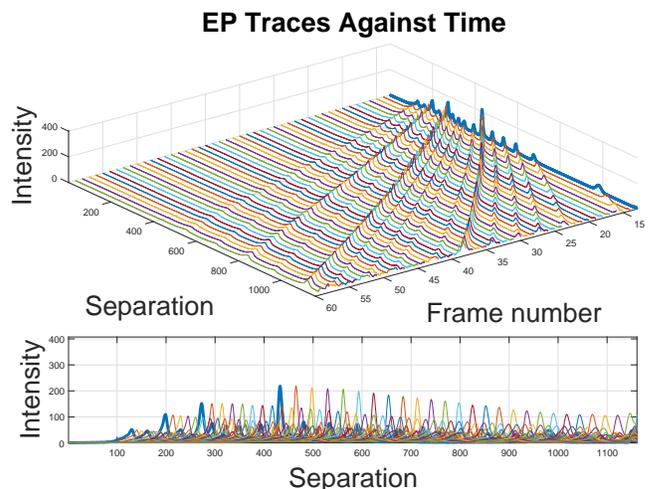


Fig. 2. Proposed algorithm utilises multiple electrophoresis snapshots.

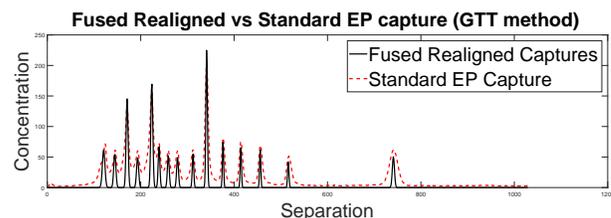


Fig. 3. Using multiple electrophoresis snapshots, the proposed algorithm (see Fig. 4) offers superior resolution of fragment components (see Section IV-B).

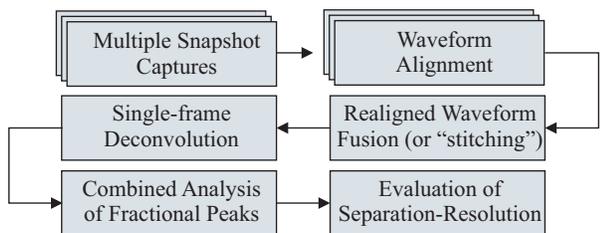


Fig. 4. An overview of the complete algorithm, showing waveform realignment, fusion, and deconvolution for resolution quantification.

at some fixed point in space. In optical snapshot imaging, molecules are electrophoresed for a constant time, in which the electrophoresis (EP) time is calibrated such that the sample has expanded to fit the imaging sensor’s entire field of view (FOV), thereby maximising the use of the sensor’s spatial resolution (see example in Fig. 1). The entire EP trace is therefore imaged after some fixed duration. In both cases, the calibration is usually performed using a standard DNA ladder for benchmarking [2], [5], [6]. Finish-line methods generally have the best resolution as a result of molecules running to a constant distance [8], but is generally more complicated to implement. Finish-line methods using multi-pixel detectors have also been developed to enhance signal-to-noise ratio (SNR) [9], [10], thereby improving sensitivity.

Automated EP products are now ubiquitous in genomics labs [11]. Most commercial devices use either capillary electrophoresis (CEP) with finish-line measurement, or gel electrophoresis (GEP) using snapshot imaging. While CEP has excellent resolution, automated CEP technologies can be slow, expensive, and have limitations compared to snapshot GEP. For example: commercial CEP products usually run samples sequentially rather than in parallel, thus taking more time to analyse multiple samples; channels reused for several samples are prone to *carry-over* or potential contamination between samples [12]; and use of a single-detector means a molecule is observed only once, so high dynamic range techniques used in GEP cannot be utilised without a multi-pixel detector [9], [10]. Snapshot GEP does not suffer these restrictions. However, to benefit from the advantages of automated snapshot GEP, techniques must be developed that build on snapshot imaging and provide the resolution of finish-line CEP.

In GEP, improvements in resolution have typically been achieved by various approaches involving the alteration of the physical conditions of the separation, for example using doping and nanoparticles [5], [6], or by altering how the analytes migrate through the gel (gradient gel electrophoresis). This paper develops a signal processing framework (Fig. 4) for multi-snapshot imaging (MSI), without introducing new significant hardware expense. By introducing an approach for improving resolution through detection alone, and signal processing techniques for information fusion, this novel technique is applicable to existing separation methods without altering any of the physical conditions of the separation.

In the proposed approach, information from multiple images (Fig. 2) are realigned and fused together to provide a single high-resolution image superior to that obtained at a standard

EP time (Fig. 3). This technique is timely due to advances in imaging technology, hardware processing capability, and automation on commercial products. The proposed MSI method provides several benefits for the analysis of macromolecules compared with single-snapshot imaging (SSI) and the finish-line method. These include improved resolution, improved estimation of fragment concentration and size, and reduction of static background noise or observation noise. While algorithms have looked at different aspects for improving the analysis of GEP [13]–[15], MSI has only recently been introduced in concept [16], [17], and has not previously been implemented in a commercial product. This paper presents a detailed analysis of the method.

A. Contributions

The work in this paper builds on the concept first presented in [16], [17], by adding the following substantial contributions:

- a comprehensive analysis of MSI resolution as a function of time, demonstrating the strength of the MSI approach;
- novel fusion techniques for combining the realigned traces to produce a high resolution equivalent GEP trace;
- a comprehensive evaluation of the algorithm over several datasets, comparing with a standard industrial benchmark.

B. Paper Organisation

The remainder of this paper is organised as follows: Section II motivates the use of MSI, provides a detailed description of resolution as a quantitative measure for performance, and provides an overview of the proposed approach. Section III provides details of two realignment algorithms; Section IV presents two methods for fusing the realigned waveforms. Section V presents a comprehensive set of results showing the realignment process, waveform fusion, and quantification of resolution. Finally, conclusions are presented in Section VI.

II. MOTIVATION FOR MULTI-CAPTURE IMAGING

To understand MSI, this section reviews SSI and investigates its resolution properties. An example of SSI in GEP is shown in Fig. 1(a). The central image rows (typically around ten rows are used) are extracted and averaged to create an 1-D signal, as shown in Fig. 1(b). In cases where surface tension leads to a significant meniscus, it is straightforward to realign the image rows using cross-correlation prior to the averaging process, although this is not generally needed in this work. In Fig. 1(b), the x -axis denotes either uncalibrated molecular weight, base-pair fragment length, or more generally *separation*, with smaller particles on the left, and larger ones on the right.

In the standard approach to snapshot detection for GEP, an electrophoresis measurement trace (EPMT) is made of the concentration field in space at some fixed point in time. This EP time is set to ensure that the field envelops the imaging sensor’s FOV, thereby maximising use of the sensor’s spatial resolution. However, an improvement in resolution with snapshot imaging can be obtained by running the EP for longer and taking multiple snapshots. The resulting EPMTs are shown in Fig. 2, where the peaks continue to separate more quickly

than they broaden. The frame number indicates sampling time, and here the sampling period is 60 seconds. Note that in Fig. 2 and the remaining figures in this paper, the peaks with highest molecular weight are allocated the smallest separation values.

A. Separation Resolution and Measurement

To quantify the improvement in resolution obtained from MSI, define at time t the separation resolution between any two peaks, $i \in \{p, q\}$, with means $\mu_{t,i}$ and full-width at half maximums (FWHMs) of $W_{t,i}$ as [1], [18]:

$$R_{t,p,q} \triangleq \frac{2|\mu_{t,p} - \mu_{t,q}|}{(W_{t,p} + W_{t,q})} \quad (1)$$

Suppose that the resolution of an EPMT is to be compared with the resolution at the standard EP time, defined at $t = t_0$. A robust metric is to measure the resolution between all combinations of pairs of peaks, rather than just adjacent ones, in order to maximise the amount of information used from the signals. However, non-adjacent peaks will have a larger separation compared with adjacent peaks, and consequently appear to have a higher resolution. Therefore, to compare the overall change in resolution of the EPMT as a function of time, it is better to define the average normalised resolution:

$$\Delta \hat{R}_t = \frac{1}{|\mathcal{P}_t|} \sum_{\{p,q\} \in \mathcal{P}_t} \frac{R_{t,p,q}}{R_{t_0,p,q}} \quad (2)$$

where \mathcal{P}_t is the set of all peaks in the sensor's FOV at time t , and $|\mathcal{P}_t|$ is the cardinality of \mathcal{P}_t (number of peak pair combinations). As t increases, some peaks leave the FOV or reach the end of the EP channel, and are not included in (2).

To calculate the peak centers, $\mu_{t,i}$, and FWHMs, $W_{t,i}$, non-linear regression is used to decompose the EPMTs into a linear combination of known peak shapes. This peak-deconvolution is standard in this field [18]. To demonstrate how the estimated average normalised resolution, $\Delta \hat{R}_t$, changes as a function of time, t , only three candidate peak shapes are considered for simplicity, as shown in TABLE I. These peak shapes are the *Gaussian*, *Lorentzian*, and *Voigt* profiles. As noted in [15], there are numerous chromatic peak models that could be used, including asymmetric peak shapes. However, the peaks shapes considered here are sufficient for our purposes. The model for the EPMT at time t as a function of separation x is thus:

$$\hat{c}_t[x|\boldsymbol{\theta}_t] = b_t[x|\boldsymbol{\theta}_{t,b}] + \sum_{p=1}^{P_t} \alpha_{t,p} \mathcal{F}_p[x|\boldsymbol{\theta}_{t,p}] \quad (3)$$

where: $\hat{c}_t[x|\boldsymbol{\theta}_t]$ is the spatially quantised model for the underlying continuous concentration field, $c(x, t)$; $b_t[x|\boldsymbol{\theta}_{t,b}]$ denotes a parametric model for the baseline, given here by a second-order polynomial: $b_t[x|\boldsymbol{\theta}_{t,b}] = \sum_{k=0}^2 a_k x^k$, $\boldsymbol{\theta}_{t,b} = \{a_k\}_0^2$. Using model-order selection techniques, higher-order polynomials are seen to not contribute significantly to the underlying baseline. Moreover, $\mathcal{F}_p[x|\boldsymbol{\theta}_{t,p}]$ denotes the functional form of the peak-shapes given in TABLE I; P_t is the number of peaks; and $\boldsymbol{\theta}_{t,w} = \{\alpha_{t,p}\}_{p=1}^{P_t}$ are the linear weighting coefficients. The peak-model parameters $\boldsymbol{\theta}_{t,p}$ are also defined in TABLE I, where a given model is used for all peaks in a given EPMT. The complete set of model parameters

| Peak Shape | Params $\boldsymbol{\theta}$ | Notation $\mathcal{F}(x \boldsymbol{\theta})$ | Functional Form | FWHM W |
|------------|---|---|---|---|
| Gaussian | $\boldsymbol{\theta}_G = \{\mu, \sigma\}$ | $\mathcal{G}(x \boldsymbol{\theta}_G)$ | $e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $W_G = \sqrt{8 \ln 2} \sigma$ |
| Lorentz | $\boldsymbol{\theta}_L = \{\mu, \gamma\}$ | $\mathcal{L}(x \boldsymbol{\theta}_L)$ | $\frac{\gamma^2}{\gamma^2 + (x-\mu)^2}$ | $W_L = 2\gamma$ |
| Voigt | $\boldsymbol{\theta}_V = \{\mu, \sigma, \gamma\}$ | $\mathcal{V}(x \boldsymbol{\theta}_V)$ | $G(x \boldsymbol{\theta}_G) * \mathcal{L}(x \{0, \gamma\})$ | $W_V \approx \alpha_V W_L + \sqrt{\beta_V W_L^2 + W_G^2}$ |

TABLE I
PARAMETRIC PEAK-SHAPES (NOTE $\alpha_V = 0.5346$ AND $\beta_V = 0.2166$.)

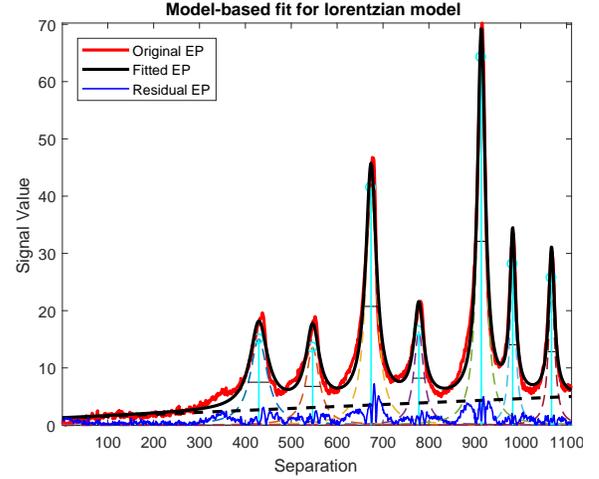


Fig. 5. Deconvolved waveforms showing: the original waveform (thick light red curve); baseline and constituent components (dotted curves); reconstructed waveform (thick dark red curve); and residual waveform (bottom dark thin blue curve).

$\boldsymbol{\theta}_t = \{\boldsymbol{\theta}_{t,b}, \boldsymbol{\theta}_{t,w}, \boldsymbol{\theta}_{t,1}, \dots, \boldsymbol{\theta}_{t,P_t}\}$ are found by nonlinear least-squares minimisation of the squared error between the concentration field $c(x, t)$ and the EPMT model $\hat{c}_t[x|\boldsymbol{\theta}_t]$:

$$\hat{\boldsymbol{\theta}}_t = \arg \min_{\boldsymbol{\theta}_t} \sum_{x \in \mathcal{R}_0} (c(x, t) - \hat{c}_t[x|\boldsymbol{\theta}_t])^2 \quad (4)$$

where each model defined in TABLE I is considered, \mathcal{R}_0 indicate the set of separation values in the sensor's FOV, and the model with least total residual error is the selected model.

The optimisation is performed using the trust-region-reflective nonlinear minimisation algorithm subject to bounds [19]. The initial conditions for this optimisation need to take on sensible values. In particular, the initial peak locations are obtained using the peak feature-extraction method from [20], the parameters controlling the FWHM and peak-heights are constrained to be positive with maximum values determined by separation length and the maximum EPMT value, respectively.

A typical deconvolution of an EPMT is shown in Fig. 5, where the dotted curved lines indicate the individual components or fractional chromatic peaks, the horizontal lines indicates the FWHM, the thick dotted line indicates the baseline model, and the residual is shown at the bottom of the graph. The fitted waveform is, due to the low-dimensional parametric shape, smoother than the original waveform.

The sequence of EPMTs shown in Fig. 2 are analysed at each time-step, t , by finding the best-fit peak-deconvolution given by solving (4). Using the estimated parameters, $\hat{\boldsymbol{\theta}}_t$, the

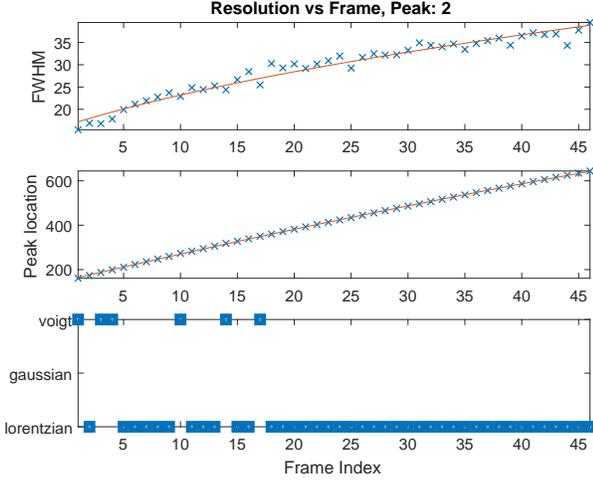


Fig. 6. The FWHM ($W_{t,p}$), peak center ($\mu_{t,p}$), and best-fit model as a function of time for the second peak in the EPMT.

expressions in TABLE I are used to find the peak centers, $\mu_{t,p}$, and FWHMs, $W_{t,p}$. An example of these parameters for the second-peak in the exemplar waveform is shown in Fig. 6 where, in this instance, it can be seen the Lorentzian usually fits the data best. However, the Voigt profile is often a better fit in earlier snapshots where the peaks can be asymmetric due to the residual effect of the initial sample loading.

The change in average-normalised resolution, $\Delta\hat{R}_t$, is calculated using (2). This *resolution improvement* as a function of time is shown in Fig. 7, where the resolution effectively doubles by continuing the EP process. The errorbars indicate variability in the resolution improvement for individual peak-pair combinations. As an exemplar, the resolution improvement is considered for the two so-called *upper-peaks*. These peaks are at the left of Fig. 5 at lowest separation values; since EP moves the peaks to the right, the upper-peaks stay in the FOV for all frames. The resolution improvement for these peaks is shown in the starred line in Fig. 7. It is crucial to realise that for the resolution to improve, the rate at which the peaks separate must be greater than the rate-of-change of the FWHM (see Section II-B). As EP continues, peaks leave the sensor's FOV, requiring a method to fuse information from different snap-shots. Such a method is proposed in Section IV.

B. Theoretical Resolution as a Function of Time

Section II-A demonstrated through an experimental example that a resolution improvement can be obtained with snapshot imaging by running the electrophoresis for longer. This section provides a theoretical justification. Consider the simplest model for describing the evolution of the concentration field, $c(x, t)$. This model, used in both the analysis of finish-line and snapshot detection methods, considers the evolution of a unit mass injected as a delta function at the origin of an EP system [1]. The concentration field, $c(x, t)$, resulting from this injection is, at time t and separation x , the solution to the averaged convection-diffusion equation for large-times,

$$\frac{\partial c}{\partial t} + \bar{U} \frac{\partial c}{\partial x} = \bar{D} \frac{\partial^2 c}{\partial x^2} \quad (5)$$

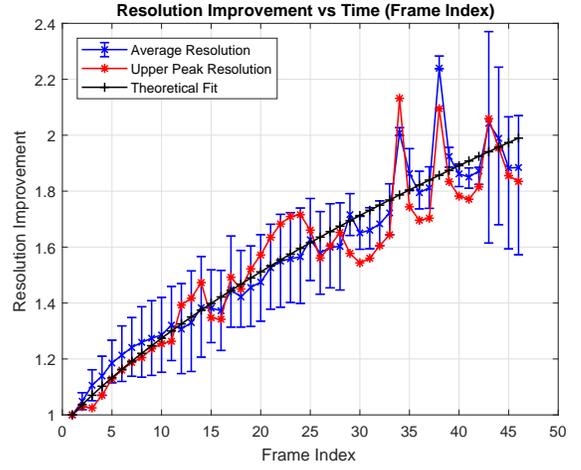


Fig. 7. Resolution improvement vs time. The average resolution improvement, $\Delta\hat{R}_t$, is shown with 90% confidence intervals and a regression analysis (8). The normalised resolution between the upper peaks is shown as an example.

where \bar{U} is the mean velocity vector of the injected molecules, and \bar{D} the dispersion coefficient. Using linear superposition, the solution in this idealised case is a Gaussian pulse [1]:

$$c_i(x, t) = \frac{1}{\sqrt{2\pi}\sigma_{t,i}} \exp \left[-\frac{(x - \mu_{t,i})^2}{2\sigma_{t,i}^2} \right] \quad (6)$$

where $c_i(x, t)$ represents the i -th peak with mean position $\mu_{t,i} = \bar{U}_i t$, and peak variance $\sigma_{t,i}^2 = 2\bar{D}_i t$. It is crucial to note that, in this idealised EPMT model, the position of the i -th peak increases linearly with time, while the peak's FWHM increases sub-linearly with time, namely that $W_{t,i} = 2\sqrt{2\ln 2} \sigma_{t,i} = 4\sqrt{(\ln 2) \bar{D}_i t}$ varies with the square root of time. This means the peaks separate more quickly than they broaden, and explains the resolution improvement as a function of time seen in Fig. 7. Therefore, the resolution between two peaks $i \in \{1, 2\}$ varies, using (1), as:

$$\begin{aligned} R_{t,1,2} &= \frac{|\mu_{t,1} - \mu_{t,2}|}{\sqrt{2\ln 2} (\sigma_{t,1} + \sigma_{t,2})} \\ &= \frac{U_{t,1} - U_{t,2}}{2\sqrt{\ln 2} (\sqrt{D_1} + \sqrt{D_2})} \sqrt{t} \propto \sqrt{t} \end{aligned} \quad (7)$$

These results can easily be extended to the case where an impulse is injected at time $t = t_0$ at separation $x = x_0$, with the times t being replaced by $t - t_0$. Inspired by these results, and allowing for practical effects in the EP process, the theoretical fit shown in Fig. 7 to model the overall improvement in average normalised resolution is of the functional form:

$$\Delta\hat{R}_t = \sqrt{1 + \alpha t^\beta} \quad (8)$$

where α and β are found by a nonlinear least squares fit [19]. In Fig. 7, $\alpha \approx 0.133$ and $\beta \approx 0.77$, for this particular dataset.

C. Resolution Enhancement through Realignment

To benefit from the improvement in resolution in time, the captures must be realigned in order to fuse the information. To show that realignment maintains the resolution improvement,

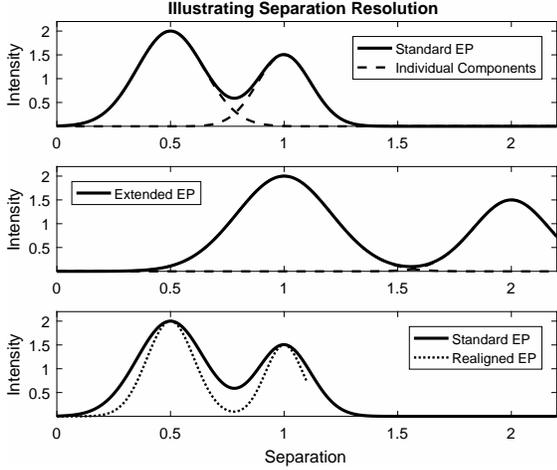


Fig. 8. Illustration of resolution change as electrophoresis continues. In this figure, the electrophoresis process moves to the right, in contrast to Fig. 1.

consider again the concentration field, $c(x, t)$, at times $t = t_0$ and $t = t_1 > t_0$, as shown in the upper two plots of Fig. 8. At $t = t_k$, $k = \{0, 1\}$, the two peaks, $i \in \{1, 2\}$, have mean positions $\mu_{t_k, i} = \bar{U}_i t_k$ and deviations of $\sigma_{t_k, i} = 2\bar{D}_i t_k$. Suppose the waveform at $t = t_1$ is rescaled in the separation axis, x , by a factor of $\Delta_{01} = \frac{t_0}{t_1}$, yielding the bottom plot in Fig. 8. The mean positions of the rescaled waveform are now:

$$\hat{\mu}_{t_1, i} = \Delta_{01} \mu_{t_1, i} = \frac{t_0}{t_1} \bar{U}_i t_1 = \bar{U}_i t_0 = \mu_{t_0, i} \quad (9)$$

meaning the rescaled peak positions match the original peak positions. Moreover, the rescaled peak-widths are:

$$\hat{\sigma}_{t_1, i} = \Delta_{01} \sigma_{t_1, i} = \frac{t_0}{t_1} \sqrt{2\bar{D}_i t_1} = \sqrt{\frac{t_0}{t_1}} \sigma_{t_0, i} \quad (10)$$

Therefore, the resolution between the two peaks, $i \in \{1, 2\}$, of the rescaled waveform is given by equations (1) and (9) as:

$$\hat{R}_{t_1, 1, 2} \triangleq \frac{|\hat{\mu}_{t_1, 1} - \hat{\mu}_{t_1, 2}|}{\sqrt{2 \ln 2} (\hat{\sigma}_{t_1, 1} + \hat{\sigma}_{t_1, 2})} = \sqrt{\frac{t_1}{t_0}} R_{t_0, 1, 2} \quad (11)$$

As $t_1 > t_0$, the resolution $\hat{R}_{t_1, 1, 2} > R_{t_0, 1, 2}$ has improved, as shown by the narrower peaks in the lower plot of Fig. 8. These results are readily extended if an impulse injected at (x_0, t_0) . In practice, due to several physical effects, the peak shapes are not Gaussian as given by (6), although other shapes such as Voigt, pseudo-Voigt, and asymmetric variants are frequently used [15]. Nevertheless, the resolution improvement from this realignment process applies irrespective of peak shape.

D. Proposed Multi-Capture Imaging Algorithm

The complete algorithm is shown in Fig. 4; an outline of this was first presented in [16]. Multiple snapshots are aligned and fused to produce a single waveform. Single-frame peak deconvolution is applied to the fused waveform and, using the properties of the resulting fractional peaks, the resolution in (2) evaluated. The following sections outline each step.

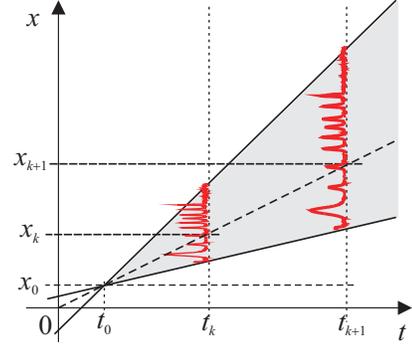


Fig. 9. Linear GEP expansion model. This is a rotated schematic of Fig. 2.

III. PROPOSED REALIGNMENT ALGORITHM

The realignment mechanism in MSI can be posed as an estimation problem. Carlson [21] produced a maximum-likelihood cross-correlation algorithm for problems involving Doppler shifts, although it assumed the full waveform is present in both measurements. In MSI, the scenario is complicated by the fact the signal leaves the sensor's FOV and thus matching is required over partial signals: direct application of correlation gives inaccurate realignment. This section derives the algorithm using an expansion model to design a signal model.

A. Linear Chromatography Expansion Model

If the EP process is continued past the standard EP time, the macromolecules continue to separate, as shown in the example in Fig. 2. If the electric field is constant, and the behaviour of the gel remains invariant, the convection-diffusion equation (5) indicates a linear expansion model, since the mean-position of the injected molecules increases linearly with time, as shown in (6). This expansion is shown diagrammatically in Fig. 9. Suppose a sample is injected such that the origin of the GEP expansion begins at (x_0, t_0) . A macromolecule at *any point* (x_k, t_k) with concentration $c(x_k, t_k)$ will later appear at position (x_{k+1}, t_{k+1}) with concentration $c(x_{k+1}, t_{k+1})$ through a simple geometric mapping. In the linear expansion case, from similar triangles (see Fig. 9), the EPMT at these two positions are related by $c(x_{k+1}, t_{k+1}) \propto c(x_k, t_k)$, and the position of a molecule at t_{k+1} and t_k are related through:

$$\frac{x_{k+1} - x_0}{t_{k+1} - t_0} = \frac{x_k - x_0}{t_k - t_0} \quad (12)$$

or, equivalently, writing $\Delta_k = \frac{t_{k+1} - t_0}{t_k - t_0}$ and $\Sigma_k = \frac{t_{k+1} - t_k}{t_k - t_0}$,

$$x_{k+1} = \Delta_k x_k + \Sigma_k \quad (13)$$

The concentration field at the two positions are proportionally related since the total concentration is conserved, even though the field has spread out. This implies values from the EPMTs at a given position x at times t_k and t_{k+1} are related:

$$c(x, t_{k+1}) = \alpha_k c\left(\frac{x - \Sigma_k}{\Delta_k}, t_k\right) \quad (14)$$

where α_k is a gain coefficient. Note that $\Sigma_k = (1 - \Delta_k) x_0$. Therefore, if $x_0 \geq 0$, and the scaling factor $\Delta \geq 1$, then the shift-in-separation $\Sigma_k \leq 0$, while if $x_0 < 0 \Rightarrow \Sigma_k > x_0$.

Given the initial injection point, $\mathcal{I}_0 = (x_0, t_0)$, the scale-in-separation can be evaluated from one frame to the next, $\Delta_k = \Delta_k(t_0, t_k, t_{k+1})$, as can the amplitude gain α_k , and the shift $\Sigma_k = \Sigma_k(x_0, t_0, t_k, t_{k+1})$. However, it is not always possible to accurately determine \mathcal{I}_0 , and therefore the position of the injection point should be assumed unknown. Moreover, while \mathcal{I}_0 can be estimated from a number of multiple-snapshots assuming linear expansion (for example, the equiphase vertexing map in [9], [10]), in practice, in GEP there is deceleration in the expansion model due to Joule heating, gel ionization, deterioration or loss of the fluorescent dyes, and other effects. Therefore, the scale-in-separation, amplitude-gain, and shift should be estimated on a per-snapshot (frame by frame) basis. These parameters can be estimated efficiently using a least-squares estimate (LSE) approach, as shown in Section III-B.

B. LSE Parameter Estimation

Consider realigning snapshots (frames) at times $t = t_P$ and $t = t_R$, which here are called the *projected* ($f_P[n]$) and *reference* ($f_R[n]$) frames to distinguish them; the projected frame will be realigned to the reference frame. First assume $t_R > t_P$, which means that the projected frame needs to be stretched to match the reference frame. This ensures the length of the stretched projected frame is longer than the reference frame, and simplifies the limits in the summations in the cost functions used below. If $t_P > t_R$, the rôles of the projected and reference frames are reversed, as described in Section III-B3.

Let $f_k[n] = c(n\delta x, t_k)$ indicate a spatially quantised version of $c(x, t)$ at separations $x = n\delta x$ for $n \in \{0, \dots, N-1\}$, with $\delta x = \frac{1}{N}x_{\max}$, where x_{\max} is the maximum separation.

1) *Linear Model*: In this model, the reference signal, $f_R[n]$, is modelled as a linear multiple of a scaled-and shifted projected frame, $f_P[n]$, with a modelling error $e[n]$:

$$f_R[n] = \alpha f_P[n; \Delta, \Sigma] + e[n] \quad (15)$$

where $f_P[n; \Delta, \Sigma] = f_P[\frac{n\delta x - \Sigma}{\Delta}]$, as given by (14). In order to estimate the unknown set of parameters $\boldsymbol{\theta} = \{\Delta, \Sigma, \alpha\}$, it makes intuitive sense to optimise their values to minimise the average square modelling error. Define the average error as:

$$E_T(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=0}^{N-1} e^2[n] = \frac{1}{N} \sum_{n=0}^{N-1} (f_R[n] - \alpha f_P[n; \Delta, \Sigma])^2 \quad (16)$$

Since $t_R > t_P$, then $\Delta \geq 1$ and the domain of $f_R[n]$ covers the domain of $f_P[n]$, and hence the summation over the N discrete-separation values. The objective is to find the set of parameters $\boldsymbol{\theta}$ that minimises $E_T(\boldsymbol{\theta})$; in other-words:

$$\boldsymbol{\theta}_{\text{opt}} = \arg \max_{\boldsymbol{\theta}} E_T(\boldsymbol{\theta}) \quad (17)$$

The gain, α , can be found analytically given Δ and Σ :

$$\alpha = \frac{\sum_{n=0}^{N-1} f_R[n] f_P[n; \Delta, \Sigma]}{\sum_{n=0}^{N-1} f_P^2[n; \Delta, \Sigma]} \quad (18)$$

In (18) the numerator corresponds to a cross-correlation term between $f_R[n]$ and $f_P[n]$, while the denominator represents the total energy in $f_P[n]$. The total error in (16) is thus:

$$\epsilon_T(\boldsymbol{\theta}) = \sum_{n=0}^{N-1} f_R^2[n] - \frac{\left\{ \sum_{n=0}^{N-1} f_P[n; \Delta, \Sigma] f_R[n] \right\}^2}{\sum_{n=0}^{N-1} f_P^2[n; \Delta, \Sigma]} \quad (19)$$

where $\epsilon_T(\boldsymbol{\theta}) = NE_T(\boldsymbol{\theta})$. The total error in (19) can be minimised with respect to $\{\Delta, \Sigma\}$: a nonlinear LSE problem.

In practice, since gel electrophoresis is actually imaging the fluorescent dyes, which can detach themselves from the macromolecules, there is often a change in an underlying baseline over large differences in sampling time. This can effect the performance of realignment using the linear model, since the model does not explicitly include this baseline. The affine model in Section III-B2 explicitly handles the baseline, but is computationally more intensive. Therefore, to take advantage of the computationally efficient linear model, a constant baseline removal process is introduced prior to realignment. In this work, since many of the EPMTs are relatively sparse, the baseline is estimated using the most common value occurring in the histogram of concentration values. In the histogram, 400 bins are used for 1000 concentration values. This relatively simple method seems to perform adequately in most cases. For many EPMTs the baseline is relatively small. Depending on the implementation, other baseline removal techniques in the literature can be used, such as rolling ball [22] and baseline estimation and denoising using sparsity (BEADS) [23].

2) *Affine Model*: The affine model attempts to address the issue raised above by explicitly incorporating a baseline model into equation (15) in addition to linear amplitude scaling:

$$f_R[n] = \alpha_0 f_P[n; \Delta, \Sigma] + \sum_{q=1}^Q \alpha_q g_q[n] + e[n] \quad (20)$$

where $\{\alpha_q\}_1^Q$ are unknown baseline coefficients, and $g_q[n]$ are Q known basis functions. Modelling the baseline as a linear combination of basis functions is still somewhat constraining, but it is shown to work well here in the realignment of EPMTs. As in Section II-A, low-order polynomials are found to be sufficient. To exploit this model, define the gain vector by $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_Q]^T$, and the augmented *projected signal*:

$$\mathbf{f}_S[n; \Delta, \Sigma] = [f_P[n; \Delta, \Sigma] \quad g_1[n] \quad \dots \quad g_Q[n]]^T \quad (21)$$

such that $f_R[n] = \boldsymbol{\alpha}^T \mathbf{f}_S[n; \Delta, \Sigma] + e[n]$. Defining:

$$E_T(\boldsymbol{\theta}) = \frac{1}{N} \sum_{n=0}^{N-1} (f_R[n] - \boldsymbol{\alpha}^T \mathbf{f}_S[n; \Delta, \Sigma])^2 \quad (22)$$

the objective is, as in (17), to find $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \Delta, \Sigma\}$ that minimises $E_T(\boldsymbol{\theta})$. The optimal gain $\boldsymbol{\alpha}$ satisfies:

$$\underbrace{\left[\sum_{n=0}^{N-1} \mathbf{f}_S[n; \Delta, \Sigma] \mathbf{f}_S^T[n; \Delta, \Sigma] \right]}_{\mathbf{R}[\Delta, \Sigma]} \boldsymbol{\alpha} = \underbrace{\sum_{n=0}^{N-1} \mathbf{f}_S[n; \Delta, \Sigma] f_R[n]}_{\mathbf{r}[\Delta, \Sigma]} \quad (23)$$

Algorithm 1 Realignment of electrophoresis traces.

```

1: procedure EP REALIGNMENT( $f_P[n], f_R[n], Q$ )
2:   if  $t_R > t_P$  then  $\triangleright f_P[n]$  stretched longer than  $f_R[n]$ 
3:      $\theta_{\text{opt}} = \text{ESTIMATEPARAMETERS}(f_P[n], f_R[n], Q)$ 
4:   else  $\triangleright f_R[n]$  stretched longer than  $f_P[n]$ 
5:      $\hat{\theta}_{\text{opt}} = \text{ESTIMATEPARAMETERS}(f_R[n], f_P[n], Q)$ 
6:      $\Sigma_{\text{opt}} = -\hat{\Sigma}_{\text{opt}}$   $\triangleright$  Reciprocate parameters
7:      $\Delta_{\text{opt}} = -1/\hat{\Delta}_{\text{opt}}$ 
8:      $\alpha_{\text{opt}} = 1/\hat{\alpha}_{\text{opt}}$   $\triangleright$  Linear case
9:      $\alpha_{1,\text{opt}} = \frac{1}{\hat{\alpha}_{1,\text{opt}}}, \alpha_{2,\text{opt}} = -\frac{\hat{\alpha}_{2,\text{opt}}}{\hat{\alpha}_{1,\text{opt}}}$   $\triangleright$  Affine case
10:   end if
11: end procedure
12: function ESTIMATE PARAMETERS( $f_P[n], f_R[n], Q$ )
13:   for all grid points over  $\{\Delta, \Sigma\}$  do
14:     Resample  $f_P[n]$  by closest rational fraction to  $\Delta$ .
15:     Calculate gain  $\alpha_{\text{opt}}$ : (18) (linear) or (23) (affine).
16:     Evaluate  $E_T(\theta)$ : (19) (linear) or (24) (affine).
17:   end for
18:    $\{\Delta_{\text{opt}}, \Sigma_{\text{opt}}\} = \arg \min_{\{\Delta, \Sigma\}} E_T(\{\alpha_{\text{opt}}, \Delta, \Sigma\})$ 
19:   return  $\theta_{\text{opt}} = \{\alpha_{\text{opt}}, \Delta_{\text{opt}}, \Sigma_{\text{opt}}\}$ 
20: end function

```

and the total error in equation (22) is:

$$\epsilon_T(\theta) = \sum_{n=0}^{N-1} f_R^2[n] - \mathbf{r}^T[\Delta, \Sigma] \mathbf{R}^{-1}[\Delta, \Sigma] \mathbf{r}[\Delta, \Sigma] \quad (24)$$

where $\epsilon_T(\theta) = NE_T(\theta)$. As in (19), the total error in (24) can be minimised with respect to $\{\Delta, \Sigma\}$ using gradient descent, a grid search, or other minimisation algorithms [19]. In this paper, an exhaustive grid search is used, to investigate the structural form of the cost functions (Fig. 12). The entire algorithm is summarised in pseudo-code in Algorithm 1. Note that if, in fact, there is no baseline present in the signal, the affine model will not perform any better than the linear model.

3) *Case if $t_P > t_R$* : The rôles of the projected and reference frames are reversed, as in lines 6–9 of Algorithm 1: the shift is reversed, the scale-in-separation inverted, and the gains modified. As such, the results above apply directly.

IV. FUSING REALIGNED TRACES

In the illustration in Fig. 8, the center graph shows the concentration field when EP is continued after the standard EP time. It is seen that the right-most peak is leaving the FOV and, therefore, when realigned, only partially overlaps the standard EP trace as shown in the bottom plot of Fig. 8. While the realigned waveform has improved resolution, there is missing information for separations $1.1 < x \leq 2.1$. Therefore, to benefit from the improvement in resolution by running EP for longer, fusion of the realigned pulses obtained in Section III-B is required. A simple fusion approach is to take, at a given x , the realigned signal corresponding to the most-recent capture (MRC), as described in Section IV-A. However, as the EP time increases, the SNR decreases due to the reduction in peak amplitude. It is therefore preferable to use all the information from all captures. Moreover, such brute-force stitching can lead to discontinuities in the fused waveform, which may be

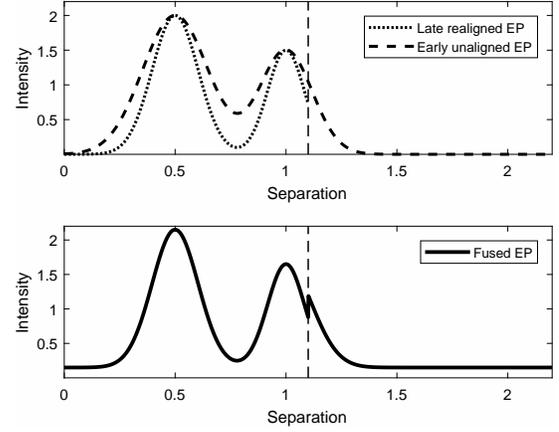


Fig. 10. Illustration of fusing traces in Fig. 8 using MRC (Section IV-A). The upper plot shows the individual realigned traces from two captures. The lower plot shows the fused trace formed of: the late realigned EPMT for separations $0 \leq x < 1.1$; the early unaligned EPMT for $1.1 \leq x < 2.2$.

unacceptable to an end user. Approaches such as waveform averaging leads to band broadening and loss of resolution. A more principled approach to the required fusion is to fit a functional shape through time (Section IV-B).

A. Fusing Traces using MRC

The algorithm for using the MRC essentially just takes the latest measurement for any given separation. This is illustrated in Fig. 10, in which the upper plot shows the early unaligned EPMT and a realigned EPMT resulting from an extended run. The MRC fused EP trace is formed by taking the late trace for separations $0 \leq x < 1.1$, and taking the early trace for $1.1 \leq x < 2.2$. The resulting EPMT shows the resulting discontinuity at $x = 1.1$. However, the combined trace is sufficient for demonstrating improvement in resolution. This fusion approach can be formally defined as follows. Denote the realigned trace at time t by $c[x, t]$, and define the region over which the trace exists by \mathcal{R}_t ; *i.e.* \mathcal{R}_t is the set of *realigned* separation values in the FOV in frame t . In general, $\mathcal{R}_\tau \subset \mathcal{R}_t$ if $\tau > t$, and define $\mathcal{R}_0 = \mathcal{R}$ where \mathcal{R} is the full separation axis. Therefore, the MRC can be defined as:

$$c_{\text{MRC}}[x] = c[x, \tau] \quad \text{where } \tau = \max t \text{ such that } x \in \mathcal{R}_t. \quad (25)$$

A key advantage of MRC is that it doesn't need a signal model, for example utilising peak position estimates. This fusion is thus very simple to implement, while being robust since it directly uses the EPMTs rather than an estimate. The disadvantage of MRC is the discontinuous resulting signal.

B. Fusing Traces Through Time (GTT)

A principled approach for fusion is to consider how the EPMT is predicted to change over time using the averaged convection-diffusion equation (5). After realignment in separation and amplitude, the concentration field is given by (6) with a suitable gain normalisation, and μ_t and σ_t^2 replaced by (9)

and (10), respectively, modified with a time-offset. Therefore, a realigned isolated pulse would be described through time by:

$$C_p(x, t) = A_p \exp \left[-\frac{(x - \mu_{r,p})^2 (t - t_0)}{2\sigma_{r,p}^2 (t_r - t_0)} \right]$$

where $\mu_{r,p}$ and $\sigma_{r,p}^2$ are the mean and variance for the p th peak at reference time $t = t_r$, t_0 is the injection time, and A_p the peak concentration. The model for the EPMT through time can thus be expressed by the 2-D function (*cf.* (3)):

$$\hat{c}[x, t | \theta] = \mathcal{B}[x, t | \theta_b] + \sum_{p=1}^P A_p \exp \left[-\frac{(x - \mu_p)^2 (1 - \alpha_p t)}{2\hat{\sigma}_p^2} \right]$$

where $\mathcal{B}[x, t | \theta_b]$ is a frame-by-frame baseline model as described in Section IV, the set of peak-parameters $\theta_c = \{A_p, \mu_p, \hat{\sigma}_p, \alpha_p\}_{p=1}^P$, and the full set of parameters are $\theta = \{\theta_b, \theta_c\}$. Thus, θ are estimated by minimising

$$\theta_{\text{GTT}} = \arg \min_{\theta} \sum_{t \in \mathcal{T}} \sum_{x \in \mathcal{R}_t} (c[x, t] - \hat{c}[x, t])^2 \quad (26)$$

where $t \in \mathcal{T} = \{t_0, t_1, \dots, t_{F-1}\}$ indicate the frame index, F is the number of frames or traces, and \mathcal{R}_t is defined in Section IV-A. This is a 2-D regression, and is analogous to the 1-D regression problem described in Section II-A. The baseline model for the resolution analysis is a plane:

$$\mathcal{B}[x, t | \theta_b] = (\gamma_0 + \gamma_1 x) (1 + \beta_0 + \beta_1 t) \quad (27)$$

where $\theta_b = \{\gamma_0, \gamma_1, \beta_0, \beta_1\}$. If the baseline for a particular trace is linear as a function of separation, x , then as the EP process continues as a function of time, this gradient increases with time. The simple model in (27) accounts for this change in baseline adequately, although higher order models could be considered. The Gaussian through-time (GTT) fused capture is then found by evaluating $\hat{c}[x, t]$ at the final time-frame:

$$c_{\text{GTT}}[x] = \hat{c}[x, t_{F-1} | \theta_{\text{GTT}}] \quad (28)$$

This can be interpreted as fitting a linear combination of Gaussian peaks with temporally varying FWHM across time, and evaluating the resulting projection at the final capture time, t_{F-1} : the frame at which the best resolution for each peak is expected. This is demonstrated diagrammatically in Fig. 11, where the estimated peak shapes have smoothly varying and decreasing FWHM across frames. The nonlinear optimisation in (26) requires, as in Section II-A, sensible initial conditions. The initial peak locations $\{\mu_p\}_{p=1}^P$ are determined from the reference frame using the same peak detection algorithm as Section II-A. Moreover, the peak variances $\hat{\sigma}_p^2$, heights A_p , and locations μ_p , are constrained to be positive. This model doesn't account for loss of fluorescent dye, nor the asymmetric peaks found in practice. However, since fusion is across time including later EPMTs, this model proves sufficient for fusion presently. Further models will be considered in future work.

V. RESULTS

To demonstrate the improvement in resolution obtained by MSI through realignment and fusion, several typical DNA

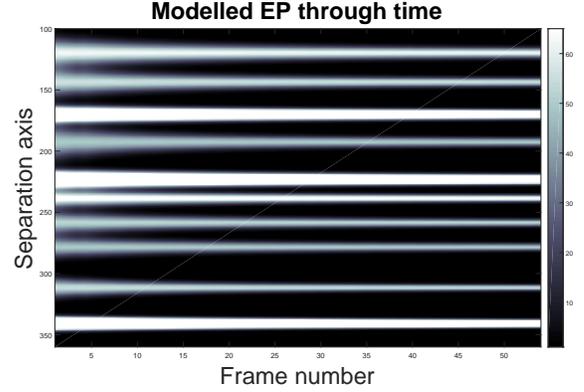


Fig. 11. Estimated fused EPMT from realigned EPMTs using a GTT model.

fragments, ladders and sparse examples are analysed.¹ DNA ladders are chosen as they have a good distribution of peak positions with respect to separation, and therefore allow a thorough assessment of the change in resolution [2], [5], [6]. Moreover, the resolution improvement for a ladder can be used to deduce the resolution improvement for an arbitrary trace. The fragments chosen are ones in which peaks have separated by the standard EP time, and therefore have no births of peaks. This aids the resolution analysis, since if peaks appear as EP continues, which is common in practice, it is necessary to track and identify individual peaks prior to deconvolution. Tracking the birth, evolution, and death of peaks is an important signal processing problem, but is beyond the scope of this paper. The results presented the realignment fusion methods which are compared with the benchmark standard EP.

A. Realignment Results

In the linear realignment method, a constant baseline is removed prior to realignment, as discussed in Section III-B1, while for the affine realignment, a quadratic polynomial models the baseline as in Section II-A. As an example of the realignment method in Section III-B, Fig. 12 shows the realignment of dataset F (see TABLE II), which is a DNA ladder labelled ‘‘Gel 206’’. In this experimental setup, snapshots are taken at multiples of 30 seconds for 1860 seconds. The standard EP time for this example is 420 seconds, corresponding to the 13th frame. The scaling Δ , shift Σ , gain α_0 , and baseline coefficients $\{\alpha_q\}_1^Q$ are estimated by minimising (24) with respect to the 13th frame, as described in Section III-B.

Anti-clockwise from the top-right, Fig. 12 shows: the reference frame (frame 13) in dotted line, and the EPMT for frame 29 in solid; the realigned trace in the top-left; a contour and surface plot of the negative cost-function as a function of shift and scale which is maximised at the best alignment. Fig. 12 shows good alignment of the signals, although some individual peaks are misaligned, likely due to non-ideal effects in the EP process. The realignment of all the frames for this dataset (F) is shown in Fig. 13, where it is clear that many peaks are aligned perfectly, while others have statistical variation. Some small peaks to the right of Fig. 13 are due to cassette noise.

¹These data sequences will be available at <http://datashare.is.ed.ac.uk/>.

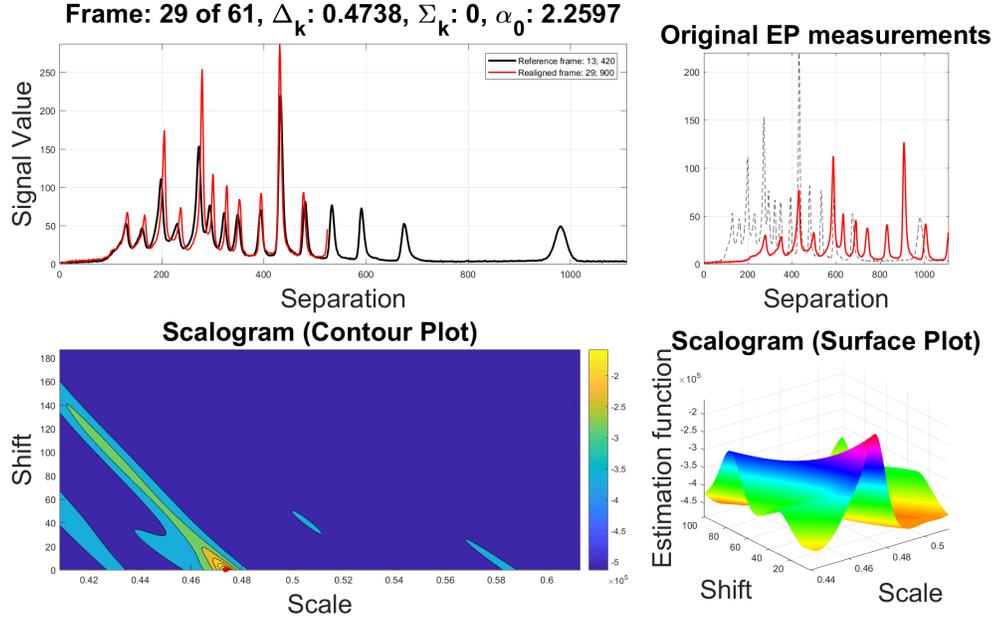


Fig. 12. The top-left figure shows the alignment of two frames, with the original EPMTs for the reference and projected frames shown in the top-right figure. The lower-left figure shows the negative error surface of equation (24) as a function of shift-in-separation and scale-in-separation, also called the scalogram. A peak in the scalogram indicates the best fit realignment parameters, and in this case there is a distinct peak at zero shift and a scale-in-separation of 0.47.

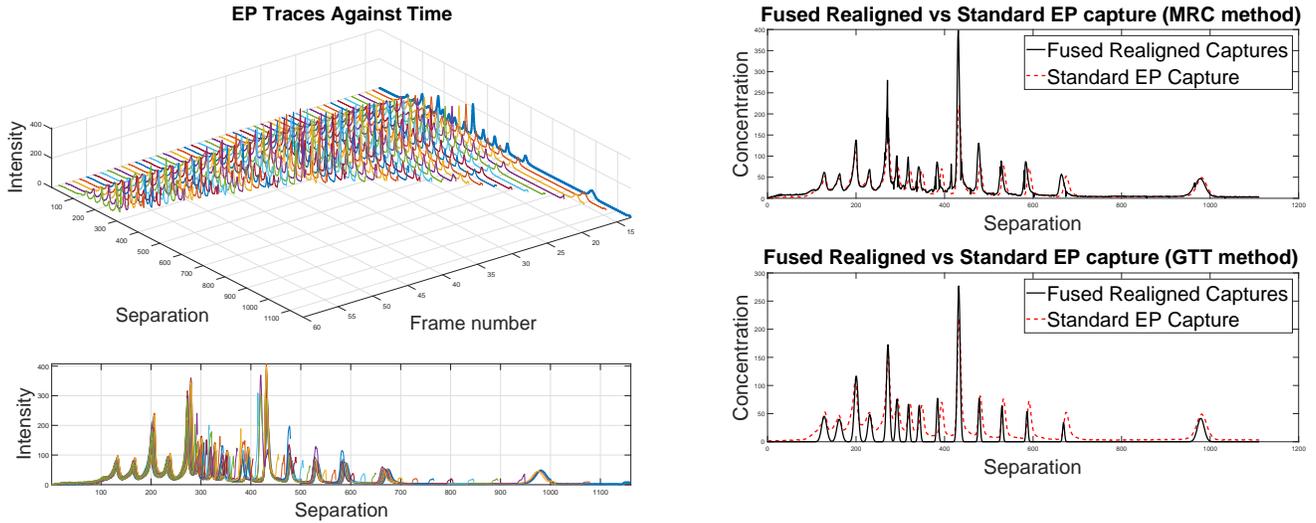


Fig. 13. Realigned electrophoresis traces using LSE and the affine model.

Fig. 14. Comparison of standard and fused EPMTs (Sections IV-A and IV-B).

B. Fusion Results

Continuing with dataset F , Fig. 14 shows the fused traces, as compared with the reference EPMT, using the MRC and GTT methods in Section IV. Note that in GTT, the baseline is removed from the plot, and thus the lower plot of Fig. 14 shows fractional peaks only. Qualitatively, the realigned-and-fused trace shows improved SNR and improved resolution, as indicated by the sharper peaks. The MRC signal (top graph) shows a less smooth signal than the GTT method, as expected when comparing a non-parametric and parametric method.

C. Resolution Improvement

To quantify resolution improvement from MSI, results from 7 datasets are presented in TABLE II. The table shows the number of frames, sampling time, number of peaks, and overall improvement given by (2) across all frames, expressed as a ratio, for the two alignment and two fusion methods. The percentage improvement per additional EP minute is also indicated, giving a fair quantitative assessment as the total EP time is different for each dataset. The resolution is estimated as per Section II-A, although for robustness, if the least-squares curve fitting fits a particular peak poorly, this peak is excluded from the calculation to prevent outliers biasing the results. This typically occurs for a single peak for the datasets with a large



Fig. 15. Resolution improvement vs time of MRC traces (Section IV-A).



Fig. 16. Resolution improvement vs time of GTT traces (Section IV-B).

number of peaks (datasets F and G). Figs 16 and 15 indicate the spread of resolution estimates across all peaks as a function of time. GTT has a substantially larger increase in resolution, although the spread of intra-peak resolution is larger than for MRC. Nevertheless, both techniques give a clear resolution improvement across all peaks, with 35% for MRC and 100% for GTT. TABLE II indicates that GTT usually outperforms MRC, except for datasets B and D , where MRC still has a resolution improvement of over 8% per minute. Moreover, the linear realignment model typically outperforms the affine model, indicating that the simple histogram baseline removal is effective compared with the polynomial baseline model. In MRC, peaks which leave the FOV earlier in the EP process (the “lower peaks”) won’t see an improvement in resolution compared with the “upper peaks” as there is no new information as the EP process continues. However, in GTT the FWHM of each peak is regressed and projected forward, which explains the improvement in resolution for all peaks.

VI. CONCLUSIONS

This paper presents a novel multi-capture snapshot imaging technique for GEP using estimation theory to realign and

fuse multiple waveforms. The proposed signal processing algorithms facilitate an improvement of the snapshot imaging approach which is typically simpler to implement in automated electrophoresis systems than the finish-line method. A further advantage of MSI is that static noise will be diminished, since static noise becomes non-coherent after realignment. Several realignment and fusion algorithms are proposed and compared. Improvement in resolution is quantified by deconvolving the EPMTs; the proposed technique improves resolution by approximately 10% per minute of additional electrophoresis, equivalent to doubling resolution over a typical run.

ACKNOWLEDGMENT

This work funded by the Royal Academy of Engineering’s Industrial Secondment Scheme, ISS1516\8\33, and the EP-SRC Impact Acceleration Account (University of Edinburgh).

REFERENCES

- [1] K. D. Dorfman *et al.*, “Beyond gel electrophoresis: Microfluidic separations, fluorescence burst analysis, and DNA stretching,” *Chemical Reviews*, vol. 113, no. 4, pp. 2584–2667, Nov. 2013.
- [2] L. Gutzweiler *et al.*, “Open microfluidic gel electrophoresis: Rapid and low cost separation and analysis of DNA at the nanoliter scale,” *Electrophoresis*, vol. 38, no. 13–14, pp. 1764–1770, Apr. 2017.
- [3] M. Dawod *et al.*, “Recent advances in protein analysis by capillary and microchip electrophoresis,” *Analyst*, vol. 142, pp. 1847–1866, 2017.
- [4] C. Birch and J. P. Landers, “Electrode materials in microfluidic systems for the processing and separation of DNA: A mini review,” *Micromachines*, vol. 8, no. 3, 2017.
- [5] J. Li *et al.*, “Enhanced resolution of DNA separation using agarose gel electrophoresis doped with graphene oxide,” *Nanoscale Research Letters*, vol. 11, Sep. 2016.
- [6] M. Zarei *et al.*, “Nanoparticle improved separations: From capillary to slab gel electrophoresis,” *TrAC Trends in Analytical Chemistry*, vol. 86, pp. 56 – 74, 2017.
- [7] R. C. Lo and V. M. Ugaz, “Microchip DNA electrophoresis with automated whole-gel scanning detection,” *Lab Chip*, vol. 8, pp. 2135–2145, 2008.
- [8] J. C. Sutherland *et al.*, “A comparison of electrophoretic resolution for snapshot and finish-line imaging,” *Analytical Biochemistry*, vol. 239, no. 2, pp. 136–144, Aug. 1996.
- [9] T. Techanukul *et al.*, “CE-based sample quality assessment prior to 2-D gel electrophoresis: Towards the standardization of gel-based proteomics,” *Journal of Separation Science*, vol. 33, no. 16, pp. 2536–2546, 2010.
- [10] A. A. Dahab *et al.*, “Determination of aristolochic acid I and II in traditional chinese medicine by HPCE with label free intrinsic imaging,” *Chromatographia*, vol. 70, no. 3, pp. 467–473, Aug 2009.
- [11] P. Maurye *et al.*, “Electrophoresis-staining apparatus for dna agarose gels with solution exchange and image acquisition,” *Instrumentation Science & Technology*, vol. 45, no. 1, pp. 49–61, 2017.
- [12] B. Mayer, “How to increase precision in capillary electrophoresis,” *Journal of Chromatography A*, vol. 907, no. 1, pp. 21 – 37, 2001.
- [13] J. W. Yoon *et al.*, *Proceedings of Bioinformatics Research and Development: First International Conference*. Germany: Springer Berlin Heidelberg, Mar. 2007, ch. Bayesian Inference for 2D Gel Electrophoresis Image Analysis, pp. 343–356.
- [14] N. Kaabouch *et al.*, “An analysis system for DNA gel electrophoresis images based on automatic thresholding an enhancement,” in *Electro/Information Technology, 2007 IEEE International Conference on*, May 2007, pp. 26–31.
- [15] V. B. D. Marco and G. G. Bombi, “Mathematical functions for the representation of chromatographic peaks,” *Journal of Chromatography A*, vol. 931, no. 1–2, pp. 1–30, Oct. 2001.
- [16] J. R. Hopgood, “Improved resolution of chromatographic peak analysis using multi-snapshot imaging,” in *24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 1533–1537.
- [17] E. H. G. Yousif *et al.*, “A new approach for multi-dimensional signal processing and modelling for signals from gel electrophoresis,” in *24th European Signal Processing Conference (EUSIPCO)*, Aug 2016, pp. 682–686.

| Dataset ID | Dataset Name | Number of Frames | Sample Interval (Seconds) | Number of Peaks | Overall Resolution Improvement | | | | Percentage Resolution Improvement Per Minute | | | |
|------------|---------------|------------------|---------------------------|-----------------|--------------------------------|--------------|--------|--------------|--|---------------|--------|---------------|
| | | | | | Linear | | Affine | | Linear | | Affine | |
| | | | | | MRC | GTT | MRC | GTT | MRC | GTT | MRC | GTT |
| A | Test Ladder | 24 | 30 | 15 | 1.742 | 1.803 | 1.705 | 1.838 | 6.439% | 6.975% | 6.125% | 7.276% |
| B | Ladder | 5 | 60 | 10 | 2.101 | 1.6 | 1.426 | 1.605 | 13.74% | 7.489% | 5.318% | 7.546% |
| C | Ladder (HFR) | 67 | 5 | 10 | 1.431 | 1.67 | 1.396 | 1.656 | 7.818% | 12.14% | 7.187% | 11.89% |
| D | Smear | 13 | 60 | 4 | 2.069 | 1.225 | 2.026 | 1.228 | 8.898% | 1.872% | 8.536% | 1.895% |
| E | Gel 183 | 22 | 60 | 10 | 1.597 | 3.696 | 1.534 | 3.608 | 2.841% | 12.83% | 2.542% | 12.41% |
| F | Gel 206 (80V) | 47 | 30 | 15 | 1.342 | 1.966 | 1.279 | 1.942 | 1.484% | 4.196% | 1.214% | 4.091% |
| G | Gel 206 (90V) | 54 | 30 | 15 | 1.706 | 2.982 | 1.514 | 2.78 | 2.661% | 7.476% | 1.937% | 6.714% |

TABLE II

RESOLUTION IMPROVEMENT FOR DATASETS A TO G. DATASET C IS RECORDED AT A HIGH-FRAME RATE (HFR), THE OTHERS AT A LOW-FRAME RATE.

- [18] A. Felinger, *Data Analysis and Signal Processing in Chromatography*. Elsevier Science, May 1998.
- [19] T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.
- [20] J. S. Morris *et al.*, "Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum," *Bioinformatics*, vol. 21, no. 9, pp. 1764–1775, 2005.
- [21] J. E. Carlson and F. F. Sjöberg, "Simultaneous maximum likelihood estimation of time delay and time scaling," in *Proceedings of the 6th Nordic Signal Processing Symposium*, Jun. 2004, pp. 260–263.
- [22] M. Kneen and H. Annegarn, "Algorithm for fitting XRF, SEM and PIXE X-ray spectra backgrounds," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 109, pp. 209 – 213, 1996.
- [23] X. Ning *et al.*, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics and Intelligent Laboratory Systems*, vol. 139, pp. 156 – 167, 2014.