(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 February 2002 (14.02.2002)

PCT

(10) International Publication Number
WO 02/13122 A2

(51) International Patent Classification[7]: G06F 19/00, G01N 27/447, C12Q 1/68

(21) International Application Number: PCT/GB01/03286

(22) International Filing Date: 20 July 2001 (20.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0019499.3 8 August 2000 (08.08.2000) GB

(71) Applicant *(for all designated States except US)*: **DIAMOND OPTICAL TECHNOLOGIES LIMITED** [GB/GB]; 6 St. Andrew Street, London EC4A 3LX (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **SIDERIS, Dimitrios** [GR/GB]; 38 Solar Court, Etchingham Park Road, London N3 2DZ (GB).

(74) Agents: **MAGGS, Michael, Norman** et al.; Kilburn & Strode, 20 Red Lion Street, London WC1R 4PJ (US).

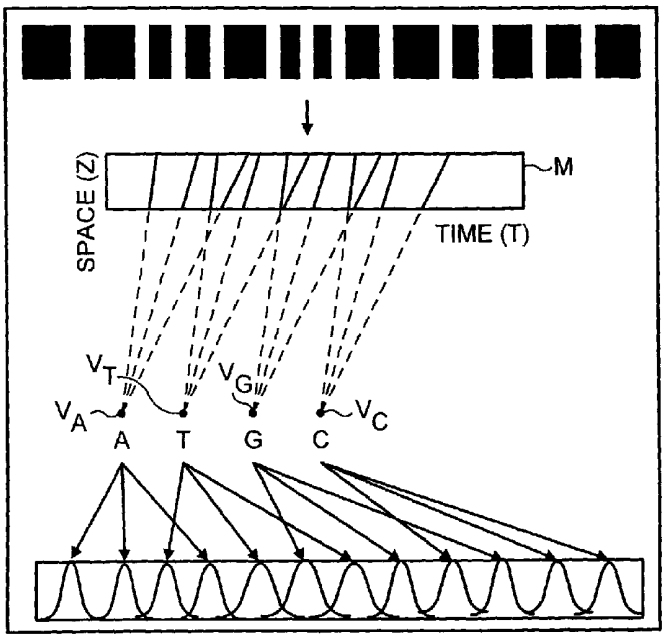
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: SYSTEM AND METHOD



(57) Abstract: A system and method of characterising, in particular sequencing, polymeric samples, the system comprising: a space-time map generator for generating a space-time map of points representative of the signal peaks of signals detected at a plurality of spaced positions as passed by the migrating components of a plurality of separately-provided sample plugs; and a vertex finder for identifying vertices from the space-time map, with a single vertex being identified for the components of each separately-provided sample plug and the velocities of the sample components being determinable from the points in the space-time map.

WO 02/13122 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SYSTEM AND METHOD

The present invention relates to a system for and method of characterising, in particular sequencing, polymeric samples, in particular, but not exclusively, as measured electrophoretically. The technique of the present invention is the MITSO™ characterisation technique.

Electrophoretic separation techniques are separation techniques in which the components of a sample plug are separated in a separation column by the differences in the migration rates of those sample components on the application of an electric field therealong, where absorption, fluoroescence, electrochemistry, conductivity, radioactivity and mass spectrometry can be all used to detect the electrophoretic separation.

By way of example, DNA samples are commonly characterised by Sanger sequencing. Whilst Sanger sequencing is in common use, the technique is limited in a number of respects, not least because the DNA bands have to be labelled according to base pair termination.

It is an aim of the present invention to provide an improved system for and method of characterising, in particular sequencing, polymeric samples, such as DNA samples, which can be used with both labelled and non-labelled sample components.

Accordingly, the present invention provides a system for characterising, in particular sequencing, polymeric samples, comprising: a space-time map generator for generating a space-time map of points representative of the signal peaks of signals detected at a plurality of spaced positions as passed by the migrating components of a plurality of separately-provided sample plugs; and a vertex finder for identifying vertices from the space-time map, with a single vertex being identified for the components of each separately-provided sample plug and the velocities of the sample components being determinable from the points in the space-time map.

Preferably, the system further comprises a velocity sorter for determining the nominal velocities associated with the signal peaks in the signals and grouping those signal peaks into sets according to nominal velocity.

Preferably, the space-time map generator is configured to utilise a corrected time component in generating the space-time map according to a function of the electric current variation.

More preferably, the correction is according to the function $t_c = \int I_0/I(t')dt'$ in the range of 0 to t, where t is the measured time, t_c is the corrected time, I is the measured current and I_0 is the reference current.

Preferably, the space-time map generator is an equiphase space-time map generator for generating an equiphase space-time map of equiphase points.

More preferably, the equiphase space-time map generator is configured to transform each data set into a set of local slopes and determine the local minima as the minimum absolute local derivatives.

In one embodiment the components are non-labelled.

In another embodiment the components are labelled.

Preferably, the components are migrated through a channel.

More preferably, the channel comprises a separation channel through which the components are electrophoretically driven.

Preferably, the sample plugs respectively comprise DNA bands having one of the base pair terminations.

The present invention also extends to an electrophoresis apparatus including the above-described system.

The present invention also provides a method of characterising, in particular sequencing, polymeric samples, comprising the steps of: generating a space-time map of points representative of the signal peaks of signals detected at a plurality of spaced positions as passed by the migrating components of a plurality of separately-provided sample plugs; and identifying vertices from the space-time map, with a single vertex being identified for the components of each separately-provided sample plug and the velocities of the sample components being determinable from the points in the space-time map.

Preferably, the method further comprises the steps of determining the nominal velocities associated with the signal peaks in the signals and grouping those signal peaks into sets according to nominal velocity.

Preferably, a time component corrected according to a function of the electric current variation is utilised in generating the space-time map.

More preferably, the correction is according to the function $t_c = \int I_0/I(t')dt'$ in the range of 0 to t, where t is the measured time, t_c is the corrected time, I is the measured current and I_0 is the reference current.

Preferably, the space-time map is an equiphase space-time map of equiphase points.

More preferably, the equiphase points are determined by transforming each data set into a set of local slopes and determining the local minima as the minimum absolute local derivatives.

In one embodiment the components are non-labelled.

In another embodiment the components are labelled.

Preferably, the components are migrated through a channel.

More preferably, the channel comprises a separation channel through which the components are electrophoretically driven.

In one embodiment the sample plugs are provided at spaced time intervals.

In another embodiment the sample plugs are provided at spaced positions.

Preferably, the sample plugs are provided at substantially the same time.

Preferably, the sample plugs respectively comprise DNA bands having one of the base pair terminations.

A preferred embodiment of the present invention will now be described hereinbelow by way of example only with reference to the accompanying drawings, in which:

Figure 1 illustrates the detector chip of an electrophoresis apparatus in accordance with a preferred embodiment of the present invention;

Figure 2 illustrates the analysis system of the apparatus of Figure 1;

Figure 3 illustrates a three-dimensional representation of the intensity-time signals of one component of a sample plug as detected at positions z_1, z_2, z_3 spaced along the separation channel of the apparatus of Figure 1;

Figure 4 illustrates the intensity-time signals of three components of a sample plug as detected at positions z_1, z_2, z_3 spaced along the separation channel of the apparatus of Figure 1;

Figure 5 illustrates a space-time map as generated from the intensity-time signals of Figure 4;

Figure 6 illustrates the velocity spectrum as determined from the vertexed space-time map of Figure 5; and

Figure 7 illustrates a space-time map as generated from the intensity-time signals from four separately-injected DNA sample plugs comprising DNA bands having different base pair terminations.

Figures 1 and 2 illustrate an electrophoresis apparatus in accordance with a preferred embodiment of the present invention.

The electrophoresis apparatus includes a detector chip 2 as microfabricated in a substrate chip, and an analysis system 3 for analysing the detection signals generated by the detector chip 2.

The detector chip 2 includes a separation channel 4, in this embodiment a meandering, gel-filled channel, through which the components of one or more sample plugs are in use driven by an applied electrophoretic voltage. The separation channel 4 has a length sufficient to allow separation of the components of the sample plugs. Preferably, the separation channel 4 has a width of from 25 to 100 μm and a length of from 20 to 300 mm. The separation channel 4 includes a plurality, in this embodiment first to fourth, spaced sample-injection ports 6, 8, 10, 12 through which sample plugs including a plurality of components, in this embodiment DNA bands having the respective base pair terminations A, T, G and C, are separately injected into the separation channel 4.

The detector chip 2 further includes a light source 14, in this embodiment a UV light source, disposed along a length of one side of the separation channel 4, and a detector 16 disposed along the length of the other side of the separation channel 4 to detect light transmitted through the separation channel 4, with the presence of the migrating

components being detected by the change in the detected light intensity as caused by absorbtion of the incident light. By detecting the sample components in this manner, the sample components need not necessarily be labelled. In this embodiment the detector 16 comprises a pixel detector array (PDA) which includes a plurality of pixels providing detecting elements for detecting the transmitted light at a plurality of positions z_1, z_2, z_3 spaced along the length of the separation channel 4 and outputting a plurality of signals S_1, S_2, S_3. For ease of description, the detector 16 is illustrated as including three detecting elements at three positions z_1, z_2, z_3. It will, however, be understood that in practice the detector 16 comprises a plurality of detecting elements at a plurality of positions z_1, z_2, z_3, ... , z_n, which each output a signal S_1, S_2, S_3, ... , S_n. In an alternative embodiment the detector 16 could be provided by a plurality of separate detectors each providing a detecting element. In another alternative embodiment labelled sample components could be used, such as sample components including fluorescent or radioactive labels, which labels would be detected by the detector 16.

The analysis system 3 comprises a data collector 18 for receiving the signals S_1, S_2, S_3 generated by the detector 16 and storing those signals S_1, S_2, S_3 as data sets, a velocity sorter 19 for determining the nominal velocities v_1, v_2, v_3 of the sample components associated with each of the signal peaks SP_1, SP_2, SP_3 of each of the signals S_1, S_2, S_3 and grouping those signal peaks SP_1, SP_2, SP_3 into sets according to nominal velocity, an equiphase space-time map generator 20 for generating an equiphase space-time map of equiphase points from the signal peaks SP_1, SP_2, SP_3 of the signals S_1, S_2, S_3, and a vertex finder 22 for identifying the vertices of the equiphase points of the grouped sets of signal peaks SP_1, SP_2, SP_3. In this embodiment the velocity sorter 19 is provided so as to be operable prior to the equiphase space-time map generator 20. In alternative embodiments the velocity sorter 19 could be provided so as to be operable after the space-time map generator 20 or the vertex finder 22.

Figure 3 is included for the purposes of illustration only and illustrates the signals S_1, S_2, S_3 as including only a single peak SP_1 from a single component of a single sample plug. In reality, however, the signals S_1, S_2, S_3 each include a plurality of signal peaks SP_{1-n}, SP_{1-n},

SP_{1-n}. Figure 4 illustrates the signals S_1, S_2, S_3 as including three signal peaks SP_1, SP_2, SP_3 from three components of a single sample plug.

The velocity sorter 19 is configured to determine the nominal velocities v_1, v_2, v_3 of the sample components associated with each of the signal peaks SP_1, SP_2, SP_3 in each of the signals S_1, S_2, S_3 and then group those signal peaks SP_1, SP_2, SP_3 into sets according to nominal velocity. The nominal velocities v_1, v_2, v_3 can be calculated as the positions z_1, z_2, z_3 of the detector elements are fixed and the elapsed time t is extractable from the signals S_1, S_2, S_3, where the nominal velocities can be expressed as $v_{1-n} = z_{1-n}/t$. By grouping the signal peaks SP_1, SP_2, SP_3 into sets according to nominal velocity, and hence sample component, subsequent analysis is facilitated as the data points associated with each sample component can be fitted without requiring the use of complex data extraction techniques. Velocity sorting is encompassed by our earlier WO-96/35946, the content of which is incorporated herein by reference.

The equiphase space-time map generator 20 is configured to determine the local minima of the signal peaks SP_1, SP_2, SP_3 in the signals S_1, S_2, S_3 detected at the detection positions z_1, z_2, z_3 and generate an equiphase map M in space-time dimensions from the determined local minima. Figure 5 illustrates the space-time map M generated from the local minima extracted from the signal peaks SP_1, SP_2, SP_3 of the signals S_1, S_2, S_3.

In this embodiment each electropherogram is transformed into a set of local slopes, where a triangular slope sequence defines a signal and the local extreme is the minimum absolute local derivative.

Also, in this embodiment the time component of the detected signals S_1, S_2, S_3 is corrected as a function of the integrated electric current variation. Owing to the variation of various factors in electrophoretic detection, the temperature being one of the most significant, the characteristics of the separation medium, in this embodiment a gel, are altered. Firstly, the resistivity of the gel changes, leading to variations in the potential difference between the electrodes and a given point in the gel and fluctuations in the electric current. Secondly,

the sieving properties of the gel change, affecting the mobility of the electrophoresed components. By monitoring the electric current, the time component of the space-time map M can be corrected as set out hereinbelow. Specifically, the time component is curved as a function of the integrated electric current variation.

The velocity of a sample component is:

$$v = dz/dt \qquad (1)$$

For a transformation of the measured time component to a corrected time component $t \rightarrow t_c$, it follows that $dt \rightarrow dt_c$ and $v \rightarrow v_c$. Thus:

$$v_c/v = dt/dt_c \qquad (2)$$

The transformation $v \rightarrow v_c$ can be defined as:

$$v_c/v = I(t)/I_0 \qquad (3)$$

where I is the measured current and I_0 is the reference current which corresponds to the frame where all velocities and time components are projected.

From equations (2) and (3), it follows:

$$dt_c = I_0/I(t)dt \quad \rightarrow \quad t_c = \int I_0/I(t')dt' \text{ for } 0 \text{ to } t \qquad (4)$$

The justification for the velocity transformation (3) is that the velocity is approximately proportional to the applied electric field, which in turn is proportional to the electric current in the separation channel 4. This correction factor has been found to work well for small current changes, with the integral of equation (4) providing for an accurate time transformation.

The vertex finder 22 is configured, in this embodiment by the use of rotational matrices, to identify the vertices V of the equiphase points of the grouped sets of signal peaks SP_1, SP_2, SP_3 as determined by the equiphase space-time map generator 20, where the components of each injected sample plug have a common vertex V by virtue of being time and/or spatially separated in the space-time dimension. All of the sample components injected in a single sample plug are uniquely identified by a single vertex V in space-time co-ordinates, thus allowing for the identification of the sample components from each of a plurality of separately-provided sample plugs. Figure 5 illustrates the vertex V as determined from the generated space-time map M. This space-time map includes only a single vertex V as all of the components were provided in a single sample plug.

By using each vertex V as a constraint to extract the velocity spectrum of the sample components, the resolution is approximately proportional to $\sqrt{n}$, where n is the number of components. In this way, the velocity of one component is calculated using the velocities of all of the other components from the same sample plug, and thus, as the number of components in a sample increases, the resolution of the analysis increases accordingly. Such space parameterisation which results in multiple vertex formation in the form of intensity enhanced regions in space-time co-ordinates is particularly suited to the cases of multiple sample injections and multiple column correlation. The power of this technique has been demonstrated on DNA samples which include large numbers of fragments (>100) having lengths of one base pair difference, thereby providing a sequencing technique having a greatly extended dynamic range.

From the determination of the vertices V in the space-time map M, high resolution of the electrophoresis data is achieved, allowing accurate determination of the velocities of the sample components as illustrated in Figure 6.

Use of the above-described electrophoresis apparatus to sequence DNA samples having the base pair terminations A, T, G and C will now be described hereinbelow.

In use, four sample plugs comprising DNA bands having different length and one of the base pair terminations A, T, G and C are separately introduced into the ports 6, 8, 10, 12 of the separation channel 4, and electrophoretically driven therealong. In one mode of use, the sample plugs are introduced simultaneously into the ports 6, 8, 10, 12 which are spatially separated along the separation channel 4. In another mode of use, the sample plugs are introduced sequentially into one of the ports 6, 8, 10, 12 so as to be time spaced. The signals S_1, S_2, S_3, ... , S_n detected by the detector 16 as the DNA bands pass the detecting elements at the detecting positions z_1, z_2, z_3, ... , z_n are collected by the data collector 18. The velocity sorter 19 then determines the nominal velocities v_1, v_2, v_3, ..., v_n of the sample components associated with each of the signal peaks SP_1, SP_2, SP_3, ..., SP_n of the signals S_1, S_2, S_3, ... , S_n and groups those signal peaks SP_1, SP_2, SP_3, ..., SP_n into sets according to nominal velocity. The equiphase space-time map generator 20 then determines the local minima of the signal peaks SP_1, SP_2, SP_3, ..., SP_n of the signals S_1, S_2, S_3, ... , S_n, and generates an equiphase space-time map M. The vertex finder 22 then identifies the vertices V_A, V_T, V_G, V_C of the determined local minima for each of the grouped sets of signal peaks SP_1, SP_2, SP_3, ..., SP_n. In this embodiment the space-time map M includes four vertices V_A, V_T, V_G, V_C as four sample plugs were separately injected into the separation channel 4, each being attributable to DNA bands having one of the base pair terminations A, T, G and C. In this way, the DNA sample can be sequenced, with the lengths of the DNA bands being determined from the migration velocities.

Finally, it will be understood that the present invention has been described in its preferred embodiment and can be modified in many different ways without departing from the scope of the invention as defined by the appended claims.

CLAIMS

1. A system for characterising, in particular sequencing, polymeric samples, comprising:
a space-time map generator for generating a space-time map of points representative of the signal peaks of signals detected at a plurality of spaced positions as passed by the migrating components of a plurality of separately-provided sample plugs; and
a vertex finder for identifying vertices from the space-time map, with a single vertex being identified for the components of each separately-provided sample plug and the velocities of the sample components being determinable from the points in the space-time map.

2. The system of claim 1, further comprising a velocity sorter for determining the nominal velocities associated with the signal peaks in the signals and grouping those signal peaks into sets according to nominal velocity.

3. The system of claim 1 or 2, wherein the space-time map generator is configured to utilise a corrected time component in generating the space-time map according to a function of the electric current variation.

4. The system of claim 3, wherein the correction is according to the function $t_c = \int I_0/I(t')dt'$ in the range of 0 to t, where t is the measured time, t_c is the corrected time, I is the measured current and I_0 is the reference current.

5. The system of any of claims 1 to 4, wherein the space-time map generator is an equiphase space-time map generator for generating an equiphase space-time map of equiphase points.

6. The system of claim 5, wherein the equiphase space-time map generator is configured to transform each data set into a set of local slopes and determine the local minima as the minimum absolute local derivatives.

7. The system of any of claims 1 to 6, wherein the components are non-labelled.

8. The system of any of claims 1 to 6, wherein the components are labelled.

9. The system of any of claims 1 to 8, wherein the components are migrated through a channel.

10. The system of claim 9, wherein the channel comprises a separation channel through which the components are electrophoretically driven.

11. The system of any of claims 1 to 10, wherein the sample plugs respectively comprise DNA bands having one of the base pair terminations.

12. A method of characterising, in particular sequencing, polymeric samples, comprising the steps of:
generating a space-time map of points representative of the signal peaks of signals detected at a plurality of spaced positions as passed by the migrating components of a plurality of separately-provided sample plugs; and
identifying vertices from the space-time map, with a single vertex being identified for the components of each separately-provided sample plug and the velocities of the sample components being determinable from the points in the space-time map.

13. The method of claim 12, further comprising the steps of determining the nominal velocities associated with the signal peaks in the signals and grouping those signal peaks into sets according to nominal velocity.

14. The method of claim 12 or 13, wherein a time component corrected according to a function of the electric current variation is utilised in generating the space-time map.

15. The method of claim 14, wherein the correction is according to the function $t_c = \int I_0/I(t')dt'$ in the range of 0 to t, where t is the measured time, t_c is the corrected time, I is the measured current and I_0 is the reference current.

16. The method of any of claims 12 to 15, wherein the space-time map is an equiphase space-time map of equiphase points.

17. The method of claim 16, wherein the equiphase points are determined by transforming each data set into a set of local slopes and determining the local minima as the minimum absolute local derivatives.

18. The method of any of claims 12 to 17, wherein the components are non-labelled.

19. The method of any of claims 12 to 17, wherein the components are labelled.

20. The method of any of claims 12 to 19, wherein the components are migrated through a channel.

21. The method of claim 20, wherein the channel comprises a separation channel through which the components are electrophoretically driven.

22. The method of any of claims 12 to 21, wherein the sample plugs are provided at spaced time intervals.

23. The method of any of claims 12 to 21, wherein the sample plugs are provided at spaced positions.

24. The method of claim 23, wherein the sample plugs are provided substantially at the same time.

25. The method of any of claims 12 to 24, wherein the sample plugs respectively comprise DNA bands having one of the base pair terminations.

26. An electrophoresis apparatus including the system of any of claims 1 to 11.

27. A system for characterising, in particular sequencing, polymeric samples substantially as hereinbefore described with reference to the accompanying drawings.

28. A method of characterising, in particular sequencing, polymeric samples substantially as hereinbefore described with reference to the accompanying drawings.

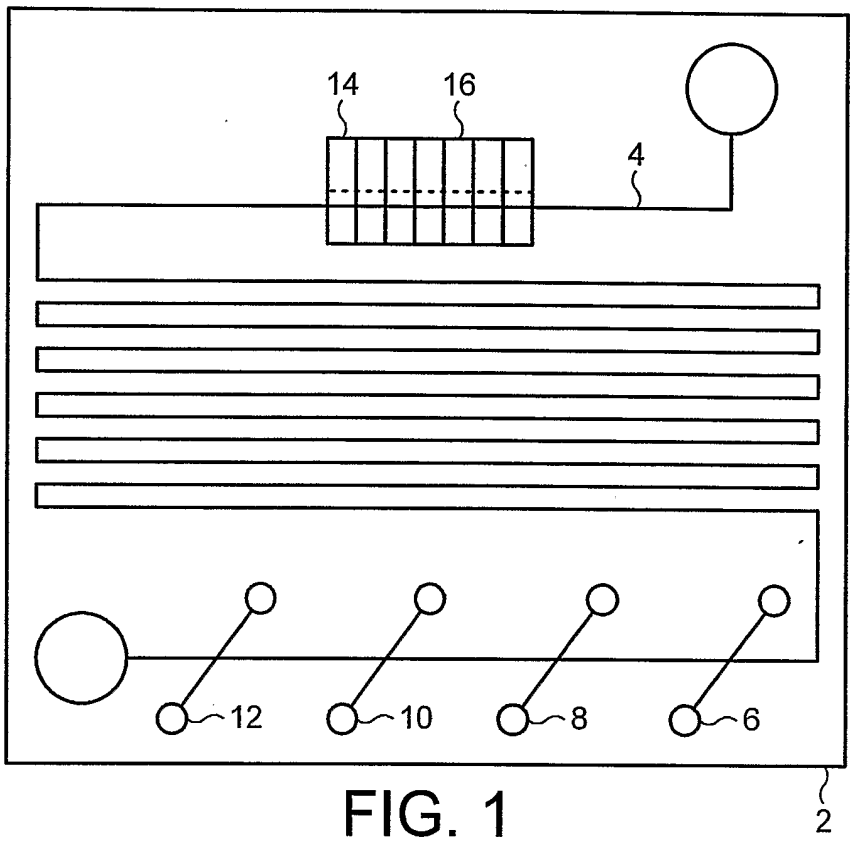

14
16
4
12
10
8
6
2

FIG. 1

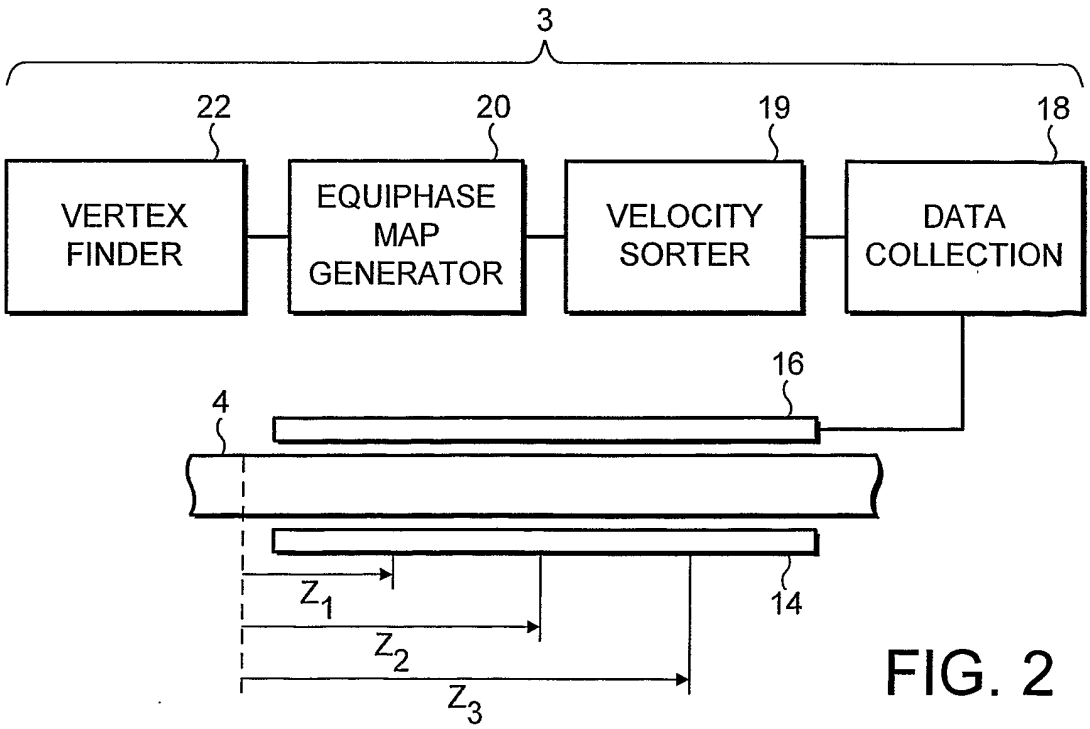

3
22
20
19
18
VERTEX FINDER
EQUIPHASE MAP GENERATOR
VELOCITY SORTER
DATA COLLECTION
16
4
14
Z_1
Z_2
Z_3

FIG. 2

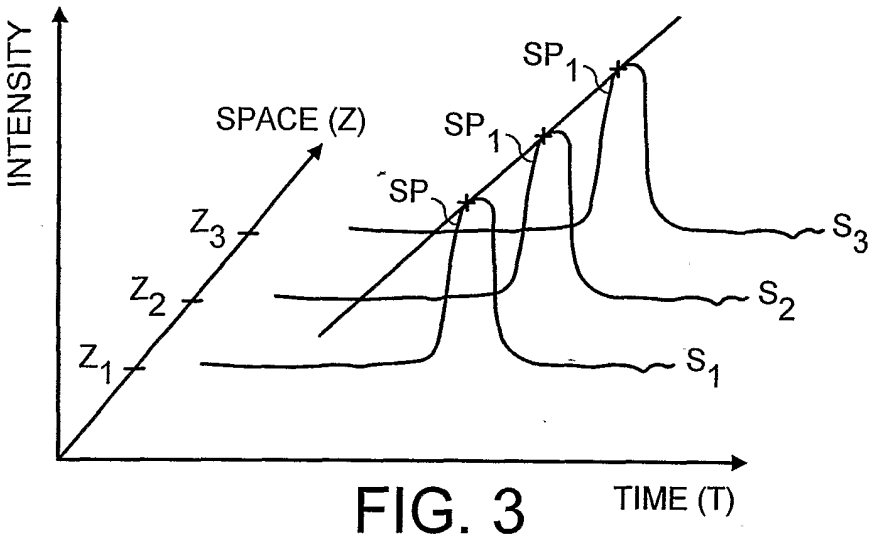


FIG. 3

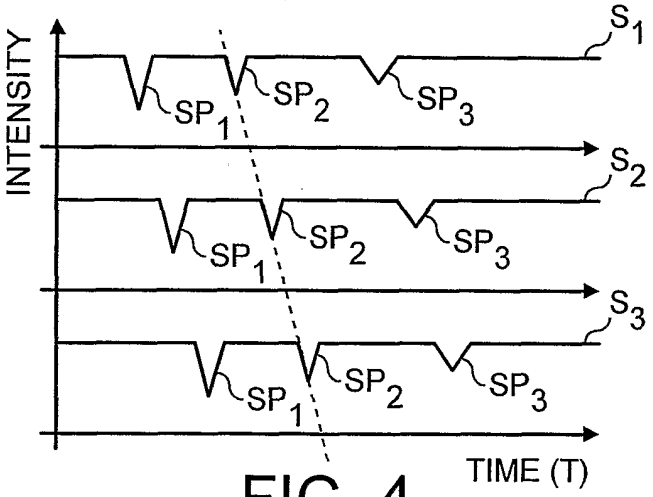


FIG. 4

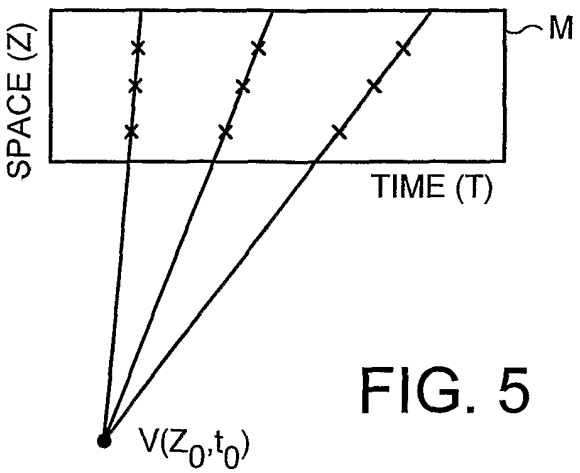


FIG. 5

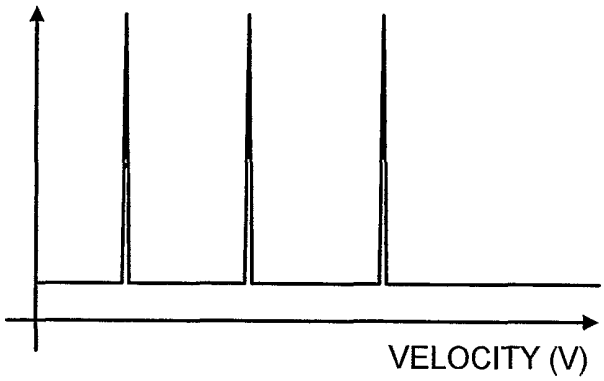


FIG. 6

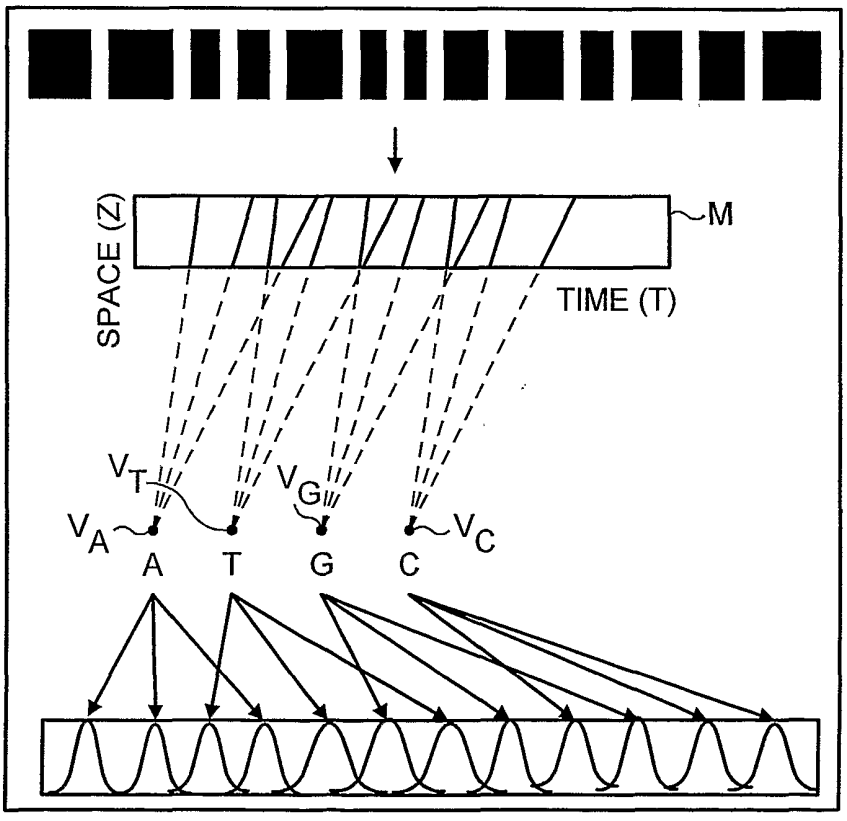


FIG. 7