



(51) International Patent Classification:

G01J 3/28 (2006.01) G01N 21/65 (2006.01)
G01J 3/44 (2006.01)

(21) International Application Number:

PCT/GB2017/052923

(22) International Filing Date:

29 September 2017 (29.09.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1616787.6 03 October 2016 (03.10.2016) GB

(71) Applicant: RENISHAW PLC [GB/GB]; New Mills, Wotton-under-Edge, Gloucestershire GL12 8JR (GB).

(72) Inventor: EDGERTON, Mark; Renishaw plc, New Mills, Wotton-under-Edge, Gloucestershire GL12 8JR (GB).

(74) Agent: MATTHEWS, Paul et al.; Renishaw plc, Patent Department, New Mills, Wotton-under-Edge, Gloucestershire GL12 8JR (GB).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: APPARATUS AND METHOD FOR REMOVING NOISE FROM MEASURED DATA

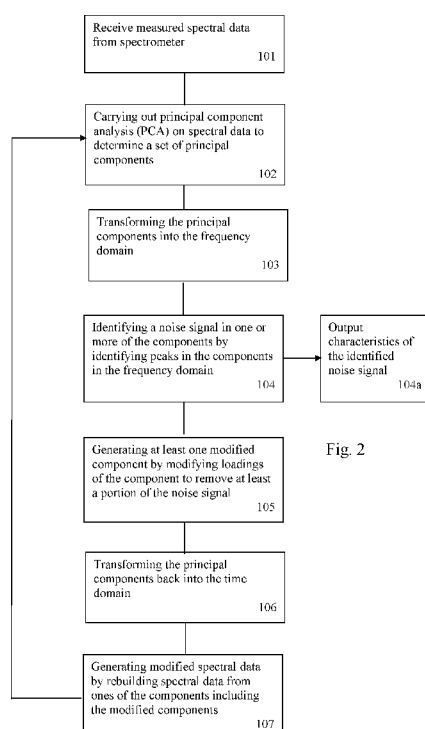


Fig. 2

(57) Abstract: This invention concerns a method and apparatus for removing noise from measured data, such as spectral data. The method comprises receiving (101) measured data comprising a plurality of sequential observations of discrete variables, carrying out (102) a method of blind signal separation to separate the measured data into a plurality of components, generating (105) at least one modified component by removing at least a portion of a noise signal present in at least one of the components and generating (107) modified data from ones of the components including the at least one modified component.



APPARATUS AND METHOD FOR REMOVING NOISE FROM MEASURED DATA

Field of the Invention

5 This invention relates to measurement apparatus and methods. It is particularly useful in spectroscopy, such as Raman, narrow-line photoluminescence, fluorescence, cathode-luminescence, UV visible (UV Vis), nuclear magnetic resonance (NMR), mid infra-red (mid-IR) or near infra-red (NIR) spectroscopy.

10 Background to the Invention

The Raman Effect is the inelastic scattering of light by a sample. In Raman Spectroscopy, a sample is irradiated by monochromatic laser light and the scattered light is then dispersed by a dispersive device, such as a diffraction
15 grating, e.g. in a monochromator, to generate a spectrum called a Raman spectrum. The Raman spectrum is detected by a detector such as a charge-coupled device (CCD). Examples of Raman spectroscopy apparatus are known from US Patents Nos. 5,442,438 and 5,510,894, which are incorporated herein by reference.

20

Different chemical compounds have different characteristic Raman spectra. Accordingly, the Raman effect can be used to analyse chemical compounds present in a sample.

25 The detected spectrum comprises the Raman spectrum together with noise generated by the system and the surrounding environment. In particular, electronic switching within the Raman apparatus and in surrounding electronic equipment can result in noise in the detected spectrum.

30 It is known to provide Raman systems for mapping a sample, for example as described in PCT/GB2016/050630. As the speed of mapping increases, a ratio of noise to the Raman signal increases.

It is desirable to provide a system for removing noise from the detected spectrum/spectra. The accurate removal of noise from the detected spectrum/spectra is particularly desirable in high speed mapping applications
5 which have a high noise to Raman signal ratio.

Summary of Invention

According to a first aspect of the invention there is provided a method of
10 removing noise from measured data, the method comprising receiving measured data comprising a plurality of sequential observations of discrete variables, carrying out a method of blind signal separation (also referred to as “blind source separation”) to separate the measured data into a plurality of components, generating at least one modified component by removing at least a portion of a
15 noise signal present in at least one of the components and generating modified data from ones of the components including the at least one modified component.

The noise may be more easily identified and removed from components of the blind signal separation method than the raw measured data, as such methods
20 attempt to separate a set of source signals, such as a source of noise, from a set of mixed signals, e.g. a plurality of spectra. This may be particularly the case for sources of fixed pattern and burst noise, which may have readily identifiable noise signatures. For example, fixed pattern noise is likely to result in at least one component describing a sinusoid and/or peaks at certain frequencies in the
25 frequency domain. It has been found, particularly for systems with a high noise to signal ratio, a component of the blind signal separation method can comprise both noise signals and useful information pertaining to the property being measured (referred to herein as the “measurement signal”). Therefore, to rebuild the measured data without the components that contain noise can result in portions of
30 the measurement signal being lost. It has been found that by modifying, rather than removing, a component identified as containing a noise signal, loss of useful portions of the measurement signal can be mitigated whilst noise is reduced.

Furthermore, the method of removing noise from the measured data is scalable with readout speed. In particular, a position of the measurement signal in the frequency domain is dependent on the spectral resolution of the spectroscopy/metrology apparatus and independent of the readout speed.

In one embodiment, the measured data is spectral data comprising a plurality of spectra captured sequentially using a spectroscopy system. In such an embodiment, the method comprises carrying out the method of blind signal separation to separate the spectral data into the plurality of components, generating the at least one modified component by removing at least the portion of the noise signal present in at least one of the components and generating modified spectral data from ones of the components including the at least one modified component. The plurality of spectra may be obtained from multiple points on a sample (so called mapping) and/or from a plurality of samples, for example which are sequentially analysed by the spectrometer as part of a manufacturing process.

In another embodiment, the measured data may be data from a metrology apparatus, for example a coordinate positioning machine (such as a machine tool or coordinate measuring machine (CMM)), a measurement probe (such as a touch trigger probe, scanning probe or optical probe), a rotary probe head, a sensor of an additive manufacturing (AM) machine (for example, sensors in the AM machine for monitoring a melt pool through an optical train of an optical scanner and/or acoustic sensors), an encoder system (such as an interferometric, magnetic or optical encoder system), a spatial laser measurement system and/or the like.

The method of blind signal separation may comprise principal component analysis. Principal component analysis generates a set of principal components ordered such that the first principal component describes a coordinate that gives rise to the greatest variance in the measured data, with each succeeding component having the next highest possible variance under the constraint that it is orthogonal (independent) to the preceding components. For measured data

comprising a high noise to measurement signal ratio, a noise signal can be present in a high order principal component, as a significant proportion of the variance in the measured data can be apportioned to noise. A high order principal component is likely to also contain portions of the measurement signal valuable for later analysis. The method may comprise modifying one of the principal components to remove at least a portion of the noise signal identified as present in the principal component and generating modified data from ones of the principal components including the modified principal component. Modification of the principal component may comprise modifying loadings of the principal components identified as comprising noise.

Other suitable methods of blind signal separation may be used, such as singular value decomposition (SVD), factor analysis, canonical correlation analysis, independent component analysis, multivariate curve resolution alternating least squares (MCR-ALS), and non-negative matrix factorization.

The noise signature(s) may be particular patterns, such as sinusoids, identifiable in the components, certain known operating frequencies of electronic components in the spectroscopy apparatus/metrology apparatus/surrounding environment, and/or signatures identified by a user. The signatures may be determined empirically, for example by experiments carried out by a user, the user inputting the determined noise signatures into an apparatus for carrying out the method, or the apparatus may be pre-programmed with typical noise signatures, for example as determined from theory and/or operating frequencies set for electronic components of the spectroscopy apparatus/metrology apparatus.

The known operating frequency may be an operating frequency of a switched-mode power supply and/or a camera that captures the measured data, such as a readout frequency. If possible, devices of the spectroscopy/metrology apparatus, which result in noise in the spectral data, may be arranged to operate at frequencies away from the frequencies at which the measurement signal appears in the components of the blind signal separation in the frequency domain. This

may simplify removal of the noise signal and reduce the risk of useful measurement information, such as spectral signals, being removed. Ideally, the frequencies should be sufficiently separated such that the noise and the measurement signals do not overlap in the frequency domain. The devices that
5 are a source for noise in the spectral data may be synchronised with the readout from the camera such that the fixed pattern noise is at the same phase as the readout. This may reduce the number of loadings in a component at which a given noise signal is present.

10 Generating the modified component may comprise mathematically transforming the component to remove the identified noise. More than one component may be modified to remove noise.

Modifying the component may comprise generating the components in a
15 frequency-domain using a time to frequency transform, such as a Fourier transform, and modifying the component in the frequency-domain to remove at least a portion of the noise signal. It will be understood that “time” and “time-domain” as used herein is not limited to data that is a function of time but refers to the domain of the original function before transformation into frequencies
20 contained within the original data by the time to frequency transform.

The step of modifying the components may comprise removing peaks attributable to noise from the components in the frequency-domain. A “peak-flattening” method by used to modify the components in the frequency-domain. For example,
25 data points of peaks of the component(s) in the frequency domain may be removed or, preferably, replaced with substitute data points interpolated from data of the component.

The method may comprise fitting an analytical curve, such as an n th order
30 polynomial, typically a 4th to 6th order polynomial, or set of spline curves, particularly cubic splines, to the data points of the component in the frequency domain and removing ones of the data points that are more than an allowable

deviation above the fitted analytical curve. The allowable deviation may be a multiple of an average deviation of the data of the component in the frequency-domain from the fitted analytical curve, such as a multiple of a root mean square (RMS) of the data. The multiple may be greater than 1 and typically less than 3.

- 5 The method may comprise replacing the removed data points with substitute data points, for example substitute data points based upon the analytical curve.

The peak flattening method may comprise carrying out an iterative process, in which each iteration comprises fitting an analytical curve to reference data,
10 determining the allowable deviation of the reference data from the analytical curve and removing data points of the reference data or of the component in the frequency-domain that are more than the allowable deviation above the analytical curve to provide the reference data for the next iteration until a termination criterion is met, wherein the reference data is initially the data of the component in
15 the frequency-domain. The reference data of the final iteration may be used as the modified component in the frequency-domain. The iterative process may be carried out on each component identified.

Fitting of the spline curve may comprise identifying a location of the anchors
20 (knots). Identification of the locations of the anchors may be based upon the reference data or component. The locations of the anchors for each iteration may be identified automatically using an algorithm. The number of anchor points predefined by a user. The number of anchor points may be predefined based on an expected width of noise peaks in the component.

25

The polynomial/cubic splines tends to be fitted much more closely to the slowly varying spread of frequencies, such as those generated by non-repeating measurement peaks, for example spectral peaks, rather than the sudden spikes of frequencies generated by fixed pattern or burst noise. Accordingly, clipping data
30 points of the reference data or component that are more than the allowable deviation above the analytical curve may remove the sharp noise peaks from the reference data whilst retaining the more slowly varying data attributable to

measurement information, such as spectral signals.

To generate the modified measured data, the components including the at least one modified component may be transformed back into the time domain using a
5 reverse time to frequency transform and the modified measured data built from the components in the time domain.

Alternatively or additionally, generating the modified component may comprise modifying the component in the time-domain. For example, a sinusoid at the
10 same frequency and phase of the noise signal identified as present in the component may be subtracted from the component in the time domain to remove the noise signal.

The method may comprise receiving a user input identifying portions of the
15 components attributable to noise.

The method may comprise an iterative process in which the method as described above is carried out in each iteration, wherein the measured data for each iteration, except for a first iteration, is the modified measured data generated in a previous
20 iteration.

According to a second aspect of the invention there is provided a method of identifying noise in measured data, the method comprising receiving measured data comprising a plurality of sequential observations of discrete variables,
25 carrying out a method of blind signal separation to separate the measured data into a plurality of components, identifying a noise signal in one or more of the components and generating an output identifying a characteristic of the noise signal.

30 The measured data may be spectral data comprising a plurality of spectra captured sequentially using a spectroscopy apparatus.

The characteristic may be a frequency of the noise signal. The user may be then identify a source of noise from the characteristic.

5 The method may comprise generating further measured data using the spectroscopy/metrology apparatus after the spectroscopy/metrology apparatus and/or localized environment to the spectroscopy/metrology apparatus has been modified to remove a source of the noise signal based upon the characteristic. For example, the identified characteristic of the noise signal may enable the user to search for and/or shield the spectroscopy/metrology apparatus from or change
10 electronics that operate at a frequency commensurate with the noise signal. Shielding the spectroscopy/metrology apparatus may comprise relocating the source of noise. Changing the electronics may comprise changing an operating/switching frequency of the electronics or switching off the electronics.

15 The method may comprise generating further measured data using the spectroscopy/metrology apparatus after the spectroscopy/metrology apparatus has been modified to change a resolution, such as spectral resolution, based upon the identified noise signal(s). For example, if a noise signal is found to be located in a component in the frequency domain amongst the measurement signal, the
20 resolution of the measured data may be increased (thus decreasing the range (spectral width) of the measured data) such that the measured signal occupies a smaller region of the component in the frequency domain that is free from the noise signal. Conversely, the resolution of the measured data may be decreased (and the range increased) if the noise signal is identified as being spaced away
25 from the measurement signal.

The method used for generating the modified measured data may depend on the width and/or position of the noise signal to be removed. Broadband noise may require modification of the localized environment to the spectroscopy/metrology
30 apparatus, whereas narrow band noise may be removed using post-processing techniques such as the peak flattening or filtering techniques as described above.

The method may comprise identifying the noise signal from the components in frequency-domain. The noise signal may be identified based upon peaks in the frequency domain having a defined characteristic, such as a peak having an amplitude above a certain threshold and/or having a defined width.

5

According to a third aspect of the invention there is provided a data carrier having stored thereon instructions, which, when executed by a processor, cause the processor to carry out the methods of the first and/or second aspects of the invention.

10

The data carrier may be a non-transient data carrier, such as volatile memory, eg RAM, non-volatile memory, eg ROM, flash memory and data storage devices, such as hard discs, optical discs, or a transient data carrier, such as an electronic or optical signal.

15

According to a fourth aspect of the invention there is provided apparatus comprising a processor, wherein the processor is arranged to carry out the methods of the first and/or second aspects of the invention.

20 According to a fifth aspect of the invention there is provided spectroscopy apparatus comprising a light source for illuminating one or more samples, collection optics for collecting light transmitted from the one or more samples when illuminated by the light source, a detector for generating signals based upon the light collected by the collection optics, wherein readout of signals from the
25 detector is synchronised with a frequency of a source of noise in the signal.

The source of noise may be a switched power supply of the spectroscopy apparatus.

30 According to a sixth aspect of the invention there is provided apparatus for removing noise from measured data comprising means for receiving measured data comprising a plurality of sequential observations of discrete variables, means

for carrying out a method of blind signal separation to separate the measured data into a plurality of components, means for generating at least one modified component by removing at least a portion of a noise signal present in at least one of the components and means for generating modified measured data from ones of the components including the at least one modified component.

According to a seventh aspect of the invention there is provided apparatus for removing noise from measured data comprising a receiving module for receiving measured data comprising a plurality of sequential observations of discrete variables, an analytical module for carrying out a method of blind signal separation to separate the measured data into a plurality of components, a component modification module for generating at least one modified component by removing at least a portion of a noise signal present in at least one of the components and a modified measured data module for generating modified measured data from ones of the components including the at least one modified component.

According to an eighth aspect of the invention there is provided apparatus for identifying noise in measured data comprising means for receiving measured data comprising a plurality of sequential observations of discrete variables, means for carrying out a method of blind signal separation to separate the measured data into a plurality of components and identifying a noise signal in one or more of the components and means for generating an output identifying a characteristic of the identified noise signal.

According to a ninth aspect of the invention there is provided apparatus for identifying noise in measured data comprising a receiving module for receiving measured data comprising a plurality of sequential observations of discrete variables, an analytical module for carrying out a method of blind signal separation to separate the measured data into a plurality of components and for identifying a noise signal in one or more of the components and an output module for generating an output identifying a characteristic of the identified noise signal.

The modified measured data, such as spectral data, may be used for identifying structural information of samples, for example structural information which gives rise to low level spectral signals.

5

For example, the samples may be samples used in as bioprocessing, protein/peptide structural analysis, microbiology, drug delivery in vitro and in vivo, cancer research/pathology, redox biology, regenerative medicine, ageing and neurodegenerative diseases, biofuel and agricultural research, lipidomics, metabolomics, developmental biology, reproductive biology and virology. In such analysis the target of interest may be present in very low quantities and therefore give rise to low level spectral signals. Accordingly, removal of noise from the spectral data may be required to identify the spectral signals of the target. The spectral analysis of the samples may be used in the diagnosis of patients.

15

The samples may include 2D materials and nanotechnology, including graphene, carbon nanotubes (CNT), graphite, diamond, and diamond-like carbon (DLC), MoS₂, hBN, and WSe₂. The method may be used in the control of processes used to manufacture such materials. For example, the number of graphene layers, and their defects, doping and strain, diamond Like Carbon (DLC) thickness, hybridised composition (sp² and sp³), carbon nanotube (CNT) diameter and functionalisation, diamond stress, purity and origin (synthetic or natural), the properties of C₆₀ and other fullerenes and the structural composition of amorphous carbons. Such materials consist of single, or just a few, atomic layers such that high sensitivity of the spectroscopy system, and therefore, spectral data with a low noise signal, is required.

25

The samples may include photovoltaics (PVs), semiconductors and/or catalysis materials. The PVs may include Si-based, CIGS, CdTe, organics, III-Vs, PV material. The structural information that may be determined is alloy fraction, electronic efficiency, strain/stress, thin film thickness, crystal structure type and orientation, crystal quality, sample uniformity and purity (e.g. defects and

30

contaminants). The spectroscopy system may be used to map the samples to determine how the material properties vary across the sample, for example entire semiconductor wafers. As the speed of mapping increases, the noise to spectral signal ratio increases, requiring a method of noise removal for high sensitivity in the mapping routine. This structural information may be used as a feedback mechanism in a manufacturing process.

Most catalysts are fundamentally unstable in air. The spectral data of the atmosphere of suitable cells for storing the catalysts can be used as part of a feedback mechanism to ensure that the catalysts are maintained in a suitable environment. For example, temperature, humidity, and gaseous environment can be controlled within the cell.

The samples may be materials used in battery technology, such as carbon materials, metal oxides, polymers and electrolytes. Mapping and imaging enables the study of the distribution of materials on electrode surfaces, or across cross-sections. As the speed of mapping increases, the noise to spectral signal ratio increases, requiring a method of noise removal for high sensitivity in the mapping routine.

Raman mapping and imaging enables the study of the distribution of materials on electrode surfaces, or across cross-sections. The resulting data can be quantified, giving metrics such as fraction estimates and particle statistics.

Brief Description of the Drawings

Embodiments of the invention will now be described, by example only, with reference to the accompanying drawings, in which:-

Figure 1 shows apparatus according to one embodiment of the invention;

Figure 2 is a flowchart illustrating a method of identifying and removing

noise form spectral data according to an embodiment of the invention;

Figure 3 is a flowchart illustrating a method of “peak flattening” according to an embodiment of the invention;

5

Figure 4a shows an initial principal component (PC) and corresponding modified PC in the “time” domain and **Figure 4b** shows the initial PC and the corresponding modified PC in the frequency domain;

10

Figure 5 is a schematic drawing of a processing apparatus according to an embodiment of the invention; and

Figure 6 is a manufacturing plant comprising apparatus according to the embodiment shown in Figure 1.

15

Description of Embodiments

Referring to Fig. 1, apparatus according to the invention comprises a Raman spectrometer connected to a computer 25 that has access to memory 29.

20

The Raman spectrometer comprises an input laser beam 10 reflected through 90 degrees by a dichroic filter 12, placed at 45 degrees to the optical path. Alternatively a holographic dichroic filter may be placed at a low angle of incidence such as 10 degrees. The laser beam then passes to an objective lens 16, which focuses it to a spot at its focal point 19 on a sample 18. Light is scattered by the sample at this illuminated spot, and is collected by the objective lens 16 and collimated into a parallel beam which passes back to the dichroic filter 12. The filter 12 rejects Rayleigh scattered light having the same frequency as the input laser beam 10, and transmits the Raman scattered light. The Raman scattered light then passes to a Raman analyser 20.

30

The Raman analyser 20 comprises a dispersive element such as a diffraction

grating. The light from the analyser 20 is focused by a lens 22 onto a suitable photo-detector 24. A photo-detector array is preferred. In the present embodiment the detector 24 is a charge-coupled device (CCD), which consists of a two-dimensional array of pixels, and which is connected to the computer 25 which
5 acquires data from each of the pixels and analyses it as required. The analyser 20 produces a spectrum spread out in a line along the CCD 24.

Samples 18 may be mounted on an X-Y table so that the focal point 19 can be scanned across it in X and Y directions, e.g. under control of the computer 25. A
10 plurality of spectra can then be gathered, each spectrum corresponding to a different point on the sample. A map can then be generated based upon properties of the sample determined from the collected spectra.

The computer 25 is programmed with software code on a suitable medium, such
15 as memory 29, comprising instructions, which when executed by a processor of computer 25 cause the computer 25 to perform the analysis routines described below. Alternatively, the data on the Raman spectrum/spectra obtained may be transferred to a separate computer having such software for this analysis. In either case, as the analysis proceeds, the values determined are stored in the computer
20 concerned, and may be further processed and output or displayed to show the concentrations of the components in the sample/samples.

To process the spectral data to determine the Raman spectrum/spectra present and therefore, the chemical component(s)/properties of the sample, the spectral data is
25 first processed to remove noise, in particular fixed pattern noise.

A switched power supply 11 is provided for powering the electronic devices of the spectrometer. A start of the readout from the CCD 24 is synchronised with the switching of the powder supply 11. This may simplify the analysis of the signals
30 for fixed pattern noise, as described below.

Referring to Figure 2, the method comprising receiving 101 spectral data

comprising a plurality of spectra sequentially recorded by the spectrometer. Principal component analysis (PCA) 102 is carried out on a plurality of the spectra resulting in a set of principal components. As required by principal component analysis, the principal components are ordered with the first principal component
5 describing the largest variance in the spectral data and each succeeding component in turn describing the highest variance in the spectral data under the constraint that it is orthogonal to preceding principal components. The light grey lines in Figure 4a show a 6th principal component 301 for a PCA of spectral data.

10 The principal components are transformed 103 into the frequency domain, for example using a fast Fourier transform (FFT). Figure 4b shows the 6th principal component 302 shown in Figure 4a in the frequency domain. Fixed pattern noise is identified 104 in the principal components in the frequency domain by identifying frequency peaks characteristic of fixed pattern noise.

15 Devices of the spectrometer, such as the switched power supply 11 and CCD detector 24 operate as known frequencies. Such known operating frequencies are stored in memory 29. The algorithm searches for peaks in the frequency domain at or around these known operating frequencies. If a peak is found then this is
20 attributed to noise and removed 105 from the principal component. The synchronisation of the readout from the CCD 24 and the switching of the power supply 11 may reduce the number of principal components comprising fixed pattern noise due the switched power supply 11 (as the fixed pattern noise form the switch power supply 11 has the same phase as the readout from the CCD).

25 Furthermore, the algorithm searches for sharp peaks in the frequency domain that are unlikely to result from spectral data. Spectra data tends to result in a more smoothly varying signal in the frequency domain with much of the signal concentrated at the lower frequencies (although some contribution will exist at the
30 higher frequencies due to high spectral peaks in the time domain). Such peaks are identified by a suitable peak detection algorithm, as is known in the art. Peaks due to noise can then be identified based upon defined signatures, such as defined

thresholds for peak maximum and/or peak width (e.g. FWHM). The thresholds may be based upon an analysis of the data in the frequency domain, for example a threshold for peak maximum may be set based upon a multiple of a RMS average of the data. In Figure 4b, the illustrated principal component comprises two
5 peaks 302a, 302b having the required signature to be attributable to noise.

In one embodiment, the method comprises generating 104a an output identifying characteristics, such as frequency, of the identified noise signal. A user may then use the knowledge of the noise signal characteristics to identify and eliminate
10 sources of noise that are obscuring the spectral signal.

In step 105 a modified set of principal components are generated by modifying the loadings of the principal components to remove the identified noise signal. In this embodiment, the principal components are modified in the frequency domain. A
15 “peak flattening” method may be used to remove the peaks from the principal component and “fill” in the gap left by peak removal. In Figure 4b, the data points of the peaks are replaced at a given point at which the peak is deemed to begin with a curve or line. The modified 6th principal component in the frequency domain is shown by the dark line 303.

20 The method used for “peak flattening” may be varied depending on the width and position of the peak to be removed. The peak flattening technique may be chosen in order to optimise the rebuild of the spectral data from the modified components.

25 The modified principal component is then transformed 106 back into the time domain using a reverse FFT. The modified 6th principal component in the time domain is shown by the dark line 304. As can be seen in the example shown in Figure 4a, removal of the fixed pattern noise reveals spectrally significant data hidden within the noise.

30 Modified spectral data is then generated 107 by rebuilding spectral data using the principal components including the modified principal components from which

fixed pattern noise has been removed.

The method is carried out as an iterative process, with the modified spectral data being fed back into step 102 and used as the initial data (i.e. “the spectral data”) for the next iteration.

The iterative process is continued until termination criteria is met. The modified spectral data of the final iteration is output as the spectral data to be used for further analysis, such as to be used for identifying properties of the sample.

10

In another embodiment, the process is not an iterative process, with steps 102 to 107 being carried out on the spectral data only once. However, a further PCA of the rebuilt spectral data may be carried out and principal components below a defined n th order removed in order to remove non-fixed pattern noise from the measured data.

15

In a further embodiment, rather than specifically identifying peaks attributable to noise, a peak flattening routine is applied to each principal component in the entire data set to remove peaks having fixed pattern noise signatures.

20

Referring to Figure 3, the modified principal components are generated using an iterative process. After determining the principal components (in step 102) and transforming the principal components into the frequency domain (step 103), for each principal component $W_k(\nu)$ in the frequency domain, a modified principal component $W'_k(\nu)$ is generated.

25

In step 202, reference data is initially set equal to the principal component $W_k(\nu)$. To this reference data, an analytical curve is fitted 203, such as an n th order polynomial or a set of cubic spline curves. An order of the polynomial to be used may be preset, for example the process may be preset to be a fifth order polynomial. In the case of fitting a set of cubic spline curves, a number of anchor points is defined by the user.

30

In step 204, a deviation of each datum point of the reference data from the analytical curve is determined and a root mean square (RMS) value is calculated for the deviations. The reference data is then modified, in this embodiment by replacing 205 data points of the reference data that are more than an allowable deviation (DEV) above the analytical curve with substitute data points that are equal to a value of the analytical curve at that frequency plus DEV. In this embodiment, the allowable deviation DEV is $M \times \text{RMS}$, wherein M is a positive real number. M may be set by the user.

10

In step 206, it is determined whether a termination criterion is met, in this embodiment, whether any points have been replaced in step 205. If points were removed from the reference data, the method proceeds to the next iteration, wherein an analytical curve is fitted to the modified reference data. The iterative process continues until no points are replaced in step 205. In step 208, the reference data of the final iteration is output as the modified principal component $W'_k(v)$.

The modified spectral data may be analysed to determine properties of the sample(s) using any suitable technique, such as the methods described in WO 2007/144664 or WO 2012/156667. The invention may enable faster speeds for mapping a surface as it may be possible to identify spectral data at higher ratios of noise to spectral signal (the amplitude of the spectral data reducing as the mapping speed increases).

25

To carry out the above described method, the computer 25 may comprise processing modules 401 to 406, as defined in hardware or software. The modules 401 to 406 perform logically discrete functions and can be called by a lower level module/unit, such as control module 407 and pass values back to the lower level module/unit. In this embodiment, the modules 401 to 406 comprise a receiving module 401 for receiving the spectral data, an analysis module 402 for carrying of PCA on a data set to determine principal components of the data set, an

30

identification module 403 for identifying noise signals in a data set, a modification module 404 for generating a modified data, such as a modified principal component, by removing at least a portion of a noise signal present in the data set, a modified measured data module 405 for generating modified spectral data by
5 rebuilding spectral data from ones of the components including the at least one modified component and an output module 406 for outputting the modified spectral data and/or a characteristic of the identified noise signal.

Referring to Figure 5, the apparatus may be used in a manufacturing process to
10 identify properties of manufactured articles 504, 507. The apparatus comprises means, in this embodiment a conveyor, for transporting material 502 to a machine 503 for transforming the material 502 into a manufactured article 504. For example, the article 504 may be a medicament, such as a tablet or a pill, a wafer of a 2D material or nanotechnology, a photovoltaic cell, semiconductor wafer and/or
15 catalysis material. The manufactured article 504 is then transported to a spectrometer 506 for obtaining at least one Raman spectrum of the article 504. In one embodiment, the spectrometer may map a surface of the article 504. A plurality of spectra obtained from the article(s) 504 are sent to processing unit 525. Processing unit 525 is arranged to carry out the method as described above
20 to remove fixed pattern noise in the spectral data. The fixed pattern noise may be internally generated by the spectrometer 506 or by other parts of the manufacturing apparatus. The processing unit 525 may comprise modules, as described with reference to Figure 5.

25 The processing unit analyses the modified spectral data to identify properties of the articles 507. The identified properties may be used to determine whether articles 507 meet specified requirements, for example a specified level of homogeneity, semiconductor doping, etc and articles that are deemed not to meet the specified requirements may be discarded and/or further processed to correct
30 the deficiency. Furthermore, operation of the machine 503 may be altered based upon the determined properties in order to alter the manufacturing process. For example, the operation may be altered to ensure that desired properties are

achieved for future articles.

Furthermore, the processing unit 525 may output characteristics, such as frequencies, of fixed pattern noise identified in the spectral data. A user may
5 analyse the output and alter the machine operation based upon the identified noise signal characteristics.

It will be understood that modifications and alterations can be made to the above described embodiments without departing from the invention as defined by the
10 claims. For example, modification of the principal component(s) may be carried out in the time domain. In some embodiments, the order of the PCA and Fourier transform may be reversed such that PCA is carried out on the spectral data in the frequency domain, with PCA and Fourier transforms being linear transformations of the data.

15

20

CLAIMS

1. A method of removing noise from spectral data, the method comprising receiving spectral data comprising a plurality of spectra captured sequentially
5 using a spectroscopy system, carrying out a method of blind signal separation to separate the spectral data into a plurality of components, generating at least one modified component by removing at least a portion of a noise signal present in at least one of the components and generating modified spectral data from ones of the components including the at least one modified component.
10
2. A method according to claim 1, wherein the method of blind signal separation comprises principal component analysis.
3. A method according to claim 1 or claim 2, comprising identifying the
15 noise signal based upon a known noise signature.
4. A method according to claim 3, wherein the noise signature is based upon a known noise frequency.
- 20 5. A method according to claim 4, wherein the known noise frequency is an operating frequency of an electronic component of the spectroscopy system.
6. A method according to any one of the proceeding claims, wherein generating the modified component comprises mathematically transforming the
25 component to remove the noise signal.
7. A method according to claim 6, wherein modifying the component comprises generating the components in a frequency-domain using a time to frequency transform and modifying the component in the frequency-domain to
30 remove at least a portion of the noise signal.
8. A method according to claim 7, wherein modifying the components comprises removing peaks attributable to noise from the components in the frequency-domain.

9. A method according to claim 8, wherein the peak is replaced with substitute data interpolated from data of the component in the frequency domain.
- 5 10. A method according to claim 8 or claim 9, comprising fitting an analytical curve to the data points of the component in the frequency domain and removing data points of the component in the frequency-domain that are more than an allowable deviation above the fitted analytical curve.
- 10 11. A method according to claim 10, wherein the allowable deviation is a multiple of an average deviation of the data points from the fitted analytical curve.
12. A method according to claim 10 or claim 11, comprising replacing the removed data points with substitute data points.
- 15 13. A method according to claim 12, wherein the substitute data points are equal to the allowable deviation above the fitted analytical curve.
14. A method according to any one of claims 10 to 13, comprising carrying
20 out an iterative process, in which each iteration comprises fitting an analytical curve to reference data, determining the allowable deviation of the reference data from the analytical curve and removing data points of the reference data or the data of the component in the frequency domain that are more than the allowable deviation above the analytical curve to provide the reference data for the next
25 iteration until termination criterion is met, wherein the reference data is initially the data of the component in the frequency domain.
15. A method according to claim 14, wherein the reference data of the final iteration is used as the modified component.
- 30 16. A method according to claim 14 or claim 15, wherein the iterative process is carried out on each component.
17. A method according to any one of claims 6 to 16, wherein generating the
35 modified component comprises modifying the component in the time-domain.

18. A method comprising an iterative process in which the method according to any one of the preceding claims is carried out in each iteration, wherein the spectral data for each iteration, other than a first iteration, is the modified spectral data generated in a previous iteration.

19. A method according to any one of the preceding claims, wherein the plurality of spectra are obtained from multiple points on a sample.

20. A method of identifying noise in spectral data, the method comprising receiving spectral data comprising a plurality of spectra captured sequentially using a spectroscopy system, carrying out a method of blind signal separation to separate the spectral data into a plurality of components, identifying a noise signal in one or more of the components and generating an output identifying a characteristic of the noise signal.

21. A method according to claim 20, wherein the characteristic is a frequency of the noise signal.

22. A method according to claim 20 or claim 21, comprising generating further spectral data using the spectroscopy system after the spectroscopy system and/or localized environment to the spectroscopy system has been modified to remove and/or modify a source of the noise identified based upon the characteristic.

23. A method according to any one of claims 20 to 22, comprising generating further spectral data using the spectroscopy apparatus after the spectroscopy apparatus has been modified to change a spectral resolution based upon the identified noise signal.

24. A method according to any one of claims 20 to 23, wherein the method comprises identifying the noise signal from the components in a frequency-domain.

25. A method according to claim 24, wherein the noise is identified based upon peaks in the frequency domain having a defined noise signature.

26. A data carrier having stored thereon instructions, which, when executed by
5 a processor, cause the processor to carry out the methods of any one of claims 1 to 25.

27. Apparatus comprising a processor, wherein the processor is arranged to carry out the methods of any one of claims 1 to 25.

10

28. Spectroscopy apparatus comprising a light source for illuminating one or more samples, collection optics for collecting light transmitted from the one or more samples when illuminated by the light source, a detector for generating signals based upon the light collected by the collection optics, wherein readout of
15 signals from the detector is synchronised with a frequency of a device which generates noise in the signal.

29. Spectroscopy apparatus according to claim 28, wherein the device is a switched power supply of the spectroscopy apparatus.

20

30. Apparatus for removing noise from spectral data comprising means for receiving spectral data comprising a plurality of spectra captured sequentially using a spectroscopy system, means for carrying out a method of blind signal separation to separate the spectral data into a plurality of components, means for
25 generating at least one modified component by removing at least a portion of a noise signal present in at least one of the components and means for generating modified spectral data from ones of the components including the at least one modified component.

30

31. Apparatus for removing noise from spectral data comprising a receiving module for receiving spectral data comprising a plurality of spectra captured sequentially using a spectroscopy system, an analytical module for carrying out a method of blind signal separation to separate the spectral data into a plurality of components, a component modification module for generating at least one

modified component by removing at least a portion of a noise signal present in at least one of the components and a modified spectral data module for generating modified spectral data from ones of the components including the at least one modified component.

5

32. Apparatus for identifying noise in spectral data comprising means for receiving spectral data comprising a plurality of spectra captured sequentially using a spectroscopy system, means for carrying out a method of blind signal separation to separate the spectral data into a plurality of components and
10 identifying a noise signal in one or more of the components and means for generating an output identifying a characteristic of the identified noise signal.

33. Apparatus for identifying noise in spectral data comprising a receiving module for receiving spectral data comprising a plurality of spectra captured
15 sequentially using a spectroscopy system, an analytical module for carrying out a method of blind signal separation to separate the spectral data into a plurality of components and for identifying a noise signal in one or more of the components and an output module for generating an output identifying a characteristic of the identified noise signal.

20

34. A method of removing noise from measured data, the method comprising receiving measured data comprising a plurality of sequential observations of discrete variables, carrying out a method of blind signal separation to separate the measured data into a plurality of components, generating at least one modified
25 component by removing at least a portion of a noise signal present in at least one of the components and generating modified data from ones of the components including the at least one modified component.

35. A method of identifying noise in measured data, the method comprising
30 receiving measured data comprising a plurality of sequential observations of discrete variables, carrying out a method of blind signal separation to separate the measured data into a plurality of components, identifying a noise signal in one or more of the components and generating an output identifying a characteristic of

the noise signal.

36. A data carrier having stored thereon instructions, which, when executed by a processor, cause the processor to carry out the methods of claim 34 or claim 35.

5

37. Apparatus comprising a processor, wherein the processor is arranged to carry out the methods of claim 34 or 35.

38. Apparatus for removing noise from measured data comprising means for receiving measured data comprising a plurality of sequential observations of discrete variables, means for carrying out a method of blind signal separation to separate the measured data into a plurality of components, means for generating at least one modified component by removing at least a portion of a noise signal present in at least one of the components and means for generating modified data from ones of the components including the at least one modified component.

10

15

39. Apparatus for removing noise from measured data comprising a receiving module for receiving measured data comprising a plurality of sequential observations of discrete variables, an analytical module for carrying out a method of blind signal separation to separate the measured data into a plurality of components, a component modification module for generating at least one modified component by removing at least a portion of a noise signal present in at least one of the components and a modified data module for generating modified data from ones of the components including the at least one modified component.

20

25

40. Apparatus for identifying noise in measured data comprising means for receiving measured data comprising a plurality of sequential observations of discrete variables, means for carrying out a method of blind signal separation to separate the measured data into a plurality of components and identifying a noise signal in one or more of the components and means for generating an output identifying a characteristic of the identified noise signal.

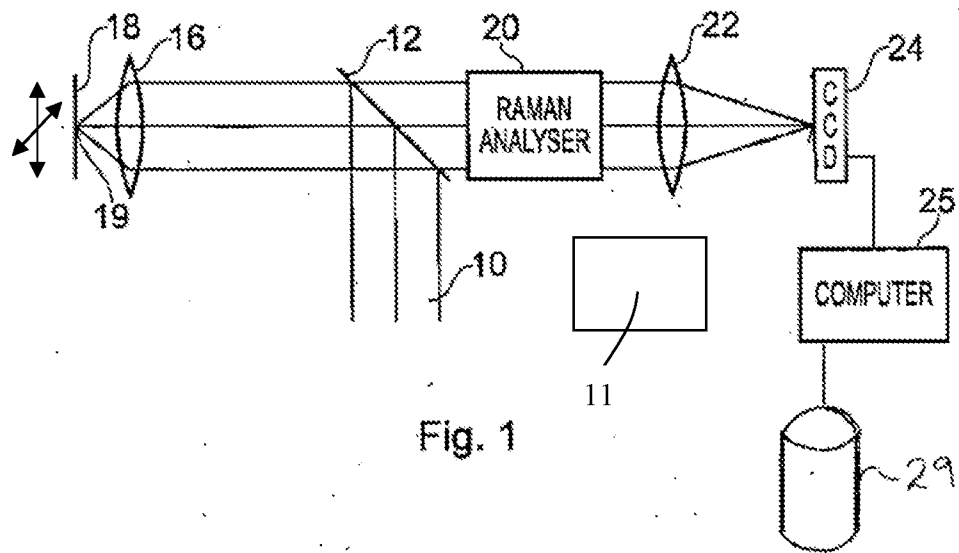
30

41. Apparatus for identifying noise in measured data comprising a receiving module for receiving spectral data comprising a plurality of sequential

observations of discrete variables, an analytical module for carrying out a method of blind signal separation to separate the measured data into a plurality of components and for identifying a noise signal in one or more of the components and an output module for generating an output identifying a characteristic of the

5 identified noise signal.

1/5



2/5

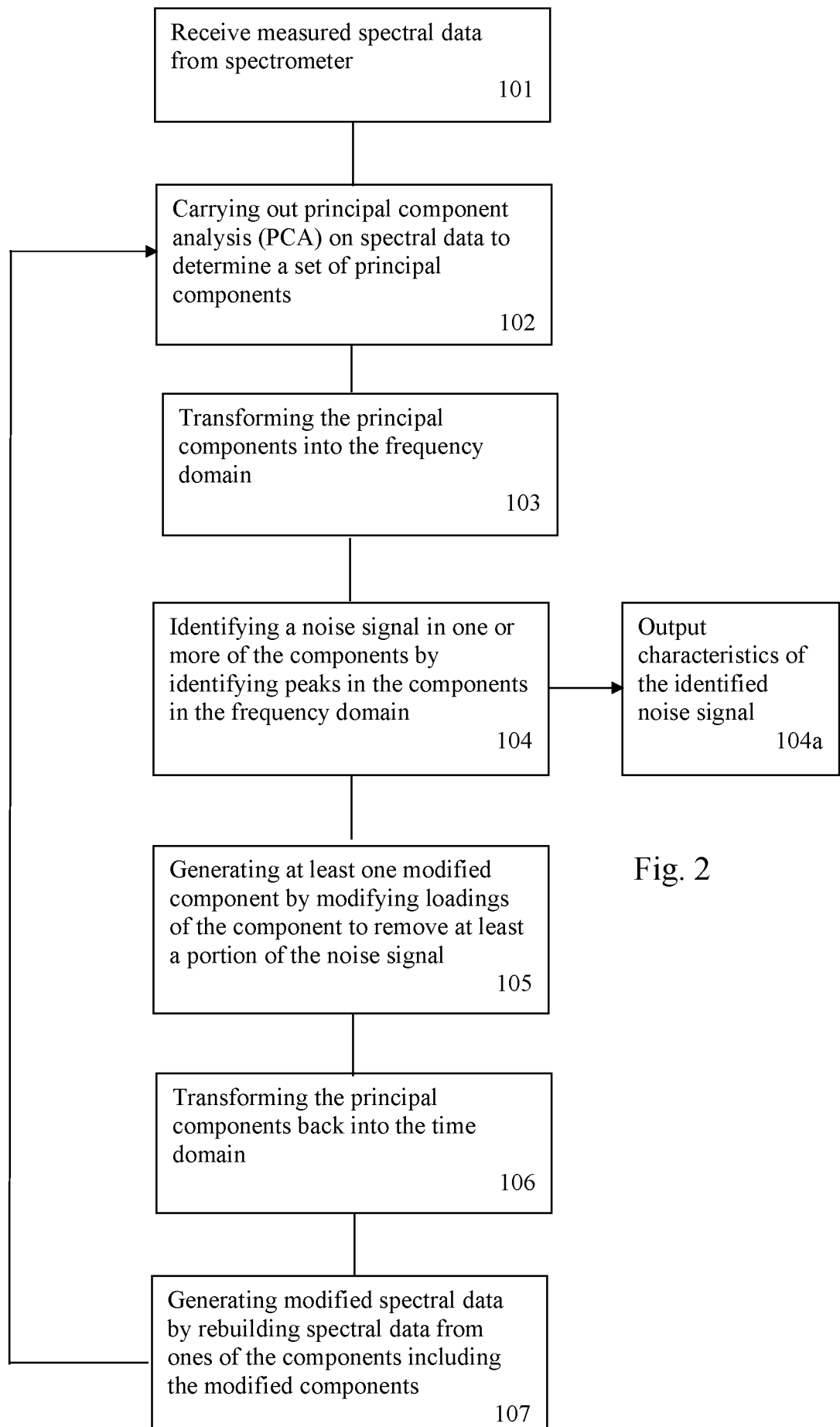


Fig. 2

3/5

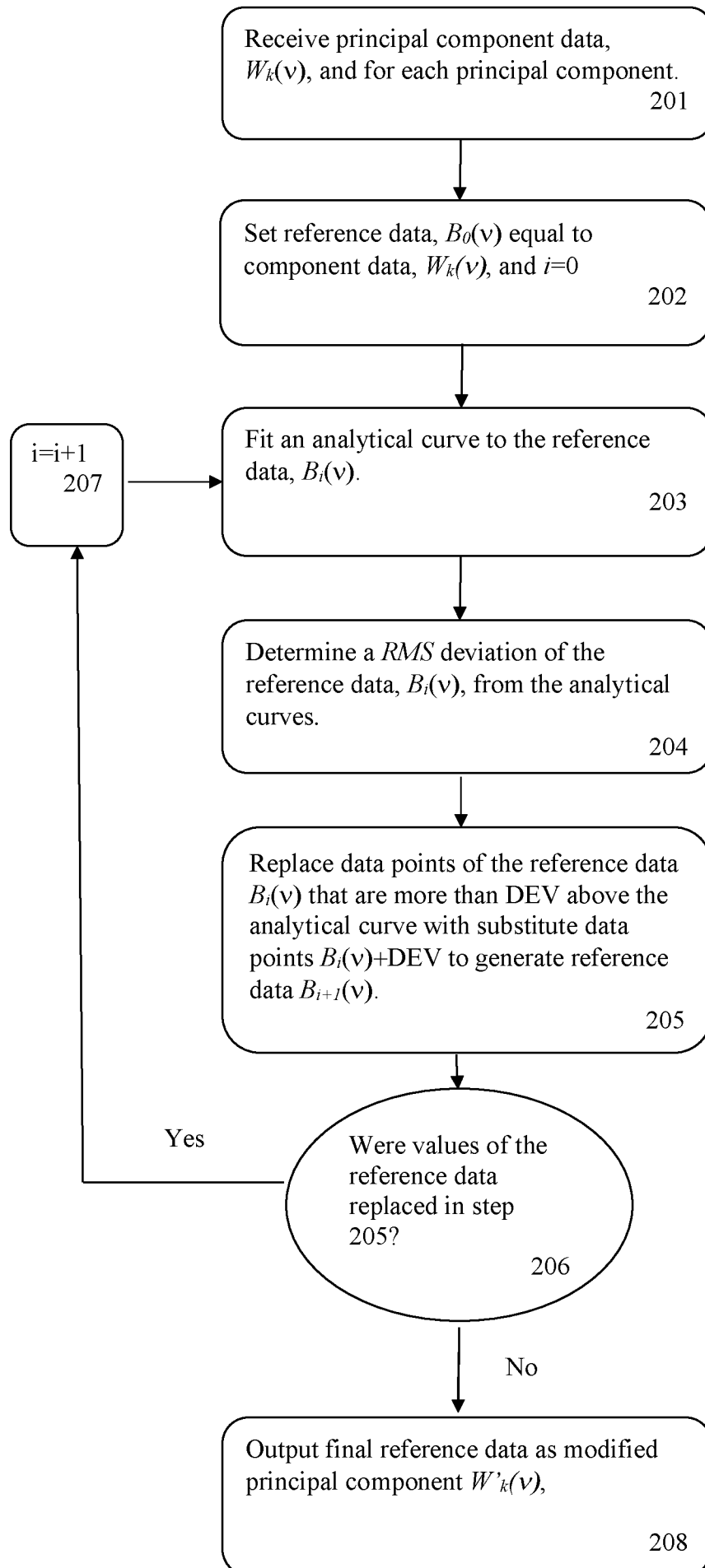
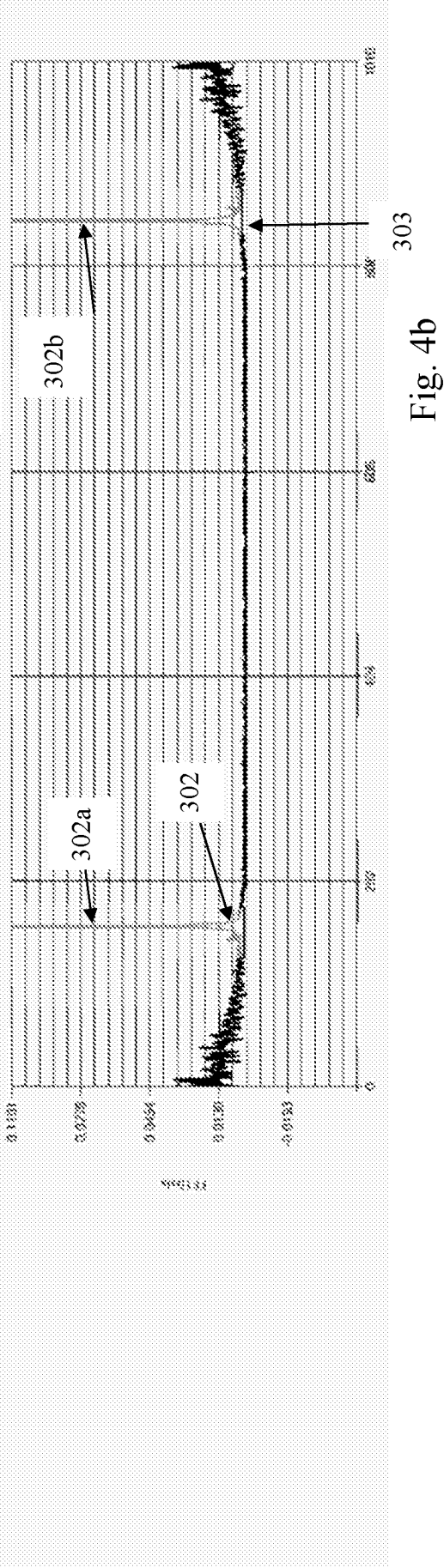
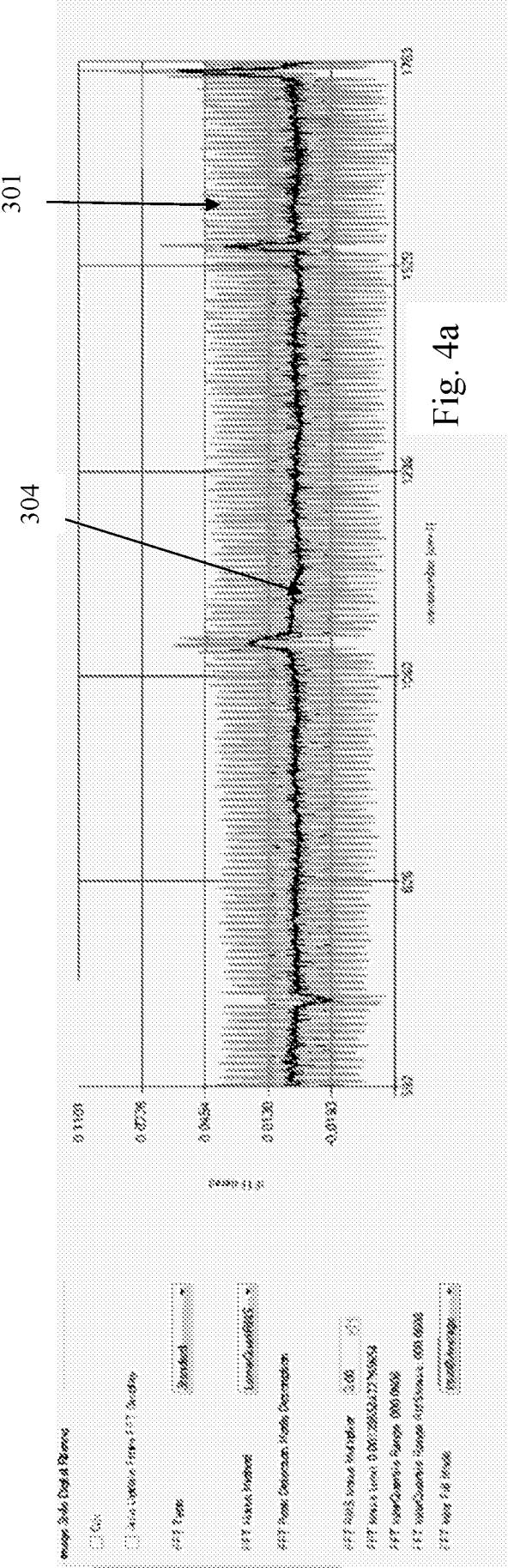


Fig. 3



5/5

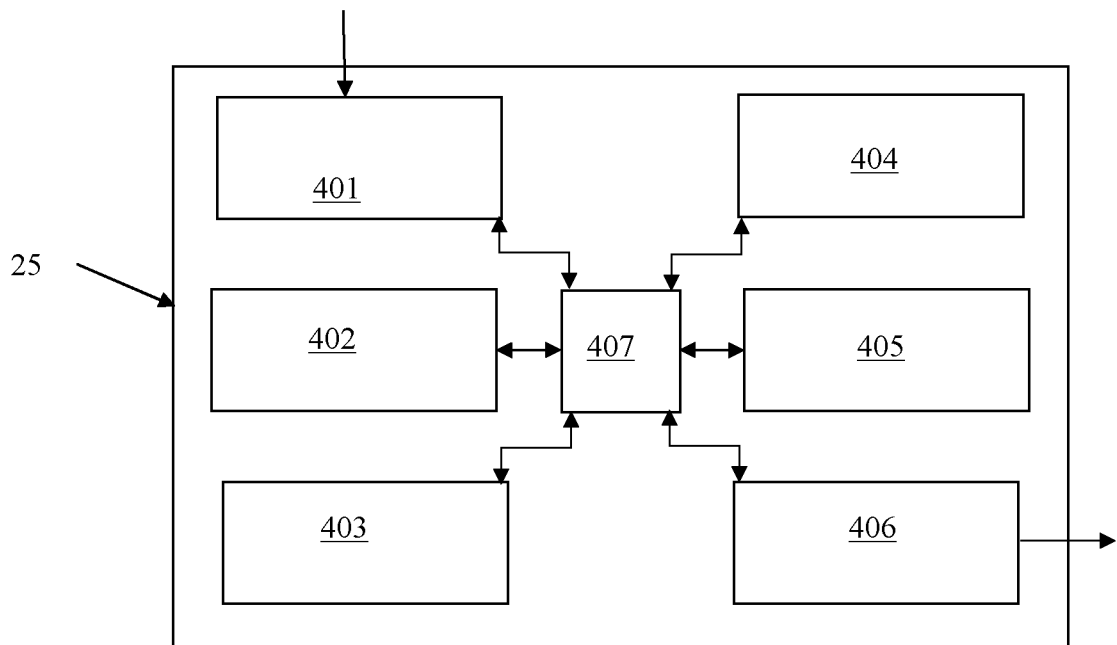


Fig. 5

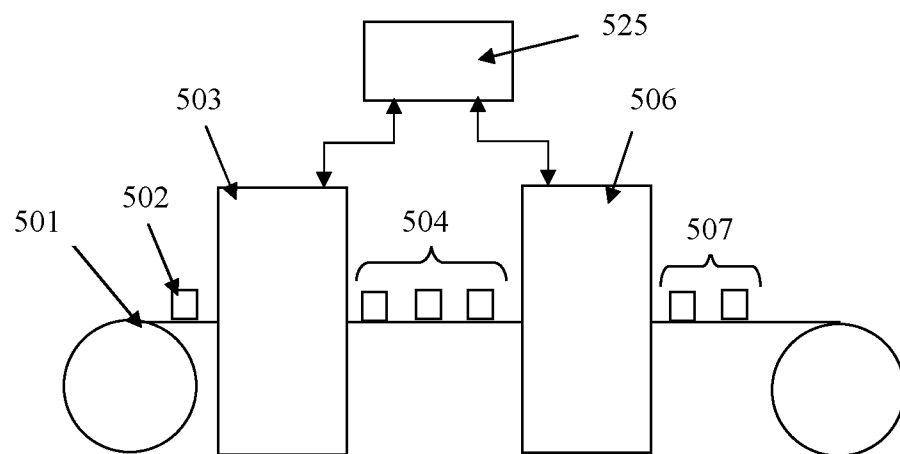


Fig. 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2017/052923

A. CLASSIFICATION OF SUBJECT MATTER INV. G01J3/28 G01J3/44 G01N21/65 ADD.				
According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G01J G01N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, WPI Data				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
X	GUANGYI CHEN ET AL: "Denoising of Hyperspectral Imagery Using Principal Component Analysis and Wavelet Shrinkage", IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 49, no. 3, 1 March 2011 (2011-03-01), pages 973-980, XP011348499, ISSN: 0196-2892, DOI: 10.1109/TGRS.2010.2075937 figure 1	1-19,26,27,30,31		
A	----- US 6 072 576 A (MCDONALD MICHAEL F [US] ET AL) 6 June 2000 (2000-06-06) column 6, lines 7-63	1,30,31		
A	----- US 8 570 507 B1 (COOPER JOHN [US] ET AL) 29 October 2013 (2013-10-29) example 1 -----	1,30,31		
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.				
* Special categories of cited documents : <table border="0"> <tr> <td> "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search 7 December 2017		Date of mailing of the international search report 13/02/2018		
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Rödiger, Christoph		

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2017/052923

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-19, 30, 31(completely); 26, 27(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-19, 30, 31(completely); 26, 27(partially)

details of a method for removal of noise in spectral data

2. claims: 20-25, 32, 33(completely); 26, 27(partially)

details of a method for identification of noise in spectral data

3. claims: 28, 29

details of a spectroscopy apparatus

4. claims: 34, 38, 39

details of a method for removal of noise in data comprising sequential observations of discrete variables

5. claims: 35-37, 40, 41

details of a method for identification of noise in data comprising sequential observations of discrete variables

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2017/052923

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6072576	A	06-06-2000	AU 727750 B2 21-12-2000 BR 9714450 A 05-12-2000 CA 2276624 A1 09-07-1998 CN 1245570 A 23-02-2000 DE 69705323 D1 26-07-2001 DE 69705323 T2 25-04-2002 EP 0948761 A1 13-10-1999 ES 2160983 T3 16-11-2001 HK 1026034 A1 02-01-2004 IL 130731 A 25-07-2004 JP 4435307 B2 17-03-2010 JP 2001508354 A 26-06-2001 KR 20000062408 A 25-10-2000 MY 117038 A 30-04-2004 NO 993261 A 26-08-1999 PL 334397 A1 28-02-2000 RU 99116247 A 20-08-2001 TW 558557 B 21-10-2003 TW 1266769 B 21-11-2006 US 6072576 A 06-06-2000 WO 9829787 A1 09-07-1998
US 8570507	B1	29-10-2013	NONE