

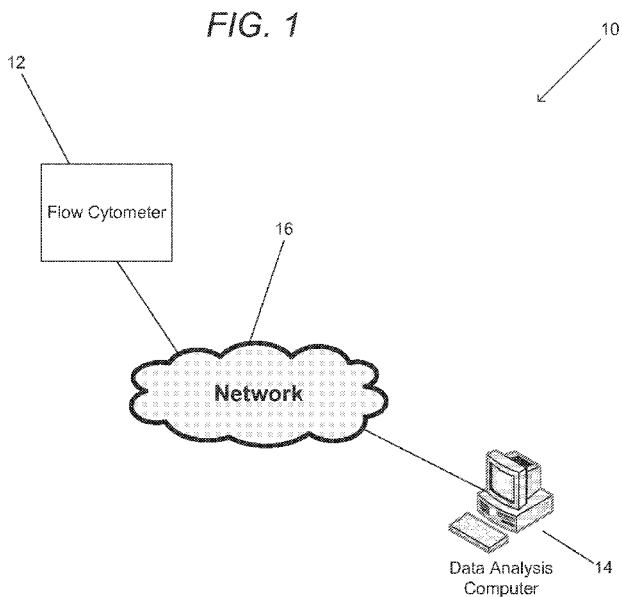


- (51) **International Patent Classification:**
G01N 21/64 (2006.01) G01J 3/42 (2006.01)
G01J 3/30 (2006.01)
- (21) **International Application Number:**
PCT/US2013/047132
- (22) **International Filing Date:**
21 June 2013 (21.06.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/662,916 22 June 2012 (22.06.2012) US
- (71) **Applicants:** DE NOVO SOFTWARE LLC [US/US]; 3250 Wilshire Boulevard, Suite 803, Los Angeles, CA 90010 (US). PURDUE RESEARCH FOUNDATION [US/US]; 1281 Win Henschel Boulevard, West Lafayette, IN 47906 (US).
- (72) **Inventors:** NOVO, David; 4711 College View Avenue, Los Angeles, CA 90041 (US). RAJWA, Bartek; 1901 WILLET CT., WEST LAFAYETTE, IN 47906 (US).
- (74) **Agent:** CHONG, Eugene, K.; KPPB LLP, 2400 E. Katella, Suite 1050, Anaheim, CA 92806 (US).

- (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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:** SYSTEMS AND METHODS FOR UNMIXING DATA CAPTURED BY A FLOW CYTOMETER



(57) **Abstract:** Systems and methods for obtaining fluorochrome abundance information by unmixing fluorescence emission data captured by a flow cytometer in accordance with embodiments of the invention are disclosed. In one embodiment, a data analysis system includes a processor, a memory, and an optical data analysis application, wherein the optical data analysis application configures the processor to obtain control optical data, generate a mixing model using the obtained control optical data and a system of linear combinations, obtain experimental optical data for particles stained with a set of fluorochromes, and estimate abundances of the fluorochromes in the set of fluorochromes using the obtained experimental optical data by solving a system of equations to unmix the optical data, where the number of equations is larger than the number of unknowns, based upon the generated mixing model using an unmixing process that accounts for increased noise variance with increased fluorochrome abundance.



SYSTEMS AND METHODS FOR UNMIXING DATA CAPTURED BY A FLOW CYTOMETER

BACKGROUND

[0001] The present invention relates generally to the field of flow cytometry and more specifically to determining the abundance of fluorochromes in flow cytometry.

[0002] Flow cytometry is a powerful cell-analysis technique, applied in various fields of life science ranging from basic cell biology to genetics, immunology, molecular biology, microbiology, plant cell biology, cancer diagnosis and environment science. Flow cytometry involves generating a stream of biological particles (bioparticles) that pass in single file through a beam of light (usually generated by one or more lasers having separate frequencies). As bioparticles pass through the beam, the beam is scattered producing light-scatter pattern dependent on particle structure, shape, and physical composition. Additionally fluorescent, phosphorescent, or Raman-scattering chemicals found in, or attached to, the bioparticle may be excited into emitting light. Depending on the type of optical process the signal emitted by the chemical species can be at a longer wavelength than the light source – in the case of linear optics, or shorter – in the case of nonlinear optical phenomena as two-photon excitation. In typical flow cytometry instruments forward-angle light scatter (size-related) and side-angle light scatter (shape- and structure-related) as well as various fluorescence emissions are collected following illumination/excitation. The data are collected, digitized, and stored on a computer where they can be further processed to discriminate subpopulations of bioparticles (cells) with similar characteristics from within the original heterogeneous sample. By analyzing the intensity of the detected light, it is possible to derive various types of information about the physical and chemical structure of each bioparticle.

[0003] A wide range of fluorochromes can be used to stain or label bioparticles. Fluorochromes are typically attached to an antibody that recognizes a target feature on or in a bioparticle, or chemical entity with affinity for a cell membrane or another structure of a bioparticle. Each fluorochrome typically has a characteristic excitation and emission spectra, and the emission spectra of different fluorochromes used in a single sample often overlap.

[0004] Common flow cytometry implementations assume that the signal from every individual fluorochrome should be collected using an independent detector. The optical pathway of flow cytometry instruments is arranged to separate signals from different fluorochromes and route them into dedicated detectors; however, owing to spectral overlap and imperfection of filters a complete separation is almost never possible. Therefore, the fluorescence emitted by every fluorochrome may be simultaneously collected by more than one detector (in an extreme case, all the detectors).

[0005] Both the absorption and emission spectra of the fluorochromes used in flow cytometry carry valuable spectral information about tagged bioparticles. The commonly used optical design of flow cytometry instruments make it desirable to have a series of efficient fluorochromes that have very specific and narrow excitation maxima within the sensitivity of an individual detector. The flow cytometry systems efficiently collect optical signals emitted by individual bioparticles, and convert them quantitatively into values that can be related to biological phenomena of interest. Flow cytometry instruments have developed from single-detector systems to devices having a plurality of detectors, hence capable of collecting signals from a number of chemical species simultaneously. More recently, the use of detector arrays with more than 30 detectors has been demonstrated. The arrays may be implemented as multianode photomultipliers (PMT), or linear charge couple devices (CCD). However, the vast majority of commercial systems are able to collect from 5 to 7 simultaneous emitted signals, plus two or more light-scatter related measurements. Flow cytometers capable of collecting multiple signals are often referred to as polychromatic flow cytometers.

[0006] In traditional polychromatic flow cytometry, the number of employed detectors is equal to the number of investigated labeled markers. Since the process of spectral overlap occurring during the measurement can be mathematically represented as linear mixing, the abundances (or values linearly correlated with abundances) are calculated by an unmixing operation that multiplies the measured data vectors (or raw fluorescence observations) by the inverse of the mixing (“spillover”) matrix. Although the mixing matrices are a priori unknown, they can be easily approximated by employing single-stained controls – that is by performing measurements of samples labeled by one fluorochrome at a time, and normalizing the resultant spectra in an appropriate fashion. This process leading to the recovery of abundances is known as flow cytometry compensation, and is well described in flow cytometry literature.

[0007] The data generated by a flow cytometer is typically formatted in accordance with the Flow Cytometry Standard (FCS), which enables analysis of the data using software applications such as FCS Express provided by De Novo Software LLC of Los Angeles, CA. Analysis that can be performed using such software involves the generation of one dimensional and two dimensional plots. In addition to plots, the software enables the generation of plot overlays and gates, which are used to generate statistics describing the observed populations of bioparticles. The analysis strategy and results derived from it are stored in an electronic document referred to as a layout file. Layout files can also optionally contain the raw data used to generate the results.

SUMMARY OF THE INVENTION

[0008] Systems and methods for obtaining fluorochrome abundance information by unmixing fluorescence emission data captured by a flow cytometer in accordance with embodiments of the invention are disclosed. In one embodiment, a data analysis system is configured to analyze optical data captured by a flow cytometer with respect to a plurality of particles stained with a plurality of fluorochromes, where an optics and detection system within the flow cytometer separates optical emission with respect to spectral ranges and where at least one detector is used to capture a number of optical measurements that is greater than the plurality of fluorochromes used to stain the

plurality of particles, where the data analysis system includes a processor, a memory connected to the processor and configured to store an optical data analysis application, wherein the optical data analysis application configures the processor to: obtain control optical data for at least one particle stained with at least one fluorochrome selected from a set of fluorochromes, where the control optical data is captured by the flow cytometer configured so that an optics and detection system within the flow cytometer separates optical emission with respect to a predetermined set of spectral ranges, generate a mixing model using the obtained control optical data and a system of linear combinations, obtain experimental optical data for particles stained with the set of fluorochromes, where the experimental optical data is captured by the flow cytometer configured so that an optics and detection system within the flow cytometer separates optical emission with respect to the predetermined spectral ranges using at least one detector configured to capture a number of optical measurements and the number of optical measurements is greater than the number of fluorochromes in the set of fluorochromes, and estimate abundances of the fluorochromes in the set of fluorochromes using the obtained experimental optical data by solving an overdetermined system of equations to unmix the optical data, based upon the generated mixing model that accounts for increased noise variance with increased fluorochrome abundance.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 is a system diagram of a data analysis system for acquiring and analyzing flow cytometry data in accordance with an embodiment of the invention.

[0010] FIG. 2 is a flow chart illustrating a process for acquiring fluorescence emission data using a flow cytometer to estimate fluorochrome abundances using an unmixing process in accordance with an embodiment of the invention that assumes the error increases with the size of the observed fluorescence signal.

DETAILED DESCRIPTION OF THE INVENTION

[0011] Turning now to the drawings, systems and methods for obtaining fluorochrome abundance information by unmixing fluorescence emission data captured by a flow cytometer in accordance with embodiments of the invention are illustrated. In several embodiments, the flow cytometer is configured as an over-determined system in which the number of detectors that capture fluorescence emission data is greater than the number of fluorochromes used to stain the bioparticles observed by the flow cytometer. In several embodiments, the unmixing process used to estimate fluorochrome abundances from the captured fluorescence emission data specifically addresses the fact that variance in the observed signal is not equal along the dynamic range of the signal but is related to fluorochrome abundance and depends on the magnitude of observed values.

[0012] A variety of unmixing processes in accordance with embodiments of the invention can be utilized that estimate fluorochrome abundances from fluorescence emission data in ways that assume noise is related to fluorochrome abundance including (but limited to) processes that approximate fluorochrome abundances utilizing a percentage error estimation via weighted least squares (WLS), processes that utilize a maximum likelihood-based solution directly employing Poisson regression to obtain fluorochrome abundances, processes that involve direct minimization of deviance, and/or minimization of Pearson residuals, and processes that approximate fluorochrome abundances by employing a Bar-Lev/Enis class of transformations. In various embodiments, the unmixing process utilizes a regression process in which a distance metric applied to a given optical measurement is weighted by a function of the given optical measurement. In many embodiments, the regression process is based upon a noise model including (but not limited to) a Poisson distributed noise, gamma distributed noise, Pólya distributed noise, and a negative binomial distributed noise.

[0013] In several embodiments, flow cytometers are configured so that individual detectors capture broader bandwidths of the emission spectrum to improve the performance of the unmixing process. In a number of embodiments, residuals

generated during the unmixing process can be utilized to gate the flow cytometry data during analysis. Data gather utilizing a variety of unmixing processes in accordance with embodiments of the invention is illustrated and described in the publication titled "Generalized Unmixing Model for Multispectral Flow Cytometry Utilizing Nonsquare Compensation Matrices" published in the Journal of the International Society for Advancement of Cytometry (Cytometry Part A 83A:508-520, 2013), the disclosure of which is incorporated by reference herein in its entirety.

[0014] Although much of the discussion that follows involves discussion of unmixing fluorescence emissions data, unmixing processes in accordance with embodiments of the invention can be utilized to estimate abundance information from any of a variety of optical data captured by flow cytometers including but not limited to fluorescence signals, Raman signals, and phosphorescence signals.

[0015] In order to better appreciate the significance of considering the relationship between signal variance and the observed signals in a flow cytometer system, the limitations of conventional unmixing processes when applied to over-determined systems are discussed below.

The Standard Model of Spectral Overlap

[0016] The linear-mixture model assumes that multiple signals measured from every particle can be expressed as a linear combination of spectral signatures. Accordingly, the standard mixing model can be represented using a basic linear spectral mixture equation:

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{e} \quad (1)$$

where \mathbf{r} is the normalized vector of length L of observations (digitized readouts from the detectors) for a bioparticle, where L the number of signals output by the detectors employed in the flow cytometry system

\mathbf{M} is an $L \times p$ spectral-signature matrix (p being the number of fluorochromes used in an experiment), which is equivalent to a mixing (spillover, spectral) matrix following appropriate normalization,

α is the vector of length p of fluorochrome abundances (or fractional abundances) for the p fluorochromes used to stain the bioparticles, and

\mathbf{e} is the vector of length L that denotes noise.

[0017] In contrast to imaging, the cytometry formulation of the problem usually does not refer to fractions (fractional abundances) but to an absolute value of abundance, which is often (however incorrectly) called “compensated fluorescence.” It is important to note that the basic spectral-mixing model as expressed by Eq. (1) in the general case is nonidentifiable, and consequently one cannot find a unique solution unless additional constraints and conditions are imposed. In remote sensing, and other imaging applications it is common to state explicitly that \mathbf{e} represents additive Gaussian noise with an expected value of zero. In the flow cytometry literature regarding compensation this is not stated; however, the praxis of compensation implicitly makes such an assumption.

[0018] If the mixing matrix \mathbf{M} is square and no additional constraints are imposed, the choice of distribution does not affect the solution, and the vector α can easily be found, producing the result known from standard cytometry practice:

$$\alpha = \mathbf{rM}^{-1} \quad (2)$$

[0019] If the error \mathbf{e} represents additive Gaussian noise with an expected value of zero, and the number of detectors is larger than the number of collected signals, the spectral unmixing is performed by solving a least-squares problem:

$$\min_{\alpha \in \square} \left\{ (\mathbf{r} - \mathbf{M}\alpha)^T (\mathbf{r} - \mathbf{M}\alpha) \right\} \quad (3)$$

[0020] Assuming no additional constraints, the linear unmixing process that attempts to recover a least-square approximation value of α is represented by the following closed-form equation:

$$\hat{\alpha} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{r} \quad (4)$$

[0021] Unfortunately, it is a common observation with flow cytometry data that the resulting least-square approximation value of α obtained using the above approach typically includes negative values for abundances that have no physical interpretation (as it is obvious that it is impossible to have negative fluorescence signal, or negative abundance). The problem is particularly acute with respect to weak fluorescence emission (such as but not limited to autofluorescence), which can be pushed below zero when minimizing for least squares error.

[0022] In order to avoid a solution that includes negative fluorochrome abundances, constraints can be imposed upon the unmixing process to ensure that all abundances are nonnegative. This constraint results in the following minimization problem for each particle:

$$\min_{\alpha} \left\{ (\mathbf{r} - \mathbf{M}\alpha)^T (\mathbf{r} - \mathbf{M}\alpha) \right\} \quad \text{s.t.} \quad \alpha \geq 0 \quad (5)$$

[0023] Additionally, it is common in many applications to require that the fractional abundances sum to 100% of the total signal.

[0024] The above expression does not have a closed form solution and so numerical techniques can be utilized to solve for the estimated fluorochrome abundances. Due to the constraints, impossible solutions are eliminated. A comparison of the unmixed results using equation (5) with simulated data demonstrates that the computed result is not, however, a good approximation of the true fluorochrome abundances, if the distribution of noise is not strictly Gaussian.

Signal Variance in Flow Cytometry Systems

[0025] The unconstrained and constrained least squares approximation techniques described above have the implicit assumption that the variance of the signal is stable across the whole range of observation values and will return the maximum-likelihood result if the noise distribution is Gaussian. Understanding the source of the observed negative values explains why systems and methods in accordance with embodiments of the invention can obtain significantly better approximations for fluorochrome

abundances from over-determined fluorescence emission data. Specifically, systems and methods in accordance with embodiments of the invention do not assume that variance of the signal is stable across the whole range of values. Instead, signal variance is assumed increasing with fluorochrome abundance. As is discussed further below, alternative noise models including (but not limited) to a noise model based upon a Poisson distribution can be utilized to model signal variance in a flow cytometry system. Based on these noise models, and a variety of unmixing processes can be utilized in accordance with embodiments of the invention that assume signal variance increases with increased fluorochrome abundance to achieve more accurate estimates of fluorochrome abundances.

Modeling Signal Variance in Flow Cytometry Systems

[0026] Flow cytometry involves detection of photons emitted by fluorescence molecules on the surface or inside of bioparticles. The detection of emitted photons can be considered to involve Poisson processes. Photons emitted by fluorochromes can be considered to arrive at random time intervals, where the probability that n photons strike a detector in a time interval t is closely approximated by a Poisson distribution. However, even if variance in the photon emission is assumed to be zero and that the photons arriving at the photocathode of a photomultiplier (a commonly used light detector in cytometry) are equally spaced in time, the number of emitted photoelectrons is not constant, as the probability of photoelectron emission is also governed by a Poisson process. Therefore, the expected variance of the signal is not stable, but increases with the abundances of the fluorochromes (i.e. the number of random photon emissions).

[0027] In addition, the probability that an emitted photon arrives at a specific detector is dependent upon the energy of the given photon and the filter arrangement used in the flow cytometer. In practice, owing to spectral overlap, two different fluorochromes can emit photons which are very close to each other or identical in terms of energy. Accordingly, a randomly emitted photon may arrive at a detector with a

probability P_1 , but may end up in another detector with a probability $(1-P_1)$. Therefore the mixing process occurs before the measurement is performed at the detector.

[0028] In the ideal case in which no additional noise sources are present and the detector offers 100% efficiency, the simplest model of the fluorescence emission data can be expressed as:

$$\mathbf{r} \sim \text{Poisson}(\mathbf{M}\alpha) \quad (6)$$

[0029] As noted above, the consequence of the model in equation (6) is that the expected variance of the signal is not stable, and increases with the fluorochrome abundances.

[0030] The goal of unmixing is to gain knowledge regarding the contribution of different fluorochromes to the total measured signal. The visualization approach commonly used in flow cytometry involving scatter plots, as well as the traditional terminology describing samples as “positive” or “negative,” suggests that practitioners are interested in minimizing the error of estimation for low-intensity signals (“negative” population) just as much as for high-intensity signals (“positive” population) when both are present in the mixture. The reason is that for Boolean classification of cells the “negative” and “positive” categories are equally important. Accordingly, unmixing processes in accordance with embodiments of the invention specifically address the fact that variance in the observed signal is not equal along the dynamic range but depends on the magnitude of observed values. Therefore, systems and methods in accordance with many embodiments of the invention consider the magnitude of an error in relation to the size of the observed fluorescence signal. Otherwise the error minimization can focus on estimating the “positive” sub-populations, at the cost of neglecting the correct estimation of abundances in “negative” sub-populations.

[0031] A variety of techniques for unmixing fluorescence emission signals are discussed below including (but limited to) processes that approximate fluorochrome abundances utilizing a percentage error estimation via weighted least squares (WLS), processes that utilize a maximum likelihood-based solution directly employing Poisson regression to obtain fluorochrome abundances, processes that involve direct minimization of deviance, and/or minimization of Pearson residuals, and processes that approximate fluorochrome abundances by employing a Bar-Lev/Enis transformation. Various processes for unmixing fluorescence emission data to obtain fluorochrome abundances in accordance with embodiments of the invention are discussed further below.

Unmixing Using Weighted Least-Squares and Percentage Error Estimation

[0032] In several embodiments, the unmixing process assumes that the observations came from a normal distribution. However, an additional assumption is made that signal variance in the Gaussian model grows with signal intensity. Consequently, measurements with lower variance have proportionally more influence on abundance estimates than measurements with higher variance. In a number of embodiments, the unmixing process involves performing a percentage errors minimization process.

[0033] In several embodiments, a mean absolute percentage error (MAPE) minimization is performed. A MAPE minimization defines percentage error as $(\text{observed value} - \text{predicted value})/(\text{predicted value})$. Since the predicted value is the value that the process aims to find, the minimization can be performed as an iterative process. In certain embodiments, an iterative reweighted least squares (IRLS) process is used to perform the iterations.

[0034] An alternative formulation of MAPE defines this value as (observed value – predicted value)/(observed value). Owing to this reformulation, a closed-form solution which minimizes MAPE can be found. Using this alternative formulation, the error E_p can be redefined as:

$$E_p = \frac{1}{n} \left\| \frac{\mathbf{r} - \mathbf{M}\boldsymbol{\alpha}}{\mathbf{r}} \right\|_1$$

where n is the number of elements in vector \mathbf{r} , and

$\frac{\mathbf{x}}{\mathbf{y}}$ is the Hadamard division (or element by element division) of the vectors

\mathbf{x} and \mathbf{y} .

[0035] The minimization problem can be rewritten as:

$$\min_{\boldsymbol{\alpha}} \{ (\mathbf{W}\mathbf{r} - \mathbf{W}\mathbf{M}\boldsymbol{\alpha})^T (\mathbf{W}\mathbf{r} - \mathbf{W}\mathbf{M}\boldsymbol{\alpha}) \}$$

where \mathbf{W} is a diagonal matrix with $1/r_j$ values:

$$\mathbf{W} = \begin{pmatrix} \frac{1}{r_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{r_L} \end{pmatrix}$$

[0036] The term that is minimized can be rewritten as:

$$(\mathbf{W}\mathbf{r} - \mathbf{W}\mathbf{M}\boldsymbol{\alpha})^T (\mathbf{W}\mathbf{r} - \mathbf{W}\mathbf{M}\boldsymbol{\alpha}) = (\mathbf{W}\mathbf{r})^T \mathbf{W}\mathbf{r} - (\mathbf{W}\mathbf{r})^T \mathbf{W}\mathbf{M}\boldsymbol{\alpha} - (\mathbf{W}\mathbf{M}\boldsymbol{\alpha})^T \mathbf{W}\mathbf{r} + \boldsymbol{\alpha}^T \mathbf{M}^T \mathbf{W}^2 \mathbf{M}\boldsymbol{\alpha}$$

[0037] In order to find a closed-form solution, the above term is differentiated with respect to the abundances vector and equated to zero:

$$-(\mathbf{W}\mathbf{r})^T \mathbf{W}\mathbf{M} + \mathbf{M}^T \mathbf{W}^2 \mathbf{M}\boldsymbol{\alpha} = \mathbf{0}$$

[0038] The solution for the above equation provides the following estimation of α :

$$\hat{\alpha} = (\mathbf{M}^T \mathbf{W}^2 \mathbf{M})^{-1} \mathbf{M}^T \mathbf{W}^2 \mathbf{r}$$

[0039] The weights in the matrix \mathbf{W} are inversely proportional to the signal, providing a simple solution that recognizes that the increase of variance (uncertainty) increases with the signal. The MAPE minimization yields a closed form solution that can be utilized in an unmixing process in accordance with embodiments of the invention. An alternative to MAPE and other least squares approximation methods is to utilize a generalized linear model, which explicitly allows for various non-Gaussian distributions of the random component.

Unmixing Processes Utilizing Generalized Linear Models

[0040] Generalized linear model processes in accordance with embodiments of the invention attempt to fit observed fluorescence emission data by the method of maximum likelihood estimation instead of least squares approximation techniques. Accordingly, generalized linear models can be utilized to perform unmixing where the noise is assumed not normally distributed.

[0041] Signal formation can be seen as a stochastic Poisson process. Therefore, it is expected that distribution of noise will be well approximated by a Poisson distribution. However, this simplest approximation typically does not represent the experimental reality very well. Flow cytometry instruments are usually not equipped with detectors capable of counting photons and reporting them directly, but rather convert light into analog electronic signals (even though this information is subsequently digitized). Therefore the “raw” readout is represented as real rather than natural numbers.

[0042] The assumption of purely Poisson-distributed flow cytometer data is also a problem for *in-silico* experiments and simulations. If no mixing occurs, a simulation utilizing a Poisson random-number generator produces only integer values illustrating the number of photons, and then the number of photoelectrons generated at each detector. The mixing process indeed generates real numbers as an artifact of matrix

multiplication, but these are subsequently truncated when processed by a Poisson random-number generator when the detection step is simulated. A simplistic solution for the purpose of simulation might involve the addition of white noise to the Poisson signal. However, the true fluorescence emission signals collected using a flow cytometer would be continuous even if no readout noise was present.

[0043] Simulating the true continuous distribution of analog signals produced by a photo multiplier tube is quite difficult, as the Poisson model is not appropriate if the secondary emission statistics are taken under consideration. It has been demonstrated that these effects can be described by the Pólya distribution. However, assuming a completely noiseless secondary emission process in which gain does not vary for different photoelectrons, the fluorescence emission data can be approximated using a simple continuous generalization of a Poisson distribution.

[0044] Therefore, the flow cytometer data acquisition can be simulated using a formulation of a Poisson distribution in which the factorial is replaced by a function Gamma:

$$P_{\mu}^{\text{cont}}(y) = \frac{\mu^y \exp(-\mu)}{\Gamma(y+1)}$$

[0045] The resultant distribution is a Gamma distribution with shape parameter $a=y+1$, and scale parameter $s=1$.

$$\text{Gamma}(x; a, s) = \frac{1}{s^a} \frac{1}{\Gamma(a)} x^{a-1} \exp\left(-\frac{x}{s}\right),$$

$$\text{Gamma}(\mu; y+1, 1) = \frac{\mu^y \exp(-\mu)}{\Gamma(y+1)}.$$

[0046] Therefore,

$$\text{Gamma}(\mu; y+1, 1) = P_{\mu}^{\text{cont}}(y).$$

[0047] The continuous Poisson distribution P^{cont} can be expressed as an exponential distribution:

$$P_{\mu}^{\text{cont}}(y) = \frac{\mu^y \exp(-\mu)}{\Gamma(y+1)} = \exp\{y \log \mu - \mu - \log(\Gamma(y+1))\}$$

[0048] The log-likelihood function L of this distribution is:

$$\mathcal{L}(\mathbf{y}; \boldsymbol{\mu}) = \sum_i y_i \log \mu_i - \mu_i - \log(\Gamma(y_i + 1))$$

[0049] In order to recover the fluorochrome abundances $\boldsymbol{\alpha}$ the function L is minimized with respect to the regression parameters.

[0050] The deviance D can be understood as a generalization of the residual sum of squares used in the case of linear models. Consequently, in the case of the continuous Poisson distribution P^{cont} , the deviance is

$$\mathcal{D}(\mathbf{y}, \boldsymbol{\mu}) = 2(\mathcal{L}(\mathbf{y}; \mathbf{y}) - \mathcal{L}(\mathbf{y}; \boldsymbol{\mu})) = 2 \sum_i \left[y_i \log \left(\frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right] \quad (8)$$

[0051] Despite the log function being the canonical link for Poisson generalized linear model, the specific problem of multispectral flow cytometry involves use of identity-link Poisson regression. This is motivated by the fact that the Poisson parameters of the observed fluorescence signal are linear functions of the fluorochrome abundances vector $\boldsymbol{\alpha}$. Consequently, finding the maximum-likelihood estimates of $\boldsymbol{\alpha}$ involves a Poisson regression with an identity-link function, rather than a log link function. Use of the identity link also means that the simplest and very common approach to a Poisson regression by iterative reweighted least squares often suffers from lack of convergence. The problem can be solved by implementing a modified IRLS approach as discussed by Marschner in Marschner, I. C. (2010) Stable Computation of Maximum Likelihood Estimates in Identity Link Poisson Regression, *Journal of Computational and Graphical Statistics* 19, 666–683 (the relevant disclosure of which is incorporated by reference). In other embodiments, various approaches to stable computation of maximum-likelihood estimates in identity-link Poisson regression can be utilized in accordance with embodiments of the invention to estimate fluorochrome abundances in accordance with embodiments of the invention.

[0052] In several embodiments, following Equations (1) and (8) an identity-link Poisson regression is used in which deviance is expressed as:

$$\mathcal{D} = 2\mathbf{j}^T \left(\mathbf{r} \circ \log \left(\frac{\mathbf{r}}{\mathbf{M}\boldsymbol{\alpha}} \right) - (\mathbf{r} - \mathbf{M}\boldsymbol{\alpha}) \right), \quad (9)$$

where \mathbf{j} is an $L \times 1$ sum vector of 1, and \mathbf{j}^T is its transpose (the sum vector is used to find the sum of the elements of the computed vector),

$\log(\mathbf{X})$ is the element-wise logarithm of \mathbf{X} ,

$\frac{\mathbf{x}}{\mathbf{y}}$ is the Hadamard division (or element by element division) of vectors \mathbf{x}

and \mathbf{y} , and

the operator \circ denotes element-wise multiplication (Hadamard product).

[0053] In contrast to the least squares approach used with Gaussian regression, the minimization of deviance in the Poisson regression problem has no general closed-form solution. Therefore, the $\boldsymbol{\alpha}$ vector is found using optimization methods such as Nelder and Mead method, Broyden, Fletcher, Goldfarb and Shanno algorithm and others

[0054] The minimization of the objective function in Eq. (9) does not guarantee the nonnegativity of $\boldsymbol{\alpha}$. Therefore an additional nonnegativity constraint assuring concordance with the physical model may be imposed.

[0055] Furthermore, additional constraints can be used, such as a sum-to-one equivalent constraint can be added as a soft penalty:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \left\{ 2\mathbf{j}^T \left(\mathbf{r} \circ \log \left(\frac{\mathbf{r}}{\mathbf{M}\boldsymbol{\alpha}} \right) - (\mathbf{r} - \mathbf{M}\boldsymbol{\alpha}) \right) + \lambda \left| \|\mathbf{r}\|_1 - \|\boldsymbol{\alpha}\|_1 \right| \right\}$$

s.t. $\boldsymbol{\alpha} > 0$

[0056] The penalty parameter λ allows control of the level of certainty in the model. This parameter can be set to 0 or to some very low value if the accuracy (or completeness) of \mathbf{M} is suspect. In other words, in an experimental setting in which not all the fluorochromes present are known, the entire signal will not be unmixed utilizing

only the spectra describing the known fluorochromes. Although specific processes for estimating fluorochrome abundances based upon a Poisson generalized linear model approach are described above, any of a variety of processes based upon a Poisson and/or continuous Poisson signal model can be utilized in accordance with embodiments of the invention.

Unmixing Processes Using Pearson Residuals

[0057] The deviance is not the only measure of goodness of fit employed with generalized linear models. Pearson residuals are another commonly used measure of overall fit for generalized linear models. Pearson residuals are defined to be the standardized difference between the observed and the predicted values. Therefore, the Pearson residual is the raw residual divided by the square root of the variance function

[0058] The minimization of the sum of squared Pearson residuals for the Poisson regression problem provides the following approximation of α :

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ \mathbf{j}^T \left(\frac{(\mathbf{r} - \mathbf{M}\alpha)^2}{\mathbf{M}\alpha} \right) \right\} \quad \text{s.t.} \quad \alpha > 0,$$

Unmixing Processes Involving Variance-Stabilizing Transformations

[0059] Unmixing processes in accordance with several embodiments of the invention can perform unmixing of Poisson-distributed measurements using a least squares estimation processes following the transformation of the mixing model into approximately Gaussian. The optimal transformation proposed by Bar-Lev and Enis, or Anscombe and Freeman–Tukey transformations (belonging to a wider class of variance-stabilization functions described by Bar-Lev and Enis), can be used for this purpose.

[0060] The Bar-Lev-Enis transformation is defined as

$$\mathcal{A}_{a,b}(x) = (x + 2a - b)(x + a)^{-\frac{1}{2}},$$

$$\mathcal{A}_{a,b,c}(x) = \mathcal{A}_{a,b}(x) + (x + c)^{-\frac{1}{2}}.$$

[0061] The transformation has been shown to exhibit optimal variance-stabilizing performance for a Poisson distribution for

$$a = \frac{3}{8} + \frac{3^{-\frac{1}{2}}}{2}, \quad b = \frac{3}{8} + \frac{3^{\frac{1}{2}}}{4}$$

[0062] Therefore, fluorochrome abundance estimates can be obtained by solving the following expression:

$$\hat{\alpha} = \arg \min_{\alpha} \left\| \mathcal{A}_{a,b}(\mathbf{r}) - \mathcal{A}_{a,b}(\mathbf{M}\alpha) \right\|_2^2 \quad \text{s.t.} \quad \alpha > 0$$

[0063] Although a variety of unmixing techniques are described above, any of a variety of techniques for estimating fluorochrome abundances using unmixing processes that account for increases in signal variance with increases in fluorochrome abundances can be utilized in accordance with embodiments of the invention. Systems that utilize unmixing processes in accordance with embodiments of the invention, modifications that can be performed to convention flow cytometers to enhance performance using the unmixing processes, and additional flow cytometer data analysis techniques that are enabled using the residuals generated during the unmixing processes are discussed further below.

Systems for Analyzing Flow Cytometer Data

[0064] Flow cytometry systems including data analysis systems in accordance with embodiments of the invention capture fluorescence emission data for bioparticles labeled with multiple fluorochromes using an over-determined system of detectors. The flow cytometry systems can then utilize an unmixing process that accounts for the increase in signal variance with fluorochrome abundance to estimate fluorochrome abundances with respect to each bioparticle.

[0065] A data analysis system in accordance with an embodiment of the invention is illustrated in FIG. 1. The data analysis system 10 includes a flow cytometer 12. As noted above, the flow cytometer is configured as an over-determined system. Stated another way, the flow cytometer is configured so that the number of signals produced by

the detectors is greater than the number of fluorochromes staining the bioparticles observed by the detectors. In many embodiments, the flow cytometer utilizes an optics and detection system to separate optical emission with respect to a predetermined set of spectral ranges using a number of detectors. As can be readily appreciated, any conventional flow cytometer including an appropriate number of detectors can be configured as an over-determined system in accordance with embodiments of the invention. In the illustrated embodiment, the flow cytometer is configured to provide data to a data analysis computer 14 via a network 16. In many embodiments, the data analysis computer is a personal computer, server, and/or any other computing device with the storage capacity and processing power to analyze the data output by the flow cytometer. The analysis computer includes a processor, memory and/or a storage system containing an optical data analysis application that includes machine readable instructions that configures the computer to generate a mixing model from control samples and to apply an unmixing process to fluorescence emission data captured by the flow cytometer based upon the mixing model and the assumption that signal variance in the fluorescence emission data increases with the signal. Although a specific data analysis system is illustrated in FIG. 1, any of a variety of data analysis systems can be utilized to analyze fluorescence emission data captured by a flow cytometer configured as an over-determined system in accordance with embodiments of the invention. In many embodiments, data acquisition systems can be included and/or attached to a flow cytometer and used to perform unmixing processes in accordance with embodiments of the invention.

Processes for Unmixing Fluorescence Emission Data

[0066] Processes for performing unmixing of fluorescence emission data in ways that account for the relationship between the noise in the fluorescence emission data and the fluorochrome abundances are described extensively above. A process for estimating fluorochrome abundances in accordance with an embodiment of the invention is illustrated in FIG. 2. The process 20 includes obtaining (22) control

fluorescence emission data for single stained controls. Fluorescence emission data is obtained (24) for bioparticles stained with multiple fluorochromes. The fluorescence emission data is obtained using a number of detectors configured to produce a number of fluorescence emission observations that is greater than the number of fluorochromes used to stain the bioparticles. The control fluorescence emission data can be obtained using the flow cytometer used to capture the fluorescence emission data. In several embodiments, however, the control fluorescence emission data can be the theoretical spectrum of a fluorochrome, a reference spectrum for a fluorochrome, and/or a spectrum obtained using another instrument.

[0067] The control fluorescence emission data is utilized to generate (26) a mixing model, which is used in the estimation of fluorochrome abundances from the fluorescence emission data. In many embodiments, fluorochrome abundances are estimated (28) by performing an unmixing process similar to the unmixing processes described above that account for the increase in the variance in the noise in fluorescence emission data with increased fluorochrome abundance.

[0068] Although a specific process is described above with respect to FIG. 2, any of a variety of flow cytometry processes that involve the capture of fluorescence emission data using an over-determined system and the unmixing of the fluorescence emission data using a process that accounts for the increase in the variance in the noise in fluorescence emission data with increased fluorochrome abundance can be utilized in accordance with embodiments of the invention. As can readily be appreciated, the use of unmixing processes in accordance with embodiments of the invention can prompt modification of the conventional manner in which flow cytometers are configured to capture fluorescence emission data and the manner in which data captured by a flow cytometer is analyzed. Techniques for configuring flow cytometers and processes for analyzing fluorescence emission data captured by flow cytometers in accordance with embodiments of the invention are discussed further below.

Modifying Flow Cytometer Configuration for Over-Determined Operation

[0069] The optical pathways employed in majority of current commercial flow cytometers use a set of bandpass and dichroic filters to separate the signal into appropriate wavelength ranges. Fluorescence emission passes through bandpass filters of a desired wavelength or another dichroic filter to be eventually recorded by a photodetector. The resultant electronic signal is then digitized and the digitized value stored. The photodetectors employed are typically photodiodes photomultiplier tubes (PMTs), avalanche photodiodes, or CCD arrays. As noted above, conventional approaches to the detection of fluorescence emission data have involved using fluorochromes with distinct emission peaks and tuning the bandpass filters of the detector system so that a single detector is tuned to detect fluorescence emissions in band corresponding to the emission peak of a single fluorochrome. When the same flow cytometer is configured as an over-determined system in accordance with embodiments of the invention, the effectiveness of the system in estimating fluorochrome abundances can be improved by increasing the bandwidth of the fluorescence emission spectra observed by each of the detectors. Instead of only observing the emission peaks of the individual fluorochromes and discarding the information contained between the peaks, the detector system can be configured to capture as much information concerning the emission spectra of the fluorochromes as is allowed by the instrument. When a flow cytometer is configured in this way, the additional information can be used to increase the accuracy of the estimate of the fluorochrome abundances obtained through the unmixing process.

Utilizing Residuals in Data Analysis

[0070] A feature of using an over-determined system to obtain fluorescence emission data in flow cytometry is that the process of estimating the fluorochrome abundances produces a residual. The residual for a specific bioparticle provides information concerning how well the estimated fluorochrome abundances, multiplied by the mixing matrix, reconstruct the observed fluorescence emission data. This

information can be extremely useful as a diagnostic tool. In many embodiments, a data analysis computer can be configured using software that enables the analysis of flow cytometry data using gates that gate the flow cytometry data based upon residuals determined during estimation of fluorochrome abundances. As can be readily appreciated, the ability to analyze subpopulations of bioparticles based upon how well the actual observed values from the detectors match the estimated fluorochrome abundances can be extremely useful in isolating or excluding subpopulations of bioparticles when analyzing flow cytometry data. For example, if a certain subpopulation of cells have estimated observations that are much farther from the true observations than the rest of the population, this likely means that the experimentally determined mixing matrix based on the single stained controls is not appropriate for these cells. As such, there must be additional physiological processes occurring in these cells to render the mixing matrix invalid. The difference between the estimated and true observation can be calculated in many ways, either as a least squares residual, a more generalized deviance or many other techniques used for assessing the difference between two vectors.

[0071] While the above description contains many specific embodiments of the invention, these should not be construed as limitations on the scope of the invention, but rather as an example of one embodiment thereof. For example, systems and methods in accordance with embodiments of the invention can be utilized in the unmixing of any optical signal captured by a flow cytometer including but not limited to a fluorescence signal, a Raman signal, and a phosphorescence signal. Accordingly, the scope of the invention should be determined not by the embodiments illustrated, but by the appended claims and their equivalents.

WHAT IS CLAIMED IS:

1. A data analysis system configured to analyze optical data captured by a flow cytometer with respect to a plurality of particles stained with a plurality of fluorochromes, where an optics and detection system within the flow cytometer separates optical emission with respect to spectral ranges and where at least one detector is used to capture a number of optical measurements that is greater than the plurality of fluorochromes used to stain the plurality of particles, the data analysis system comprising:

a processor;

a memory connected to the processor and configured to store an optical data analysis application, wherein the optical data analysis application configures the processor to:

obtain control optical data for at least one particle stained with at least one fluorochrome selected from a set of fluorochromes, where the control optical data is captured by the flow cytometer configured so that an optics and detection system within the flow cytometer separates optical emission with respect to a predetermined set of spectral ranges;

generate a mixing model using the obtained control optical data and a system of linear combinations;

obtain experimental optical data for particles stained with the set of fluorochromes, where the experimental optical data is captured by the flow cytometer configured so that an optics and detection system within the flow cytometer separates optical emission with respect to the predetermined spectral ranges using at least one detector configured to capture a number of optical measurements and the number of optical measurements is greater than the number of fluorochromes in the set of fluorochromes; and

estimate abundances of the fluorochromes in the set of fluorochromes using the obtained experimental optical data by solving an overdetermined system of equations to unmix the optical data, based upon the generated mixing model that accounts for increased noise variance with increased fluorochrome abundance.

2. The data analysis system of claim 1, wherein the optical data analysis application further configures the processor to obtain control optical data for the at least one particle stained using a single fluorochrome selected from the set of fluorochromes.

3. The data analysis system of claim 1, wherein the optical data can be captured from optical signals that can be selected from the group consisting of fluorescence signals, Raman signals, and phosphorescence signals.

4. The data analysis system of claim 1, wherein each of the number of detectors are tuned to capture optical emissions over a spectrum as wide as allowed by the flow cytometer.

5. The data analysis system of claim 1, wherein the optical data analysis application further configures the processor to estimate fluorochrome abundances by utilizing a percentage error estimation via a weighted least squares method.

6. The data analysis system of claim 1, wherein the optical data analysis application further configures the processor to estimate fluorochrome abundances by utilizing a percentage errors minimization process.

7. The data analysis system of claim 6, wherein the optical data analysis application further configures the processor to estimate fluorochrome abundances by utilizing a mean absolute percentage errors minimization process using:

$$\hat{\alpha} = (\mathbf{M}^T \mathbf{W}^2 \mathbf{M})^{-1} \mathbf{M}^T \mathbf{W}^2 \mathbf{r}$$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles, \mathbf{M}^T is the transpose of the matrix \mathbf{M} , \mathbf{r} is a normalized vector of length L of optical data observations, and \mathbf{W} is a diagonal matrix with $1/r_j$ values such that:

$$\mathbf{W} = \begin{pmatrix} \frac{1}{r_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{r_L} \end{pmatrix}$$

8. The data analysis system of claim 1, wherein the optical data analysis application further configures the processor to estimate fluorochrome abundances by utilizing a maximum likelihood-based Poisson regression using:

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ 2\mathbf{j}^T \left(\mathbf{r} \circ \log \left(\frac{\mathbf{r}}{\mathbf{M}\alpha} \right) - (\mathbf{r} - \mathbf{M}\alpha) \right) + \lambda \left(\|\mathbf{r}\|_1 - \|\alpha\|_1 \right) \right\}$$

s.t. $\alpha > 0$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{j} is an $L \times 1$ sum vector of 1 where L is the number of optical data observations, and \mathbf{j}^T is the transpose of the vector \mathbf{j} , \mathbf{r} is a normalized vector of length L of optical data observations, operator \circ denotes element-wise multiplication, α is a vector of length p of fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations

and p is the number of fluorochromes used to stain the particles, and λ is a penalty parameter that allows for control of the level of certainty in the model.

9. The data analysis system of claim 1, wherein the optical data analysis application further configures the processor to estimate fluorochrome abundances by minimizing Pearson residuals using:

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ \mathbf{j}^T \left(\frac{(\mathbf{r} - \mathbf{M}\alpha)^2}{\mathbf{M}\alpha} \right) \right\} \quad \text{s.t. } \alpha > 0,$$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{j} is an $L \times 1$ sum vector of 1 where L is the number of optical data observations, and \mathbf{j}^T is the transpose of the vector \mathbf{j} , and \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles.

10. The data analysis system of claim 1, wherein the optical data analysis application further configures the processor to estimate fluorochrome abundances by utilizing a Bar-Lev/Enis class of transformations and using:

$$\hat{\alpha} = \arg \min_{\alpha} \left\| \mathcal{A}_{a,b}(\mathbf{r}) - \mathcal{A}_{a,b}(\mathbf{M}\alpha) \right\|_2^2 \quad \text{s.t. } \alpha > 0$$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{r} is a normalized vector of length L of optical data observations, \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles, α is the vector of length p where p is the number of fluorochromes used to stain the particles, and the Bar-Lev/Enis transformation is defined as:

$$\mathcal{A}_{a,b}(x) = (x + 2a - b)(x + a)^{-\frac{1}{2}},$$

$$\mathcal{A}_{a,b,c}(x) = \mathcal{A}_{a,b}(x) + (x + c)^{-\frac{1}{2}}.$$

11. The data analysis system of claim 1, wherein the processor being configured by the data analysis application to use an unmixing process that accounts for increased noise variance with increased fluorochrome abundance further comprises using a regression process in which a distance metric applied to a given optical measurement is weighted by a function of the given optical measurement.

12. The data analysis system of claim 11, wherein the regression process is based upon a noise model selected from the group consisting of Poisson distributed noise, gamma distributed noise, Pólya distributed noise, and negative binomial distributed noise.

13. The data analysis system of claim 1, wherein the processor being configured by the data analysis application to use an unmixing process that accounts for increased noise variance with increased fluorochrome abundance further comprises using a regression process in which a distance metric applied to a given optical measurement is weighted by a function of the predicted value for the given optical measurement.

14. The data analysis system of claim 13, wherein the data analysis application utilizes an iterative percentage errors minimization process.

15. The data analysis system of claim 1, wherein the least one detector configured to capture a number of optical measurements are multiple CCD detectors.

16. The data analysis system of claim 1, wherein the least one detector configured to capture a number of optical measurements is a single CCD array detector.

17. A method for analyzing optical data captured by a flow cytometer with respect to a plurality of particles stained with a plurality of fluorochromes, where an optics and detection system within the flow cytometer separates optical emission with respect to spectral ranges and where at least one detector is used to capture a number of optical measurements that is greater than the plurality of fluorochromes used to stain the plurality of particles, using a data analysis system, the method comprising:

obtaining control optical data for at least one particle stained with at least one fluorochrome selected from a set of fluorochromes using the data analysis system, where the control optical data is captured utilizing the flow cytometer configured so that an optics and detection system within the flow cytometer separates optical emission with respect to a predetermined set of spectral ranges;

generating a mixing model using the obtained control optical data and a system of linear combinations using the data analysis system;

obtaining experimental optical data for particles stained with the set of fluorochromes using the data analysis system, where the experimental optical data is captured utilizing the flow cytometer configured so that an optics and detection system within the flow cytometer separates optical emission with respect to the predetermined spectral ranges using at least one detector configured to capture a number of optical measurements and the number of optical measurements is greater than the number of fluorochromes in the set of fluorochromes; and

estimating abundances of the fluorochromes in the set of fluorochromes using the obtained experimental optical data by solving an overdetermined system of equations to unmix the optical data using the data analysis system, based upon the generated mixing model that accounts for increased noise variance with increased fluorochrome abundance.

18. The method of claim 17, wherein the obtaining control optical data for at least one particle stained with at least one fluorochrome selected from a set of fluorochromes using the data analysis system further comprises selecting a single fluorochrome from the set of fluorochromes using the data analysis system.

19. The method of claim 17, wherein the optical data can be captured from optical signals that can be selected from the group consisting of fluorescence signals, Raman signals, and phosphorescence signals using the data analysis system.

20. The method of claim 17, wherein each of the number of detectors are tuned to capture optical emissions over a spectrum as wide as allowed by the flow cytometer using the data analysis system.

21. The method claim 17, wherein the estimating abundances of the fluorochromes in the set of fluorochromes using the data analysis system further comprises utilizing a percentage error estimation via a weighted least squares method using the data analysis system.

22. The method of claim 17, wherein the estimating abundances of the fluorochromes in the set of fluorochromes using the data analysis system further comprises utilizing a percentage errors minimization process using the data analysis system.

23. The method of claim 22, wherein the estimating abundances of the fluorochromes in the set of fluorochromes using the data analysis system further comprises using the data analysis system to utilize a mean absolute percentage error minimization process and a formula defined such that:

$$\hat{\alpha} = (\mathbf{M}^T \mathbf{W}^2 \mathbf{M})^{-1} \mathbf{M}^T \mathbf{W}^2 \mathbf{r}$$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles, \mathbf{M}^T is the transpose of the matrix \mathbf{M} , \mathbf{r} is a normalized vector of length L of optical data observations, and \mathbf{W} is a diagonal matrix with $1/r_j$ values such that:

$$\mathbf{W} = \begin{pmatrix} \frac{1}{r_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{r_L} \end{pmatrix}$$

24. The method of claim 17, wherein the estimating abundances of the fluorochromes in the set of fluorochromes using the data analysis system further comprises using the data analysis system to utilize a maximum likelihood-based using a Poisson regression and a formula defined such that:

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ 2\mathbf{j}^T \left(\mathbf{r} \circ \log \left(\frac{\mathbf{r}}{\mathbf{M}\alpha} \right) - (\mathbf{r} - \mathbf{M}\alpha) \right) + \lambda \left(\|\mathbf{r}\|_1 - \|\alpha\|_1 \right) \right\}$$

s.t. $\alpha > 0$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{j} is an $L \times 1$ sum vector of 1 where L is the number of optical data observations, and \mathbf{j}^T is the transpose of the vector \mathbf{j} , \mathbf{r} is a normalized vector of length L of optical data observations, operator \circ denotes element-wise multiplication, α is a vector of length p of fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles, and λ is a penalty parameter that allows for control of the level of certainty in the model.

25. The method of claim 17, wherein the estimating abundances of the fluorochromes in the set of fluorochromes using the data analysis system further comprises using the data analysis system to minimize Pearson residuals and to utilize a formula defined such that:

$$\hat{\alpha} = \arg \min_{\alpha} \left\{ \mathbf{j}^T \left(\frac{(\mathbf{r} - \mathbf{M}\alpha)^2}{\mathbf{M}\alpha} \right) \right\} \quad \text{s.t. } \alpha > 0,$$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{j} is an $L \times 1$ sum vector of 1 where L is the number of optical data observations, and \mathbf{j}^T is the transpose of the vector \mathbf{j} , and \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles.

26. The method of claim 17, wherein the estimating abundances of the fluorochromes in the set of fluorochromes using the data analysis system further comprises using the data analysis system to utilize a Bar-Lev/Enis class of transformations and a formula such that:

$$\hat{\alpha} = \arg \min_{\alpha} \left\| \mathcal{A}_{a,b}(\mathbf{r}) - \mathcal{A}_{a,b}(\mathbf{M}\alpha) \right\|_2^2 \quad \text{s.t. } \alpha > 0$$

where $\hat{\alpha}$ is a vector of length p of the estimated fluorochrome abundances where p is the number of fluorochromes used to stain the particles, \mathbf{r} is a normalized vector of length L of optical data observations, \mathbf{M} is an $L \times p$ spectral-signature matrix where L is the number of optical data observations and p is the number of fluorochromes used to stain the particles, α is the vector of length p where p is the number of fluorochromes used to stain the particles, and the Bar-Lev/Enis transformation is defined as:

$$\mathcal{A}_{a,b}(x) = (x + 2a - b)(x + a)^{\frac{1}{2}},$$

$$\mathcal{A}_{a,b,c}(x) = \mathcal{A}_{a,b}(x) + (x + c)^{\frac{1}{2}}.$$

27. The method of claim 17, wherein the processor being configured by the data analysis application to use an unmixing process that accounts for increased noise variance with increased fluorochrome abundance further comprises using a regression process in which a distance metric applied to a given optical measurement is weighted by a function of the given optical measurement.

28. The method of claim 27, wherein the regression process is based upon a noise model selected from the group consisting of Poisson distributed noise, gamma distributed noise, Pólya distributed noise, and negative binomial distributed noise.

29. The method of claim 17, wherein the processor being configured by the data analysis application to use an unmixing process that accounts for increased noise variance with increased fluorochrome abundance further comprises using a regression process in which a distance metric applied to a given optical measurement is weighted by a function of the predicted value for the given optical measurement.

30. The method of claim 29, wherein the data analysis application further utilizes an iterative percentage errors minimization process.

31. The method of claim 17, wherein the least one detector configured to capture a number of optical measurements further comprises multiple CCD detectors.

32. The method of claim 17, wherein the least one detector configured to capture a number of optical measurements further comprises a single CCD array detector.

1/2

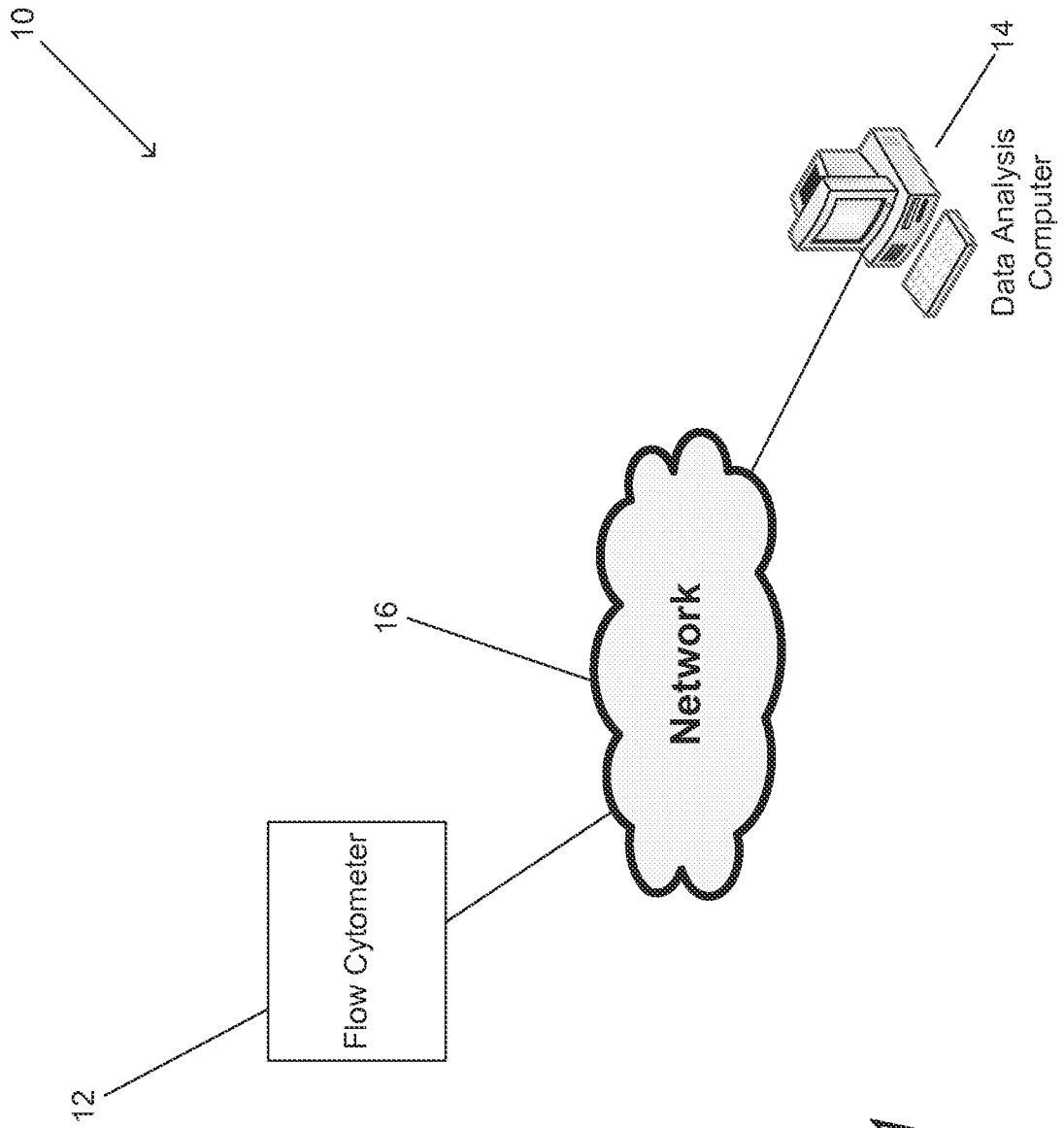


FIG. 1

2/2

20
↘

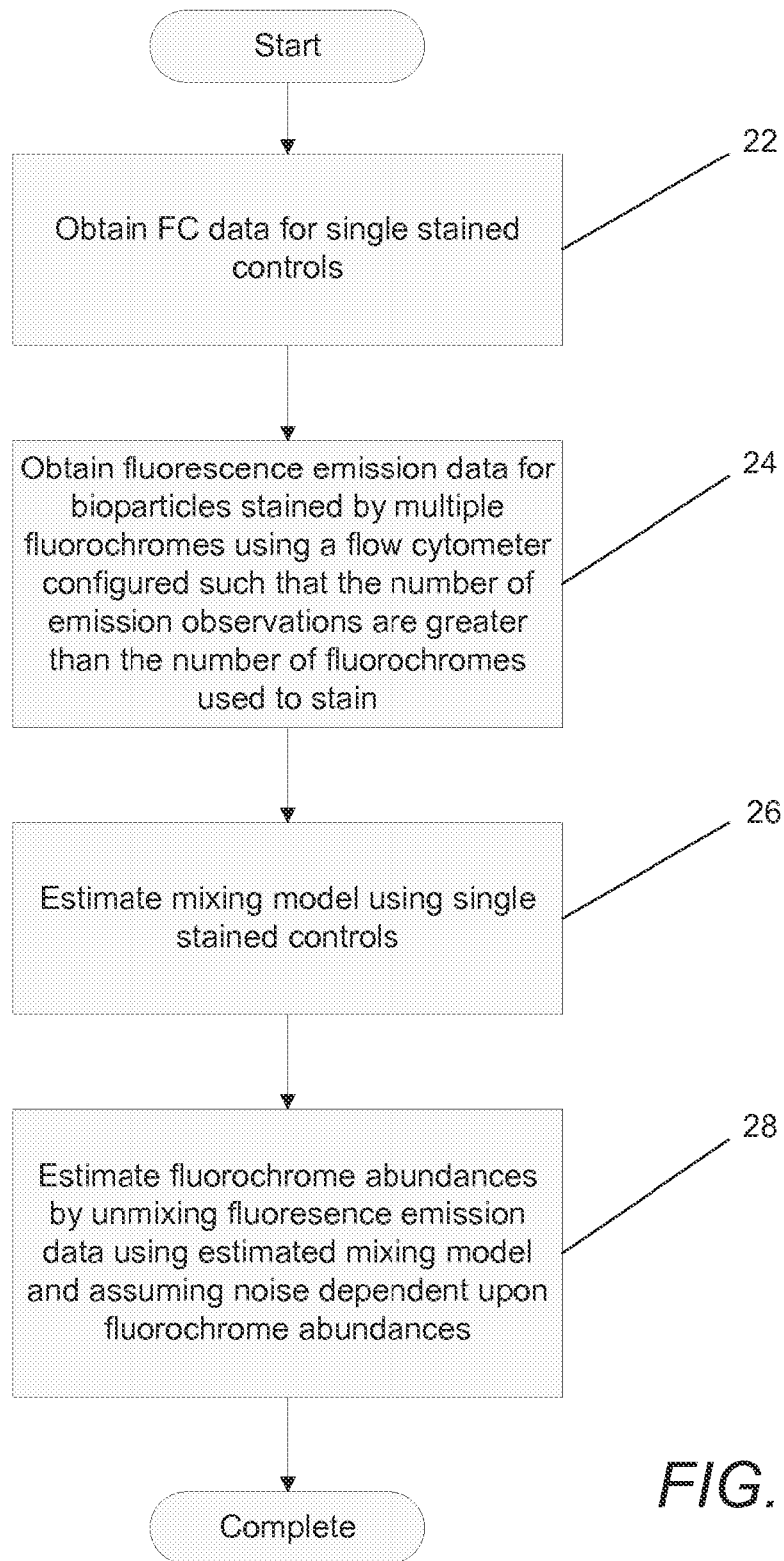


FIG. 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US13/47132

<p>A. CLASSIFICATION OF SUBJECT MATTER IPC(8) - G01N 21/64; G01J 3/30, 3/42 (2013.01) USPC - 356/311, 317; 436/10 According to International Patent Classification (IPC) or to both national classification and IPC</p>																	
<p>B. FIELDS SEARCHED</p> <p>Minimum documentation searched (classification system followed by classification symbols) IPC(8) Classification(s): G01N 21/64; G01J 3/30; G06M 11/00 (2013.01) USPC Classification(s): 356/300, 311, 317; 436/8, 10</p> <p>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched</p> <p>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) MicroPatent (US Granted, US Applications, EP-A, EP-B, WO, JP, DE-G, DE-A, DE-T, DE-U, GB-A, FR-A); DialogPro (Derwent, INSPEC, NTIS, PASCAL, Current Contents Search, Dissertation Abstracts Online, Inside Conferences); IP.com; IEEE.com; Google Scholar cytometry overdetermined least weighted squares, Flow cytometer, Fluorochrome, spectra detector *determined system of equations</p>																	
<p>C. DOCUMENTS CONSIDERED TO BE RELEVANT</p> <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>Y --- A</td> <td>US 2010/0302536 A1 (BALL, J. et al.) December 2, 2010 figures 1-4, paragraphs [0013]-[0037].</td> <td>1-7, 11-23, 27-32 --- 8-10, 24-26</td> </tr> <tr> <td>Y --- A</td> <td>HUTH, Y. et al. Fourier Transformed Spectral Bio-imaging for Studying the Intracellular Fate of Liposomes. Cytometry Part A, Vol. 57A, No. 1, pages 10-21 [online], 2004, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://onlinelibrary.wiley.com/doi/10.1002/cyto.a.10105/pdf > <DOI: 10.1002/cyto.a.10105>.</td> <td>1-7, 11-23, 27-32 --- 8-10, 24-26</td> </tr> <tr> <td>Y</td> <td>TOFALLIS, C. Least Squares Percentage Regression. Journal of Modern Applied Statistical Methods , Vol 7 , No. 2 , pp. 526-534 [online], 2009, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1406472>.</td> <td>5-7, 11-14, 21-23, 27-30</td> </tr> <tr> <td>P, X</td> <td>NOVO, D. et al. Generalized Unmixing Model for Multispectral Flow Cytometry Utilizing Nonsquare Compensation Matrices. Cytometry Part A, Vol. 83A, No. 5, pages 508-520 [online], March 22, 2013, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://onlinelibrary.wiley.com/doi/10.1002/cyto.a.22272/abstract > <DOI: 10.1002/cyto.a.22272>.</td> <td>1-32</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	Y --- A	US 2010/0302536 A1 (BALL, J. et al.) December 2, 2010 figures 1-4, paragraphs [0013]-[0037].	1-7, 11-23, 27-32 --- 8-10, 24-26	Y --- A	HUTH, Y. et al. Fourier Transformed Spectral Bio-imaging for Studying the Intracellular Fate of Liposomes. Cytometry Part A, Vol. 57A, No. 1, pages 10-21 [online], 2004, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://onlinelibrary.wiley.com/doi/10.1002/cyto.a.10105/pdf > <DOI: 10.1002/cyto.a.10105>.	1-7, 11-23, 27-32 --- 8-10, 24-26	Y	TOFALLIS, C. Least Squares Percentage Regression. Journal of Modern Applied Statistical Methods , Vol 7 , No. 2 , pp. 526-534 [online], 2009, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1406472 >.	5-7, 11-14, 21-23, 27-30	P, X	NOVO, D. et al. Generalized Unmixing Model for Multispectral Flow Cytometry Utilizing Nonsquare Compensation Matrices. Cytometry Part A, Vol. 83A, No. 5, pages 508-520 [online], March 22, 2013, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://onlinelibrary.wiley.com/doi/10.1002/cyto.a.22272/abstract > <DOI: 10.1002/cyto.a.22272>.	1-32
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.															
Y --- A	US 2010/0302536 A1 (BALL, J. et al.) December 2, 2010 figures 1-4, paragraphs [0013]-[0037].	1-7, 11-23, 27-32 --- 8-10, 24-26															
Y --- A	HUTH, Y. et al. Fourier Transformed Spectral Bio-imaging for Studying the Intracellular Fate of Liposomes. Cytometry Part A, Vol. 57A, No. 1, pages 10-21 [online], 2004, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://onlinelibrary.wiley.com/doi/10.1002/cyto.a.10105/pdf > <DOI: 10.1002/cyto.a.10105>.	1-7, 11-23, 27-32 --- 8-10, 24-26															
Y	TOFALLIS, C. Least Squares Percentage Regression. Journal of Modern Applied Statistical Methods , Vol 7 , No. 2 , pp. 526-534 [online], 2009, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1406472 >.	5-7, 11-14, 21-23, 27-30															
P, X	NOVO, D. et al. Generalized Unmixing Model for Multispectral Flow Cytometry Utilizing Nonsquare Compensation Matrices. Cytometry Part A, Vol. 83A, No. 5, pages 508-520 [online], March 22, 2013, [retrieved 2013-10-15]. Retrieved from the Internet <URL: http://onlinelibrary.wiley.com/doi/10.1002/cyto.a.22272/abstract > <DOI: 10.1002/cyto.a.22272>.	1-32															
<p><input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/></p>																	
<p>* Special categories of cited documents:</p> <table border="0"> <tr> <td>"A" document defining the general state of the art which is not considered to be of particular relevance</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</td> </tr> <tr> <td>"E" earlier application or patent but published on or after the international filing date</td> <td>"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</td> </tr> <tr> <td>"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)</td> <td>"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</td> </tr> <tr> <td>"O" document referring to an oral disclosure, use, exhibition or other means</td> <td>"&" document member of the same patent family</td> </tr> <tr> <td>"P" document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </table>			"A" document defining the general state of the art which is not considered to be of particular relevance	"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	"E" earlier application or patent but published on or after the international filing date	"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	"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)	"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	"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	"P" document published prior to the international filing date but later than the priority date claimed						
"A" document defining the general state of the art which is not considered to be of particular relevance	"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																
"E" earlier application or patent but published on or after the international filing date	"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																
"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)	"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																
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family																
"P" document published prior to the international filing date but later than the priority date claimed																	
<p>Date of the actual completion of the international search 15 October 2013 (15.10.2013)</p>		<p>Date of mailing of the international search report 13 NOV 2013</p>															
<p>Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201</p>		<p>Authorized officer: Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774</p>															