# High speed classification of individual bacterial cells using a model-based light scatter system and multivariate statistics

Murugesan Venkatapathi,[1,*] Bartek Rajwa,[2] Kathy Ragheb,[2] Padmapriya P. Banada,[3] Todd Lary,[4] J. Paul Robinson,[2] and E. Daniel Hirleman[1]

[1]School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907, USA
[2]Purdue University Cytometry Laboratories at Bindley Bioscience Center, Purdue University, West Lafayette, Indiana 47907, USA
[3]Molecular Food Microbiology Laboratory in the Department of Food Science, Purdue University, West Lafayette, Indiana 47907, USA
[4]Cellular Analysis Technology Center, Beckman Coulter, Inc., Miami, Florida 33196, USA

*Corresponding author: mvenkata@purdue.edu

Received 14 May 2007; revised 29 November 2007; accepted 30 November 2007; posted 4 December 2007 (Doc. ID 83027); published 7 February 2008

We describe a model-based instrument design combined with a statistical classification approach for the development and realization of high speed cell classification systems based on light scatter. In our work, angular light scatter from cells of four bacterial species of interest, *Bacillus subtilis, Escherichia coli, Listeria innocua*, and *Enterococcus faecalis*, was modeled using the discrete dipole approximation. We then optimized a scattering detector array design subject to some hardware constraints, configured the instrument, and gathered experimental data from the relevant bacterial cells. Using these models and experiments, it is shown that optimization using a nominal bacteria model (i.e., using a representative size and refractive index) is insufficient for classification of most bacteria in realistic applications. Hence the computational predictions were constituted in the form of scattering-data-vector distributions that accounted for expected variability in the physical properties between individual bacteria within the four species. After the detectors were optimized using the numerical results, they were used to measure scatter from both the known control samples and unknown bacterial cells. A multivariate statistical method based on a support vector machine (SVM) was used to classify the bacteria species based on light scatter signatures. In our final instrument, we realized correct classification of *B. subtilis* in the presence of *E. coli, L. innocua*, and *E. faecalis* using SVM at 99.1%, 99.6%, and 98.5%, respectively, in the optimal detector array configuration. For comparison, the corresponding values for another set of angles were only 69.9%, 71.7%, and 70.2% using SVM, and more importantly, this improved performance is consistent with classification predictions. 

*OCIS codes:* 170.1530, 120.5820, 290.5850.

## 1. Introduction

The primary diagnostic used to classify cells in traditional flow cytometry is fluorescence. Molecular tags or labels are designed such that they have distinct fluorescence spectra and attach only to specific subpopulations of cells. Cells passing through the sensing zone are illuminated by the interrogation beam and elastically scatter some of the incident light. The fluorescent tags also absorb incident optical energy, some of which is subsequently emitted as fluorescence. The spectral signature of the fluorescence is designed such that it can be differentiated from the scattered light and also from fluorescence deriving from tags that attach to other subpopulations. The fluorescence is detected and provides a signal whereby each subpopulation can be counted or separated from the total population.

0003-6935/08/050678-09$15.00/0

Scattered light has also been used to a degree in cytometry, though historically in a subservient role to fluorescence. Since all cells scatter light when they are in the incident beam, a scattered light detector signals the presence of a particle or cell in the optical sensing zone. Further, it has often been assumed that scattering is monotonically dependent on particle or cell size, and for that reason scattering has also been used as an indication of the variability of sizes in the cell population [1]. Most commercially available flow cytometers sense both forward (FALS) and right-angle (RALS) light scatter intensities. However, as we detail in this paper, the correlation of scatter intensity with respect to the size of the cells is actually quite strongly dependent on the detector geometry and projected solid angle as well as on the shape and refractive index of the cells. A similar observation involving scatter measurements of polystyrene spheres in a flow cytometer was made previously [2] and also shown to be expected according to Lorenz–Mie theory [3].

Since the angular variation of light scatter from particles and cells depends on their size, shape, optical properties, and internal structure, scattering can in principle be used to good advantage in classification. Also, interrogation beams used in light scatter are of low enough power that there is negligible alteration of the biochemical composition of the cells, i.e., scattering diagnostics can be truly noninvasive. Light scatter from individual cells in liquid suspensions as a function of angle has been studied before [4,5] and has been found to be sensitive to differences in cell properties. As a general rule, near-forward scatter (light scattered within a few degrees of the incident beam direction) exhibits strong dependence on the size and shape of the particle projection into the beam as well as on the effective refractive index. Scatter at larger angles (e.g., side and back scatter) offers imprints of the internal nonhomogeneities and surface characteristics of the particles. In larger mammalian cells ($>10$ mm) the nuclei and mitochondria are larger than optical wavelengths, and hence scattering from these intracellular structures becomes important even at forward angles [5], though that is not the case for the smaller bacterial cells used in this study.

To effectively use a correlation of light scatter with the morphological properties of cells, more than one forward- or side-scatter measurement is required in single particle instruments like flow cytometers. In theory, the angular scattering signature from an unknown particle can be used for biological analysis and classification by applying two different approaches. One could apply inverse scattering methods (that are not always mathematically tractable) to infer information on the size, shape, and optical properties of the cell, which in turn could be translated into biological parameters. Alternatively, a simpler black box approach using feature recognition and clustering algorithms alone can in principle distinguish and classify cells based on the scatter from known biological groups. To realize either of these approaches, it will generally be necessary to measure the scatter from each cell with a high resolution over a large angular span, resulting in a large number of scatter measurements [6,7]. Unfortunately, this is not simple to achieve in traditional flow cytometers because of interference with fluorescence detectors (which typically measure at angles 70°–100° (measured from the laser beam propagation vector) and the high speed of operation required to analyze a significant number of cells in biological applications. Also, in cell sorting applications, each cell has to be classified in a few microseconds, which severely limits the amount of scatter information that can be collected and processed.

The need for high speed, especially important for classification of rare cells, limits the number of angles at which the scattered light signature can be sampled. In other words, oversampling incurs a large penalty, and it is necessary to position the detection apertures at optimal or near-optimal locations so that the maximum amount of information can be extracted from the minimum number of scattered light detectors. Thus it is crucial to use model-based design to select viable detector angles if the number of scatter detectors must be limited. This is especially true for biological samples that not only have a large variation in size and refractive indices within each phenotype or species but may also have contamination.

To demonstrate the effectiveness of our approach, we have optimized detector configurations using simulated scatter measurements at a few angles and then applied multivariate statistical methods to classify actual bacteria in corresponding experiments. We also show a strong variation in classification rates for different sets of angles. The analytical scatter models are used to study scatter from the bacterial cells only in the forward angles ($<45°$ from the transmitted laser beam), allowing us to model the cells as a homogeneous material with an effective refractive index representative of the cells. One of the important conclusions of this study is that the variations of the physical properties of the bacteria have to be included for effective optimization of high speed scatter measurements. Typical models of scatter from bacteria use a representative size and refractive index, and this approach is insufficient for many cases, as shown in this work. This is because even small changes in size and refractive indices can result in order of magnitude changes in the scatter. In addition to the changes in physical properties of bacteria considered here, it should be noted that instrument noise and small variations in the orientation of bacteria in a flow cytometer manifest as minor sources of noise in scatter measurements of a bacteria species.

This paper has three sections; Section 2 describes the experimental techniques and the system used for multiangle scatter measurement of single bacterial cells, Section 3 describes the scatter models and optimization of detector configurations for bacteria, and Section 4 discusses the multivariate statistical meth-

ods applied and the results of the bacteria classification experiments.

## 2. Experimental Methods

All the analyses were run on a Cytomics FC500 flow cytometer (Beckman-Coulter, USA) equipped with a 488 nm air-cooled argon laser. The particles and cells are forced through square microchannels (250 μm sides) one by one across the laser beam using hydrodynamic forces between the sheath fluid and the fluid containing the samples (Fig. 1). The cytometer was retrofitted with an enhanced scatter detection system (Beckman-Coulter, USA) that measures scatter at four distinct forward angles and replaces the traditional forward scatter detector. This scatter measurement system (Fig. 1) consists of four ring detectors and an axial light loss (extinction cross section) detector in a single assembly. The assembly can be moved along the laser beam axis to change the angles of measurement. The four angles of measurement in each experiment cannot be chosen independently because the four rings in the detector are fixed with respect to each other. Uniformly spaced optical fibers in each ring (12 to 34 per ring) are used to direct the light scattered into each ring to a separate avalanche photodiode. The scattering signals from each ring detector were preamplified and then amplified to achieve 10 bit resolution. The CXP software (Beckman-Coulter) was used to acquire the data on the flow cytometer. WinMDI (freeware from Joe Trotter) was used to display and analyze the flow cytometry data.

Four different nonpathogenic bacterial cultures from species of importance in biosafety were selected because they span a critical range of size and shape. *Escherichia coli* K12, *Listeria innocua* F4248, and *Bacillus subtilis* ATCC 6633 are rod shaped bacteria, and *Enterococcus faecalis* CG110 appears as cocci (round-shaped) in chains. The cultures, obtained from the laboratory of A. Bhunia in the Food Science Department at Purdue University, were grown in brain heart infusion (BHI) broth for 16–18 h at 37 °C, 140 rpm in a shaker incubator. The cultures were washed (once by centrifuging at 3000 g) and resuspended in sterile phosphate buffered saline (PBS—pH 7.6) before analysis.

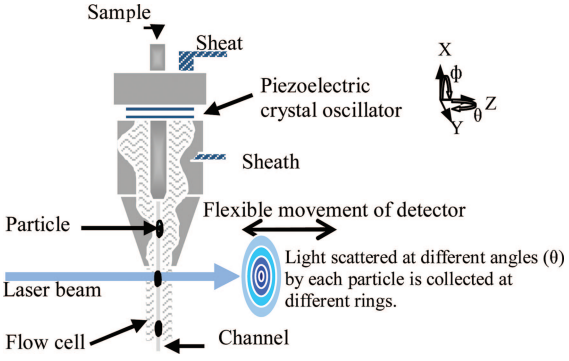


Fig. 1. (Color online) Schematic of the flow cytometer system used in experiments.

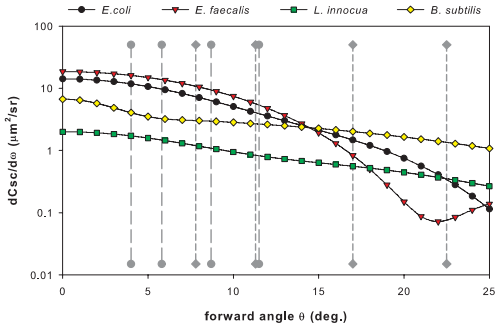


Fig. 2. (Color online) Forward scattering signatures for representative bacteria of the four species of interest, plotted as average differential scattering cross section (dCsc/dω averaged over φ) versus forward scattering angle θ. The vertical lines indicate the center angles for the four ring apertures in the detector array. The two sets of four lines correspond to the two different measurement configurations, A (diamond ends) and B (circular ends).

## 3. Scatter Models

The dimensions of the laser beam used (Gaussian distributions of electric field amplitude, dimensions ~20 μm × 120 μm) were smaller than the width of the microchannel (250 μm square cross section), and hence the beam was not distorted by the edges of the channel as in the case of cylindrical microchannels [8]. The analytical scatter models assume the particles are in isolation in the sheath fluid, and the angular scatter distribution is calculated and integrated over the area of the forward scatter detectors. Since the cells are much smaller than the incident laser beam, an incident plane wave is assumed [9]. Also the cells are modeled as homogeneous particles with effective refractive indices using the discrete dipole approximation (DDA), and results based on this assumption are valid for forward angles as discussed above. The DDA was first formulated in 1973

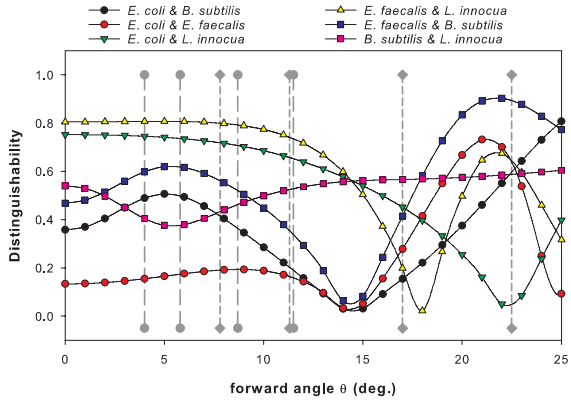


Fig. 3. (Color online) Variation of distinguishability with forward angle θ (averaged over all azimuth angles *f*) for representative bacteria of the four species. The two sets of four vertical lines correspond to the two different measurement configurations, A (diamond ends) and B (circular ends).

[10] and later extended to applications in astrophysics [11,12]. In this approach, the particle is divided into volume elements (dipoles) with dimensions of one-tenth of a wavelengths or less. Since the dimension of dipole volumes are much less than the wavelength, each such volume of the particle can be represented by a single dipole with suitable properties. The polarizations of these dipoles, due to both the incident laser beam and the other dipoles in the scattering volume, are then calculated using coupled matrix equations. The resulting scattered field from the whole particle is then calculated as the sum of the fields created by the polarizations of each dipole element in the scattering volume. A numerical algorithm developed by the group of one of the authors [13–15] using this DDA approach was used to model the light scatter from the bacteria species.

Light scatter from four different bacteria species, *E. coli, L. innocua, B. subtilis*, and *E. faecalis* was studied. *L. innocua, E. coli*, and *B. subtilis* are rod-shaped bacteria; *L. innocua* measures ~2 μm in length and ~0.6 μm in width [16]. *E. coli* is typically 2 μm in length and about 1 μm in diameter [17]. *E. faecalis* is a spheroid-shaped bacteria or cocci in chains, with each coccus measuring about 1.38 μm and 1.29 μm [18]. *B. subtilis* forms long rods with oval-shaped endospores, and the typical length of the vegetative cell may vary from 2 to 8 μm depending on the growth media used. A phase contrast light microscope (Leica Microsystems USA, Bannockburn, IL) was used to measure the average dimensions of the vegetative cell as 4.3 μm by 0.54 μm . Typically, the volume of the bacterial cells changes by as much as 4% during sporulation, while the refractive index increases [19] from 1.39 to 1.51. The spores that are approximately half of the size of the vegetative cell are left intact if the cell lyses. Similarly, cell death results in a decrease of the relative refractive index to a minimum [20]. The analytical model of the bacterial cells in this work assumes a nominal effective refractive index of 1.394 as has been observed previously [17]. The angular variation of scatter was also corrected for refraction of the scattered partial waves across the flow cell on the way to the detectors. Also,

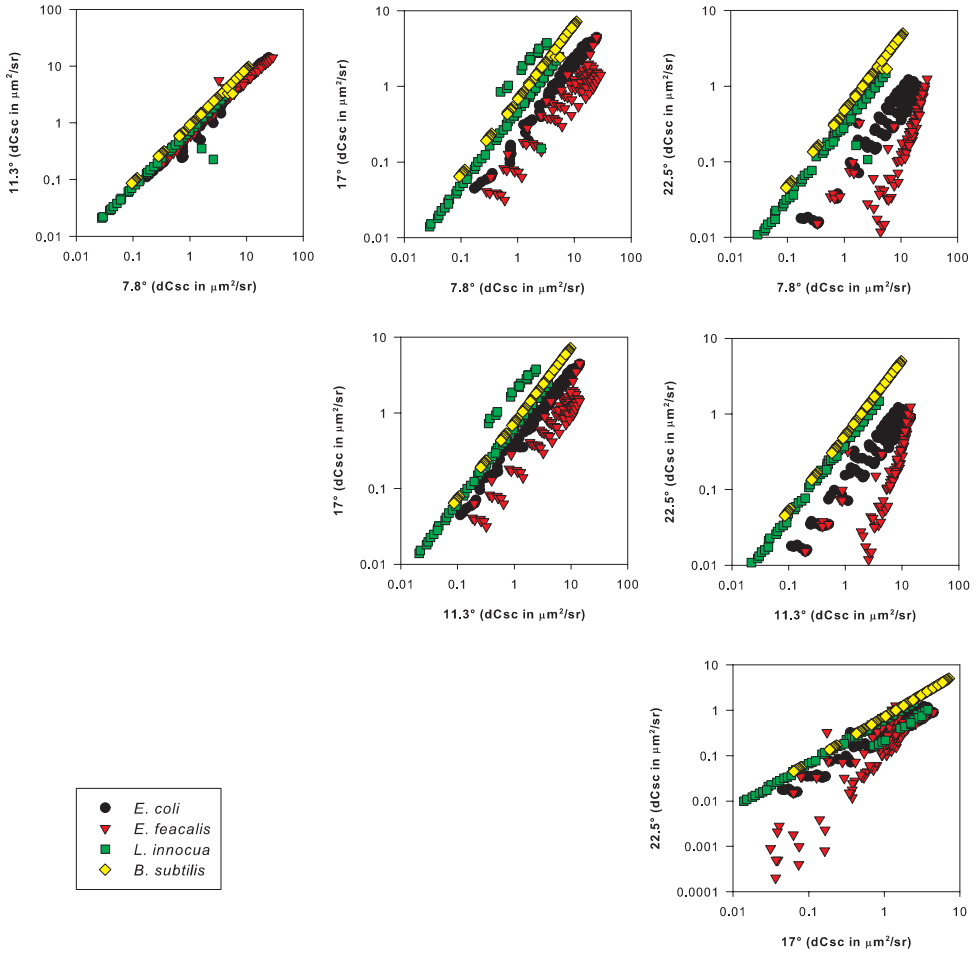


Fig. 4. (Color online) Predicted scatter plots of ring-averaged forward scattering intensities (average dCsc/dω over *f*) for detector configuration A for four bacterial species (note: size of subgroups are not to scale). Each data point represents the predicted signal pair for bacteria from a population distribution governed by Eq. (2) with the following: refractive index mean of 1.394 and standard deviation of 2% and a normal volume distribution with standard deviation of 5%.

the longer axes of the bacteria are assumed to be approximately aligned with the axis of flow due to the hydrodynamic forces in a flow cytometer [17]. Nevertheless, it is to be noted that the scatter measurement is insensitive to small rotations of the particles in the *XY* plane because of the circular geometry of detectors. The variation of refractive indices and sizes results in dispersion of each bacterial species population in the light scatter measurement space, and hence the scatter data must eventually be processed using multivariate statistical methods. But in the first step of the instrument design phase, we assume representative physical properties for each bacterial species and endeavor to locate the few available scatter detectors at optimal locations for effective classification. The computational results for angle-resolved scatter (averaged over azimuthal angles) of the four bacteria species are shown in Fig. 2. The differential scattering cross section ($I_{\text{scat}}r^2/I_o$) is independent of the distance between the detector and the sample, and is used to choose the angles for maximum discrimination of the bacteria. As can be seen from Figs. 2 and 3, the angles represented by vertical dashed lines with diamond ends (configuration A: nominal 7.8°, 11.3°, 17°, and 22.5°) are more appropriate, while the angles highlighted by vertical dashed lines with circular ends (configuration B: nominal 4°, 5.8°, 8.7°, and 11.5°) are not appropriate for distinguishing between all the four types of bacteria.

For the purposes of supporting the design phase of instrument development it is important to have a simple model for predicting the performance of various design configurations. To do so, we define distinguishability at a particular scattering angle or for a particular detector as the absolute value of the ratio of the difference in scattering cross sections to the sum of scattering cross sections for a pair of representative bacteria (see Fig. 4). A distinguishability close to 1.0 indicates that the expected scattering signal difference between representative bacteria from two species is approximately equal to the sum of the two signals, so differentiation is maximum, and a value of 0 indicates there is no difference at this angle.

For an array of detectors, the distinguishabilities for each detector can be summed to provide a simple distinguishability characteristic $D$ of the array and of representative bacteria from a pair of bacterial species or subspecies, as given by

$$D = \sum_{4\text{angles}} \left| \left(\text{dCsc}^i - \text{dCsc}^j\right) / \left(\text{dCsc}^i + \text{dCsc}^j\right) \right|, \quad (1)$$

where $i, j$ represent different bacteria.

This detector-array-summed distinguishability parameter provides a semi-quantitative indicator of the expected performance of different detector array constellations in differentiating between paired species of bacteria and therefore can be used in instrument design. While predicted distinguishability allows us to compare different detector geometries, it does not predict classification rates between pairs of bacteria. Table 1 shows distinguishability as defined by Eq. (1) for the six bacteria pairs for detector configurations A and B, respectively.

This model (Table 1) predicts that in changing from instrument design A to B there will be an improvement in classification between two of the bacteria species pairs (*L. innoccua* ↔ *E. coli* and *L. innoccua* ↔ *E. faecalis*) and a reduction in the ability to differentiate between the other four species pairs. Though it is true in general that the classification rates in experiments (Table 2) are greater in configuration A as predicted here, detailed comparisons with experiments would show that the above predictions are inconsistent except in the case of *B. subtilis*. In fact, the prediction of increased distinguishability of *L. innocua* in configuration B contradicts the experimental results. In one situation (differentiating between *E. coli* and *E. faecalis*) the distinguishability is significantly low (0.67) in configuration B, indicating that any noise in the scatter measurements will make differentiation of the two species relatively very dif-

**Table 1. Performance Predictions for Advanced Cytometers based on Multiangle Scatter Detector Configurations A and B[a]**

| | Distinguishability $D$ Computed from Eq. (1)[b] | | Classification Rates Predicted Using Full Model with Bacteria Population Distributions from Eq. (2)[c] | |
|---|---|---|---|---|
| Bacteria Species Pairs | A | B | A | B |
| *E. faecalis* versus *B. subtilis* | 2.25 | 2.01 | 100% | 87% |
| *E. faecalis* versus *L. innocua* | **2.42** | **3.12** | **95%** | **79%** |
| *E. faecalis* versus *E. coli* | 1.34 | 0.67 | 77% | 61% |
| *E. coli* versus *B. subtilis* | 2.15 | 1.49 | 100% | 81% |
| *E. coli* versus *L. innocua* | **1.88** | **2.82** | **95%** | **79%** |
| *B. subtilis* versus *L. innocua* | 2.11 | 1.79 | 98% | 74% |

[a]Two cases where the simple distinguishability predictions and the classification rate predicted by the full simulation differ in trend are in bold.

[b]Assuming a single representative bacteria size and effective refractive index.

[c]$n_o = 1.394$, $\sigma_n = 2\%$, $\sigma_v = 5\%$ (rounded to nearest integer %).

Table 2. Classification Rates of Bacteria for Multiangle Light Scatter Detector Configurations A and B

| Bacteria Species Pairs | Predicted Classification Rates based on Full Model with Bacteria Population Distributions from Eq. (2)[a] | | Measured Classification Rates[b] | |
|---|---|---|---|---|
| | A | B | A | B |
| *E. faecalis* versus *B. subtilis* | 100% | 87% | 98.5% | 70.2% |
| *E. faecalis* versus *L. innocua* | 95% | 79% | 81.6% | 77.0% |
| *E. faecalis* versus *E. coli* | 77% | 61% | 68.7% | 57.9% |
| *E. coli* versus *B. subtilis* | 100% | 81% | 99.1% | 69.9% |
| *E. coli* versus *L. innocua* | 95% | 79% | 86.3% | 74.8% |
| *B. subtilis* versus *L. innocua* | 98% | 74% | 99.6% | 71.7% |

[a] $n_o = 1.394$, $\sigma_n = 2\%$, $\sigma_v = 5\%$. Percentages rounded to nearest integer.
[b] Average uncertainties are $\pm 0.5\%$.

ficult. This distinguishability estimate ignores the variation of scatter within each species due to intraspecies differences in size and refractive indices, nor does it consider instrument noise. Even the apparent validity of this simple model for *B. subtilis* is in fact a mirage. For example, this simple model (Fig. 3) would indicate that a single detector at 7.8° should be much less useful than one at 22.5° to differentiate

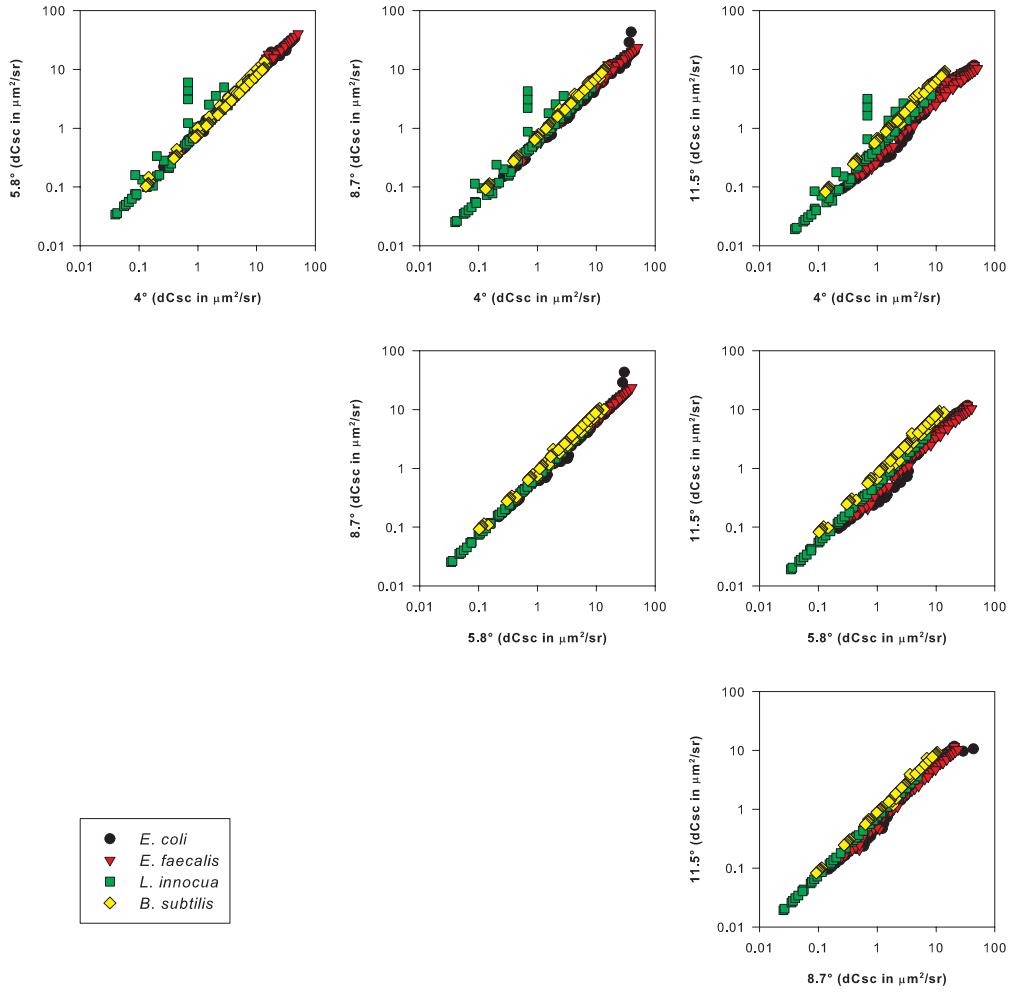


Fig. 5. (Color online) Predicted scatter plots of ring-averaged forward scattering intensities (average dCsc/dω over *f*) for detector configuration B for four bacterial species (note: sizes of subgroups are not to scale). Each data point represents the predicted signal pair for bacteria from a population distribution governed by Eq. (2) with the following: refractive index mean of 1.394 and standard deviation of 2% and a normal volume distribution with standard deviation of 5%.

*B. subtilis* from *E. faecalis*, but reality is different [Fig. 6(b)]. We find that for typical intraspecies variations in physical properties of biological particles in realistic classification applications, the simple model is not useful in predicting relative performance of various detectors. For that reason, we are convinced that a much more complete model that includes the variation of scattering signatures with intraspecies variations in physical properties must be used.

For a model to be useful in instrument design it must simulate all the variations in scattering signatures as well as simulate the classification process itself. In order to do so, the variance of scatter at different angles due to the variation in refractive index and size has to be considered. Hence a normal distribution of size and refractive index of these bacteria species was used in addition to their nominal sizes and refractive indices to make predictions of classification rates. The standard deviation of the normal distribution in refractive index was 2% with a mean ($\mu$) of 1.394. Similarly the standard deviation in volume of the bacteria was assumed to be 5% to account for variation in size of the vegetative cells. This variation in volume was modeled by corresponding isotropic changes in the dimensions of the bacteria. This normal population distribution of each bacteria species was divided into 91 subgroups (13 divisions along refractive index and 7 in volume varying from $\mu - 3s$ to $\mu + 3s$) and was mapped on to the four dimensional space of the two measurement configurations using DDA models. Using these transformed populations in this measurement space, an estimate of the classification rate was made. While the two dimensional sections of the four dimensional measurement space (Figs. 4 and 5) can show a qualitative difference between the two measurement scenarios, the estimated classification rates (Table 1) comprise the set of variables that could be used as optimization parameters in an instrument design process.

The population of bacteria in each grid of the input parameter space was weighted to its clone in the scatter measurement space, providing us a population density for each subgroup. The size of each subgroup (a four dimensional hypercuboid) in the measurement space varies, and it was determined by its average distance to the neighboring subgroups (adjacent population groups in the space of refractive index and volume). This then allows us to use the set of all scatter measurement simulation points that fall within the overlapping (hyper) volumes to define a conservative estimate of the misclassification rate. This model does not take the instrument noise into account, but uses a conservative gating system and hence was expected to give a reasonable estimate of classification rates. Once scatter intensities of each subgroup are computed using DDA, they are weighted by the population density ($w$) using

$$w_{ij} = \iint_{\text{subgroup}} \frac{1}{\sigma_n\sqrt{2\pi}} \exp\left[\frac{-(n_i - n_o)^2}{2{\sigma_n}^2}\right] \times \frac{1}{\sigma_v\sqrt{2\pi}} \exp\left[\frac{-(V_j - V_o)^2}{2{\sigma_v}^2}\right], \quad (2)$$

where $n_i$, $V_j$, are the refractive index and volume of the subgroup $ij$. $n_o$, $\sigma_n$ are the mean and standard deviation of the refractive indices; $V_o$, $\sigma_v$ are the mean and standard deviation of the volume of bacteria.

#### 4. Classification Methods and Experimental Results

The measured scatter at four angles in the form of multidimensional data from each of the four known bacteria was partitioned into training sets and the unknown cells to be classified using a random data partitioning algorithm. The classification problem was to determine the rate of correct classification of the cells partitioned as unknown particles, by using the training sets. Figure 6 shows the distribution (500 data points in each species) of measured scatter intensities from *B. subtilis* and *E. faecalis*. The traditional forward and side scatter measurements shown in Fig. 6(a) are useless in classification of these bacteria, while scatter measurements at 7.8° and 22.5° in Fig. 6(b) show better separation, as can be inferred from the numerical scatter model (Fig. 3). Nevertheless, it is amply clear that these two bacteria species cannot be resolved in this two dimensional parametric space, and data have to be classified using all the four dimensional data simultaneously. Hence the SVM algorithm has been employed to classify these data sets [21]. The fact that the classification rates for these two bacteria at this set of angles are as high as 98.5% is not obvious from Figs. 6(b) and 6(c). Linear discriminant analysis (LDA) was also used to classify these data sets, and the classification results were significantly lower (70% for the above case) due to LDA limitations in classifying the bacteria groups that form complex patterns in this multidimensional data space.

Table 2 shows the final classification rates between the four bacteria using measurement configurations A and B in experiments. Typically when using such a statistical classification method, accuracy does not have to be symmetric, which means classification success for bacteria A and bacteria B from a mixture AB can indeed be different. But in the following cases, the results were almost symmetrical (within the limits of uncertainty), and hence the average classification rates between any two bacteria are tabulated.

The presented uncertainty of classification (0.5%) is the average variation of classification rates among multiple sets of bacteria, each set containing scatter measurements of approximately 5000 bacteria in each species. It was observed that these results reflect the predictions of the full scatter model, and hence highlight the potential of these models in optimizing scatter measurements of bacteria for high speed classification. Once the scatter measurement system is optimized, prior measurements of known bacteria can be used to

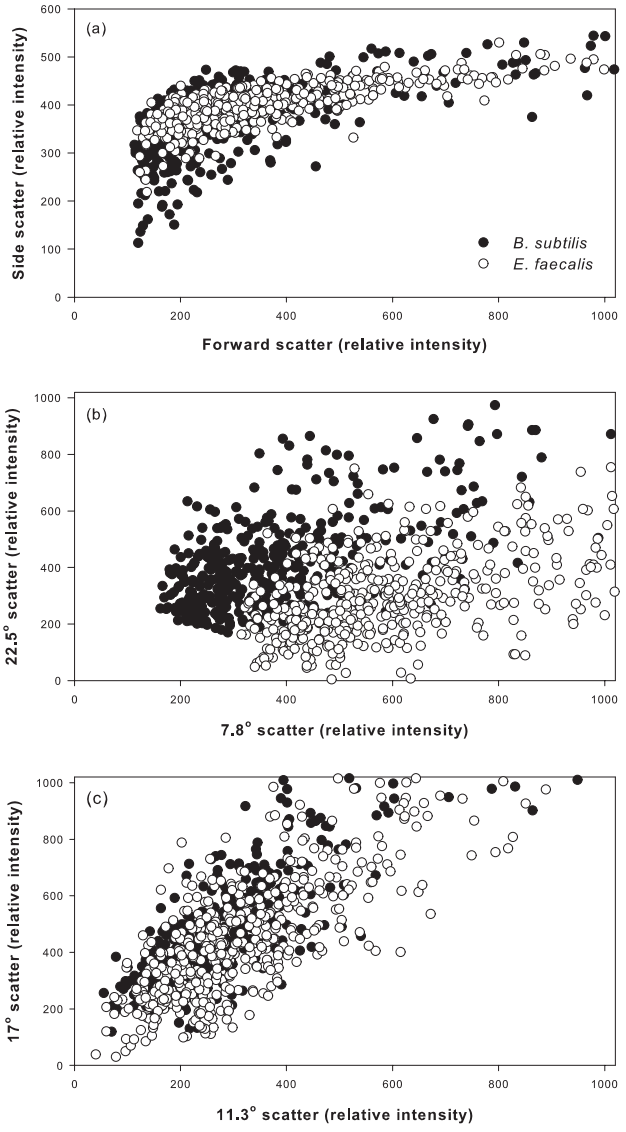


Fig. 6. (Color online) Scatter measurements of *B. subtilis* and *E. faecalis* (forward scatter was integrated from 7° to 23° and side scatter integrated from 80° to 100° as used in conventional flow cytometry).

classify the bacteria of an unknown sample with the least possible uncertainty. To remove the necessity of prior experimental data in such a classification system, more accurate scatter models of bacteria might be needed. This requires a more accurate distribution of size and refractive index in a species (can be dynamic or time varying, depending on the sample of interest) and also gain and noise functions of the instrument.

## 5. Conclusions

Light scattering models are indispensable for designing detector configurations for flow cytometers that minimize the number of detectors and maximize the performance for given classification problems. We have proven that this model-based design approach, in combination with multivariate statistical methods for classifying the data vectors obtained from the optimal detector arrays, enables label-free (scattering only) classification of many important biological samples in modified flow cytometers. The effectiveness of the approach was demonstrated by optimizing an advanced multiangle (four-ring) forward scatter flow cytometer for classification of mixtures of *E. coli, L. innocua, B. subtilis*, and *E. faecalis*. With the new instrument we demonstrated classification rates ranging from a low of 68.7% (*E. faecalis* ↔ *E. coli*) up to 99.6% (*L. innoccua* ↔ *B. subtilis*). Our work also conclusively established that scatter models that include intraspecies variation of physical properties are essential in optimizing detector geometries in most cases, classification of *L. innocua* in the presence of *E. coli* or *E. faecalis*, for example. Finally, the results show that if the gain functions, noise characteristics of the instrument, and the distribution of the relevant morphological properties (refractive index and size) of the bacteria are all known adequately, the numerical models can be used to replace prior experimental data for development of the classification system, in addition to their use for optimizing the scatter measurement configuration.

This research was supported through a cooperative agreement with the Agricultural Research Service of the U.S. Department of Agriculture project number 1935-42000-035 and the Center for Food Safety and Engineering at Purdue University.

## References

1. G. C. Salzman, S. B. Singham, R. G. Johnston, and C. F. Bohren, "Light scattering and cytometry," in *Flow Cytometry and Sorting,* 2nd ed. (Wiley, 1990).
2. A. G. Hoekstra, R. M. P. Doornbos, K. E. I. Deurloo, H. J. Noordmans, B. G. de Grooth, and P. M. A. Sloot, "Another face of Lorenz—Mie scattering: monodisperse distributions of spheres produce Lissajous-like patterns," Appl. Opt. **33,** 494–500 (1994).
3. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley Interscience, 1983).
4. P. J. Wyatt, "Cell wall thickness, size distribution, refractive index ratio and dry weight content of living bacteria (*Staphylococcus aureus*)," Nature **226,** 277–279 (1970).
5. J. R. Mourant, J. P. Freyer, A. H. Hielscher, A. A. Eick, D. Shen, and T. M. Johnson, "Mechanisms of light scattering from biological cells relevant to noninvasive optical-tissue diagnostics," Appl. Opt. **37,** 3586–3593 (1998).
6. B. J. Price, V. H. Kollman, and G. C. Salzman, "Light-scatter analysis of microalgae. Correlation of scatter patterns from pure and mixed asynchronous cultures," Biophys. J. **22,** 29–36 (1978).
7. A. V. Chernyshev, V. I. Prots, A. A. Doroshkin, and V. P. Maltsev, "Particle classification from light scattering with the scanning flow cytometer," Cytometry **37,** 215–220 (1999).
8. M. Venkatapathi, G. Gregori, K. Ragheb, J. P. Robinson, and E. D. Hirleman, "Measurement and analysis of angle resolved scatter from small particles in a cylindrical microchannel," Appl. Opt. **45,** 2222–2232 (2006).
9. J. A. Lock, "Interpretation of extinction in Gaussian-beam scattering," J. Opt. Soc. Am. A **12,** 929–938 (1995).
10. E. M. Purcell and C. R. Pennypacker, "Scattering and absorp-

tion of light by nonspherical dielectric grains," Astrophys. J. **186,** 705–714 (1973).
11. B. T. Draine, "The discrete-dipole approximation and its application to interstellar graphite grains," Astrophys. J. **333,** 848–872 (1988).
12. M. A. Taubenblatt and T. K. Tran, "Calculation of light scattering from particles and structure by the coupled-dipole method," J. Opt. Soc. Am. A **10,** 912–919 (1993).
13. R. Schmehl, B. M. Nebeker, and E. D. Hirleman, "Discrete-dipole approximation for scattering by features on surfaces by means of a two-dimensional fast Fourier transform technique," J. Opt. Soc. Am. A **14,** 3026–3036 (1997).
14. B. M. Nebeker, G. W. Starr, and E. D. Hirleman, "Evaluation of iteration methods used when modeling scattering from features on surfaces using the discrete-dipole approximation," J. Quant. Spectrosc. Radiat. Transf. **60,** 493–500 (1998).
15. B. M. Nebeker, J. L. de la Pena, and E. D. Hirleman, "Comparisons of the discrete-dipole approximation to modified double interaction model methods to predict light scattering from small features on surfaces," J. Quant. Spectrosc. Radiat. Transf. **70,** 749–759 (2001).
16. A. Gupta, D. Akin, and R. Bashir, "Detection of bacterial cells and antibodies using surface micromachined thin silicon cantilever resonators," J. Vac. Sci. Technol. B **22,** 2785–2791 (2004).
17. A. N. Shvalov, J. T. Soini, I. V. Surovtsev, G. V. Kochneva, G. F. Sivolobova, A. K. Petrov, and V. P. Maltsev, "Individual *Escherichia coli* cells studied from light scattering with the scanning flow cytometer," Cytometry **41,** 41–45 (2000).
18. C. Signoretto, M. del Mar Lleò, M. C. Tafi, and P. Canepari, "Cell wall chemical composition of *Enterococcus faecalis* in the viable but nonculturable state," Appl. Environ. Microbiol. **66,** 1953–1959 (2000).
19. A. Katz, A. Alimova, M. Xu, P. Gottlieb, E. Rudolph, J. C. Steiner, and R. R. Alfano, "*In situ* determination of refractive index and size of *Bacillus* spores by light transmission," Opt. Lett. **30,** 589–591 (2005).
20. A. P. Kononenko, K. I. Kononenko, and D. M. Mikhov, "Dependence of refractive index on physiological state of microbial population," J. Appl. Spectros. **11,** 795–797 (1969).
21. V. Vapnik, *Statistical Learning Theory* (Wiley, 1998).