

**University of Alberta**

Computational Analysis of Wide-Angle Light Scattering from Single Cells

by

Patrick Michael Pilarski

A thesis submitted to the Faculty of Graduate Studies and Research  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Electrical and Computer Engineering

© Patrick Michael Pilarski  
Fall 2009  
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

## **Examining Committee**

Christopher J. Backhouse, Electrical and Computer Engineering

Petr Musilek, Electrical and Computer Engineering

Marek Reformat, Electrical and Computer Engineering

Bruce Cockburn, Electrical and Computer Engineering

Walter F. Bischof, Computing Science

Aaron Wheeler, Chemistry, University of Toronto

*Dedicated to my parents,  
Eugene and Linda Pilarski,  
foundations and skyhooks,  
and to my partner Nicole Pagan,  
the sky itself.*

# Abstract

The analysis of wide-angle cellular light scattering patterns is a challenging problem. Small changes to the organization, orientation, shape, and optical properties of scatterers and scattering populations can significantly alter their complex two-dimensional scattering signatures. Because of this, it is difficult to find methods that can identify medically relevant cellular properties while remaining robust to experimental noise and sample-to-sample differences. It is an important problem. Recent work has shown that changes to the internal structure of cells—specifically, the distribution and aggregation of organelles—can indicate the progression of a number of common disorders, ranging from cancer to neurodegenerative disease, and can also predict a patient’s response to treatments like chemotherapy. However, there is no direct analytical solution to the inverse wide-angle cellular light scattering problem, and available simulation and interpretation methods either rely on restrictive cell models, or are too computationally demanding for routine use.

This dissertation addresses these challenges from a computational vantage point. First, it explores the theoretical limits and optical basis for wide-angle scattering pattern analysis. The result is a rapid new simulation method to generate realistic organelle scattering patterns without the need for computationally challenging or restrictive routines. Pattern analysis, image segmentation, machine learning, and iterative pattern classification methods are then used to identify novel relationships between wide-angle scattering patterns and the distribution of organelles (in this case mitochondria) within a cell. Importantly, this work shows that by parameterizing a scattering image it is possible to extract vital information about cell structure while remaining robust to changes in organelle concentration, effective size, and random placement. The result is a powerful collection of methods to simulate and interpret experimental light scattering signatures. This gives new insight into the theoretical basis for wide-angle cellular light scattering, and facilitates advances in real-time patient care, cell structure prediction, and cell morphology research.

# Preface

The work contained in this dissertation was carried out under the supervision of Dr. Christopher J. Backhouse between September 2004 and August 2009. All work was performed as a part of the Applied Miniaturisation Laboratory (AML) in the Electrical and Computer Engineering Department at the University of Alberta, Canada. This dissertation is in accordance with the “paper format” regulations of the Faculty of Graduate Studies and Research, University of Alberta, and is based on the content of the following published first-author papers, conference proceedings, and papers in preparation or submitted. For these papers, I acted as lead researcher—designing the experiments/software in collaboration with other authors, writing and evaluating the software, conducting the experiments, analyzing the data, and writing the majority of the published work. Previous publication and submission information is listed on the first page of each chapter in accordance with the guidelines put forth by the Faculty of Graduate Studies and Research; chapters recast this cited work. Bibliographic formatting follows the style guide of the Optical Society of America (OSA).

## Refereed Journal and Conference Publications

1. **P. M. Pilarski** and C. J. Backhouse, “A method for cytometric image parameterization,” *Optics Express* **14**(26), 12720–12743 (2006).
2. **P. M. Pilarski**, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Rapid simulation of wide-angle scattering from mitochondria in single cells,” *Optics Express* **16**(17), 12819–12834 (2008).
3. **P.M. Pilarski**, X.T. Su, D.M. Glerum, and C.J. Backhouse, “Computational analysis of mitochondrial placement and aggregation effects on wide-angle cell scattering patterns,” *Proceedings of SPIE* **7187**, 71870J, 12 pages (2009).
4. **P. M. Pilarski**, X. T. Su, D. M. Glerum, and C. J. Backhouse, “Robust pattern analysis techniques for the interpretation of wide-angle light scattering images from single cells,” *Journal of Biomedical Optics*, 15 pages (submitted Aug. 10, 2009).
5. **P. M. Pilarski** and C. J. Backhouse, “Theoretical foundations for robust wide-angle scattering pattern analysis,” *Optics Express*, 5 pages (submitted Aug. 12, 2009).
6. **P. M. Pilarski** and C. J. Backhouse, “Reverse monte carlo analysis applied to wide-angle cellular light scattering,” in preparation (2009).

Tools and methods developed during the course of researching this dissertation were found to be applicable to a number of other related biomedical and computer imaging problems. This resulted in a number of collaborative publications, conference presentations, and two patents pending. These are not included in the text of this dissertation, but a list of selected contributions is included here for further reading by interested parties (for conferences, the presenting author is indicated with an asterisk):

### **Refereed Journal and Conference Publications**

1. V. Sieben, C. Debes-Marun, **P. Pilarski**, G. Kaigala, L. Pilarski, and C. Backhouse, "FISH and chips: chromosomal analysis on microfluidic platforms," *IET Nanobiotechnology* **1**(3), 27–35 (2007).
2. T. Mirzayans, N. Parimi, **P. Pilarski**, C. Backhouse, L. Wyard-Scott, and P. Musilek, "A swarm-based system for object recognition," *Neural Network World* **15**(3), 243–255 (2005).

### **Refereed Conference Presentations and Published Abstracts**

1. **P.M. Pilarski\***, X.T. Su, D.M. Glerum, and C.J. Backhouse, "Disease characterization using wide-angle light scattering patterns," in proceedings of the *iCORE Alberta Electrical and Computer Engineering Graduate Research Symposium* (University of Alberta, Edmonton, 22 June 2009, poster and abstract).
2. **P. M. Pilarski\***, V. J. Sieben, C. D. Marun, and C. J. Backhouse, "Computer vision for fish screening in myeloma," in *XIth International Myeloma Workshop* (Kos, 2007, poster and abstract) and *Haematologica–The Hematology Journal* **92**(6), 108–108 (2007).
3. V.J. Sieben\*, C.D. Marun, **P.M. Pilarski**, G.V. Kaigala, L.M. Pilarski, and C.J. Backhouse, "Microchips for optimized fish screening in myeloma", in *XIth International Myeloma Workshop* (Kos 2007, oral and abstract) and *Haematologica–The Hematology Journal* **92**(6), 64-65, (2007).
4. V.J. Sieben\*, C.S. Debes-Marun, **P.M. Pilarski**, G.V. Kaigala, L.M. Pilarski, and C.J. Backhouse, "FISH AND CHIPS: Chromosomal analysis utilizing microfluidic platforms," in *Nanotechnology in Biomedicine (Keystone)* (Tahoe City, California, USA, 2007, poster).

5. C. S. Debes Marun\*, V. Sieben, **P. M. Pilarski**, T. Reiman, A. R. Belch, and L.M. Pilarski, "FISH and chips: novel point of care technology to detect chromosomal abnormalities," *Blood* **108**, 971a, (2006).
6. **P. M. Pilarski\***, V. J. Sieben, C. Debes Marun, and C. J. Backhouse, "An artificial intelligence system for detecting abnormal chromosomes in malignant lymphocytes", in *Canadian Society for Immunology, Annual Conference*, p. 126 (Halifax, Nova Scotia, 2006, oral, abstract, and poster).

### **Patents Pending**

1. "Novel methods for cellular image analysis", C.J. Backhouse, **P.M. Pilarski**, and X.T. Su; provisional patent filed December 21, 2006; patent filed December 24, 2007 (PCT/CA2007/002321).
2. "Automated FISH analysis, circulating microfluidic chip, and method for immobilizing cells to a microfluidic chip", L.M. Pilarski, C.J. Backhouse, V.J. Sieben, C. Debus Marun, **P.M. Pilarski**, G.V. Kaigala; provisional patent filed September 22, 2006; patent filed September 17, 2007 (PCT/CA2007/001641).

# Acknowledgements

First and foremost, thanks to my supervisor Christopher Backhouse. Over the past ten years, Chris opened my eyes to countless facets of engineering and science; I will carry the things I learned from him forward in all my future research endeavours. Most importantly, Chris taught me how to cultivate vision, and how to ground it in a clear, precise reality. For this I cannot thank him enough.

To my committee members along the way, my appreciation for steering me in the correct direction, your supportive words, and your excellent ideas. Appreciation as well to all my labmates, fellow students, other ECEGSA council members, departmental staff and faculty, and collaborators past and present who played key roles in both this research and my graduate career. In particular, my thanks to Govind Kaigala for his calm and always astute guidance, and to my cubicle mate, co-conspirator, and character foil Vincent Sieben. I have learned more from Vince than I can relate, in all aspects of life, and his friendship is one of the things I will remember most about my time in graduate studies.

This dissertation would not have been possible without three people. My parents Linda and Eugene Pilarski handed me the board and let me play my own game. They are my mentors, two of my best friends, and they have never stopped supporting me and helping me grow along the way; this dissertation is a ripple from the stones they placed. I will always remember our globes of lightning and captured pond creatures. To my partner Nicole Pagan: your unbounded love made this achievable; you are the brightest of lights, my beacon, and the source of all possibility. To the rest of my family, my gratitude for unwavering support and for patiently listening to me talk about my research—no small feat, I know.

Finally, thanks to all the funding agencies that supported me along the way. This work was made possible by financial support from the Natural Sciences and Engineering Research Council (NSERC) and the Canadian Institute for Photonic Innovations (CIPI), studentships/scholarships from NSERC, SPIE, Alberta Ingenuity, the Informatics Circle of Research Excellence (iCORE), and a University of Alberta Dissertation Fellowship.

*huge blue  
the road  
disappears in sky*

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Social Impact and Medical Relevance . . . . .	1
1.2	Project Description and Scope . . . . .	3
1.2.1	Publication and Presentation of Results . . . . .	4
1.3	Overview of the Remaining Chapters . . . . .	5
	References . . . . .	5
<b>2</b>	<b>Background</b>	<b>10</b>
2.1	Background on Related Areas of Study . . . . .	10
2.1.1	Cellular Light Scattering Theory . . . . .	10
2.1.2	Fixed-Angle and Wide-Angle Cytometry . . . . .	11
2.1.3	Optical Simulation Methods . . . . .	12
2.1.4	Biomedical Image Analysis . . . . .	14
2.2	Previous Work in Cellular Scattering Pattern Analysis . . . . .	15
2.2.1	The Inverse Scattering Problem . . . . .	15
2.2.2	Direct Approaches . . . . .	16
2.2.3	Parametric Methods . . . . .	18
2.2.4	Iterative Methods . . . . .	20
2.3	Key Components of the Problem . . . . .	23
2.3.1	Experimental and Simulated Data . . . . .	23
2.3.2	Feature Extraction . . . . .	24
2.3.3	Linking Methods . . . . .	26

2.4	Summary . . . . .	27
	References . . . . .	27
<b>3</b>	<b>Rapid Simulation of Wide-Angle Scattering from Mitochondria in Single Cells</b>	<b>33</b>
3.1	Introduction . . . . .	33
3.2	Background . . . . .	34
3.3	Methods . . . . .	35
3.3.1	A Theoretical Approach . . . . .	35
3.3.2	A Basis in X-ray Diffraction Theory . . . . .	37
3.3.3	Scattering Examples . . . . .	39
3.3.4	The mtPatterns Algorithm . . . . .	40
3.3.5	Isotropic Scattering from Mitochondria in the Mie Regime . . . . .	42
3.3.6	Test Images . . . . .	43
3.4	Results and Discussion . . . . .	44
3.4.1	Comparison of mtPatterns and Experimental Cytometry Images . . . . .	44
3.4.2	Extension to Scattering from Larger Spherical Mitochondria . . . . .	47
3.4.3	Extension to Scattering from Non-Spherical Mitochondria . . . . .	50
3.4.4	Observations on the Effect of Scatter Distribution and Interaction . . . . .	50
3.5	Conclusions . . . . .	51
	References . . . . .	52
<b>4</b>	<b>Theoretical Foundations for Robust Wide-Angle Scattering Pattern Analysis</b>	<b>56</b>
4.1	Introduction . . . . .	56
4.2	Background . . . . .	57
4.3	Comparison of Corrected and Uncorrected Patterns . . . . .	58

4.4	Discussion of Robust Behaviour . . . . .	60
4.5	Conclusions . . . . .	61
	References . . . . .	62
<b>5</b>	<b>Pattern Analysis Techniques for Wide-Angle Light Scattering Images</b>	<b>63</b>
5.1	Introduction . . . . .	63
5.2	Background . . . . .	64
5.2.1	Clinical Relevance and Cell Morphology . . . . .	66
5.3	Methods . . . . .	68
5.3.1	Simulation Setup . . . . .	68
5.3.2	Image Analysis . . . . .	69
5.3.3	Attribute Selection . . . . .	70
5.3.4	Classification . . . . .	70
5.3.5	Extensions to the <code>mtPatterns</code> Algorithm . . . . .	71
5.4	Results and Discussion . . . . .	71
5.4.1	Classification of Distributions . . . . .	77
5.5	Conclusions . . . . .	80
	References . . . . .	81
<b>6</b>	<b>A Method for Cytometric Image Parameterization</b>	<b>84</b>
6.1	Introduction . . . . .	84
6.2	Light Scattering Through Complex Cellular Bodies . . . . .	85
6.3	Computational Approaches to the Inverse Problem . . . . .	87
6.3.1	Recent Segmentation Work . . . . .	89
6.3.2	Computational Challenges . . . . .	91

6.4	The Computational Pipeline . . . . .	92
6.4.1	Feature Detection . . . . .	93
6.4.2	Feature Clustering . . . . .	95
6.4.3	Post-Processing . . . . .	97
6.4.4	Parameterization . . . . .	98
6.5	Analysis Methods . . . . .	100
6.6	Results . . . . .	103
6.6.1	Qualitative Assessment . . . . .	103
6.6.2	Quantitative Assessment . . . . .	106
6.7	Discussion . . . . .	109
6.7.1	Remarks on Feature Detection . . . . .	109
6.7.2	Remarks on Clustering . . . . .	110
6.7.3	Remarks on Post-Processing and Parameterization . . . . .	111
6.7.4	Remarks on Image Size Reduction . . . . .	112
6.7.5	Remarks on Versatility . . . . .	112
6.8	Conclusions . . . . .	113
	References . . . . .	114
<b>7</b>	<b>Reverse Monte Carlo Analysis for Wide-Angle Cellular Scattering</b>	<b>119</b>
7.1	Introduction . . . . .	119
7.2	Background . . . . .	120
7.3	RMC for Cellular Light Scattering . . . . .	122
7.3.1	Mutation Methods . . . . .	122
7.3.2	Acceptance of Solutions . . . . .	124
7.3.3	Fitness Metrics . . . . .	124

7.3.4	Test Cases . . . . .	125
7.4	Results . . . . .	125
7.5	Conclusions . . . . .	131
	References . . . . .	132
<b>8</b>	<b>Conclusions</b>	<b>134</b>
8.1	Impact and Contributions . . . . .	135
8.2	Future Directions . . . . .	136
8.2.1	Applied Parameterization for Pattern Decomposition . . . . .	137
8.2.2	Advanced Iterative Structure Prediction . . . . .	137
	References . . . . .	140

# List of Tables

3.1	The mtPatterns algorithm . . . . .	40
5.1	List of texture analysis attributes generated during image processing . . . . .	70
5.2	Attribute selection results using 10-fold cross-validation . . . . .	72
5.3	Attribute ranking results . . . . .	73
5.4	Classification accuracy using 10-fold cross-validation . . . . .	80
5.5	Example confusion matrices for two classifiers (best four attributes) . . . . .	80
6.1	The set of useful image parameters ( <b>P</b> ) . . . . .	98
6.2	Statistical analysis for band intensity parameters . . . . .	107
6.3	Statistical analysis for band width parameters . . . . .	107
6.4	Statistical analysis for band number/spacing parameters . . . . .	108
7.1	The RMC algorithm for scattering prediction . . . . .	122

# List of Figures

2.1	Schematic diagram of a wide-angle cytometer . . . . .	12
2.2	A graphic representation of the cellular inverse scattering problem . . . . .	15
2.3	Direct solution to the inverse scattering problem . . . . .	16
2.4	Parametric solution to the inverse scattering problem . . . . .	18
2.5	Iterative solution to the inverse scattering problem . . . . .	21
3.1	Example of how changes to spatial distribution impact the spacing of features within two-dimensional scattering patterns . . . . .	39
3.2	Diagram of the mtPatterns scattering simulation process . . . . .	41
3.3	Two angular point-spread functions $M(\theta)$ for individual scatterers . . . . .	44
3.4	Comparison of feature spacing in experimental and simulated data . . . . .	45
3.5	A series of images generated using the mtPatterns algorithm . . . . .	48
4.1	The effect of phase perturbation on the scattering pattern of a population . . . . .	59
4.2	Comparison of scattering patterns in terms of standard image descriptors . . . . .	60
5.1	Schematic of four different organelle distributions . . . . .	67
5.2	Qualitative comparison of scattering from different distributions . . . . .	74
5.3	Relationship between distribution radius and pattern complexity . . . . .	75
5.4	Relationship between distribution radius and texture energy . . . . .	76
5.5	Histograms showing S5, S5x2, W5, and E5 Law texture energy . . . . .	77
5.6	Histograms showing Law S5x4 and Haralick H2 values . . . . .	78
5.7	Histograms showing E5 and S5 Law values . . . . .	79

6.1	Schematic diagram of a wide-angle cytometer . . . . .	85
6.2	Example images containing features present in experimental and simulated wide-angle scattering patterns . . . . .	86
6.3	Frame from an animated movie of the Cythe processing pipeline . . . . .	94
6.4	Agent fixation in the Cythe pipeline . . . . .	95
6.5	An example of horizontal bridge removal . . . . .	96
6.6	Two parts of a single <code>propagate_id()</code> cycle for an active agent . . . . .	96
6.7	A visual comparison of Cythe region detection with a test image . . . . .	104
6.8	A visual example of horizontal bridge removal . . . . .	105
6.9	Extraction of a band hierarchy for noisy and noise-free images . . . . .	106
7.1	Comparison of search parameter performance for TBRMC . . . . .	126
7.2	Comparison of search parameter performance for PBRMC . . . . .	127
7.3	Comparison of scattering from peripheral target and candidate populations .	128
7.4	Comparison of a peripheral target population $P_t$ containing ten organelles with the candidate population $P_c$ using TBRMC . . . . .	129
7.5	Comparison of a peripheral target population $P_t$ containing ten organelles with the candidate population $P_c$ using PBRMC . . . . .	130

# List of Symbols and Abbreviations

2D	Two-dimensional
3D	Three-dimensional
Å	Angström
$A(\mathbf{S})$	Scattering intensity in the reciprocal (Fourier) domain
$A(\mathbf{s})$	Scattering amplitude on the receptive field
ANN	Artificial neural network (also multi-layer perceptron)
BS	Band spacing
CBIR	Content-based image retrieval
CCD	Charge-coupled device (or, the receptive part of a digital camera)
cRMC	Cellular Reverse Monte Carlo search
cv10	Ten-fold cross-validation
$d$	Distance from the scattering distribution center to the receptive field
$d_n^o$	Distance from a scattering point to the incident light source
$d_o$	Target distribution during Reverse Monte Carlo iteration
DM	Dynamic mutation rate in cRMC
DJ	Dynamic replacement (jump) rate in cRMC
EC	Evolutionary computing
F(IM)	Function to translate an image into a set of physical parameters
FB	Flashback frequency in cRMC
FDTD	Finite difference time domain
FISH	Fluorescent in situ hybridization
FLSI	Flying light scattering indicatrix
FT	Fourier transform
FFT	Fast Fourier transform
FM	Free movement rate in cRMC
$I(\mathbf{S})$	Scattering intensity in the reciprocal (Fourier) domain

$IM_t$	True scattering plot for a target population $P_t$
LOC	Lab-on-a-chip
LSS	Light scattering spectroscopy
MLP	Multi-layer perceptron
MM	Multiple myeloma
MMC	Metropolis Monte Carlo search
$mt$	Mitochondria
$\#mt$	Number of mitochondria in a distribution
$P_t$	Target (scatterer) distribution, density, or population
$P_c$	Candidate (scatterer) distribution, density, or population
PBRMC	Direct pixel-based Reverse Monte Carlo fitness comparison
$px$	Pixels
R-G	Rayleigh-Gans
RBC	Red blood cell
RMC	Reverse Monte Carlo search
RMSE	Root mean squared error
$r_n$	Scattering point source 'n'
$S$	Vector to a point in reciprocal (Fourier) domain
$s$	Vector to a point on the receptive field
$s'$	Vector to a point on the receptive field
$s_0$	Direction of incident light (unit vector)
$S(P)$	True scattering function
$S'(P)$	Approximate/inferred scattering function
$S^{-1}(IM)$	Inverse scattering function using image IM
SSE	Sum squared error
SM	Static mutation rate in cRMC
SJ	Static replacement (jump) rate in cRMC
$sr$	Steradian

TBRMC	Texture-based Reverse Monte Carlo fitness comparison
WM	Waldenström's macroglobulinemia
XRD	X-ray diffraction
$\beta$	Noise removal threshold for the Cythe algorithm
$\chi^2$	Statistical comparison metric
$\delta$	Region joining threshold for the Cythe algorithm
$\lambda$	Wavelength of light
$\phi$	Angle of observation perpendicular to the axis of incident light
$\rho(\mathbf{r})$	Distribution of scattering points
$\tau(\mathbf{r})$	Phase shift distribution for scattering points
$\theta$	Angle of observation along axis of incident light

# Chapter 1

## Introduction

Recent studies have demonstrated robust methods to record the light scattering signatures of small cell-like particles and their component macromolecules<sup>1-6</sup>, bacteria<sup>7</sup>, and whole human cells or populations of cells<sup>2,4,8-16</sup>. These light scattering patterns are rich with information about a cell's internal structure<sup>8-12,15,17-22</sup>. However, the link between scattering intensity and cell structure is still only defined for limited number of situations, and is the subject of continuing international research.

The issue is an important one. When coupled with a knowledge of disease-related changes to cellular morphology, light scattering information can be used to detect the presence, absence, or progression of a number of treatable and/or mitigable human diseases<sup>11,23</sup>. By knowing how to properly assess light scattering patterns, it is possible to infer important information about human illnesses.

This dissertation explores the link between scattering patterns and cellular structure. To do so, it presents a set of theoretical relationships, algorithmic tools and methods to simulate, interpret, and categorize information contained in complex biomedical signatures, such as those generated by laser light scattering through human cells.

### 1.1 Social Impact and Medical Relevance

Medical technology is on the cusp of a miniaturization revolution<sup>24</sup>. Emerging *lab-on-a-chip* (LOC) devices allow complex medical tests and genetic assays to be performed for a fraction of the cost of traditional methods, and in a fraction of the time<sup>24</sup>. This presents the potential for high-throughput patient screening, with reduced clinic wait times and greater accessibility to novel testing methods<sup>2,11,19,24</sup>.

One of the most promising LOC candidates for inexpensive and noninvasive diagnostics is the miniaturized wide-angle cytometer<sup>2</sup>. Wide-angle cytometers allow the rapid acquisition

of detailed two-dimensional light scattering patterns from whole cells, building on current limited-angle clinical cytometry techniques to further investigate the light-scattering properties of cells. This increase in available scattering information leads to an increase in diagnostic potential—wide-angle LOC light scattering has been shown to contain detailed information about a biological cell’s internal micro- and nano-structures<sup>1,2,19</sup>.

Once known, cell structure can be directly related to disease. Changes to intracellular makeup—for instance, the number of metabolically related organelles (mitochondria) in a cell, along with their size, function, and distribution, have been shown to indicate the progression of a number of ailments, including cancer and neurodegenerative disease<sup>11,25–28</sup>. In addition, it has been shown that the arrangement of mitochondria within a cell can indicate a patient’s response to treatments such as chemotherapy, and help evaluate the severity of their disease<sup>29,30</sup>.

With the proliferation of inexpensive imaging technology and more robust microfluidic platforms, portable wide-angle cytometry has become a very real clinical option<sup>2,11,19</sup>. This could bring complicated and expensive medical tests to users that would otherwise not have access to the necessary facilities—*e.g.* cancer screening in remote rural health-care centers, disease testing for workers in isolated environments (such as the international space-station or deep-sea research platforms), or on-the-spot neurodegenerative testing in assisted living environments.

The impact of accessible wide-angle cytometry systems extends outside the clinic. Once it can be interpreted, the information contained in two-dimensional scattering signatures provides a powerful tool to explore the micro- and nano-structural components of living cells, without the need for costly (and in many cases diffraction-limited) imaging systems. This opens the door for detailed exploratory research into human metabolic mechanisms, cell function and behaviour in the presence of diseases and/or drugs, automated pathogen screening, and customized pharmaceutical design.

However, there is still an analytical bottleneck—there are few systems to automatically interpret massive quantities of biomedical imaging data<sup>31</sup>, and none applicable for use with a wide-angle cytometry system. With an increase in the complexity and information density of biomedical signatures, it is becoming increasingly intractable for human experts to manage the diagnostic task without computational assistance<sup>31</sup>. This is especially true for visually complex two-dimensional scattering signatures of wide-angle cytometers; despite a number of recent advances, the problem of relating scattering patterns to medically relevant cellular structure is still open and largely unsolved.

## 1.2 Project Description and Scope

This dissertation describes a new set of theoretical relationships, methods, and tools for interpreting the medically relevant aspects of wide-angle scattering patterns, such as those captured by a wide-angle cytometer. To do so, it encompasses new research in light scattering simulation, biomedical pattern analysis and image processing, and iterative methods for cell-structure determination. In short, it presents a novel framework for associating wide-angle scattering patterns with aspects of biological cell structure—specifically, the arrangement, concentration, and morphology of mitochondria, organelles known to have a direct relationship to a number of human illnesses<sup>25</sup>.

There are no direct numerical methods capable of relating a scattering signature to a full geometric representation of cellular structure<sup>18,32</sup>. Thus, to identify relationships between wide-angle scattering patterns and cellular structure, it is essential to develop techniques and tools able to extract meaningful associations from available empirical evidence.

The state-of-the-art in cellular scattering pattern analysis is limited to light scattering information collected at a set of fixed angles, or a one-dimensional angular slice<sup>18</sup>. As such, there are few available tools to begin tackling the problem of wide-angle scattering pattern analysis. To date, experimental wide-angle scattering data is scarce—there are a limited number of devices currently under development world-wide, including our device at the University of Alberta<sup>2</sup>, and large clinical datasets (on the order of thousands of scattering images) have not yet been created.

The difficulty is two-fold. An understanding of wide-angle patterns and the data contained within them is critical for steering the development of the emerging optical devices. However, without already having access to fully developed cytometry devices, it is challenging to get to the heart of the relationship between a two-dimensional scattering pattern and cellular structure. To further complicate the problem, it is also challenging to obtain a large database of simulated patterns. Numerical simulation methods are currently limited by either restrictive geometries, or prohibitive computational cost<sup>19</sup>.

Previous work in cellular scattering analysis has mainly focused on the interpretation of one-dimensional scattering data, which does not address the image processing and pattern analysis aspects of the two-dimensional form of the problem. Two-dimensional patterns are extremely complex in terms of their visual composition, and also in the way this composition relates to a scatterer's structure. While this complexity leads to a corresponding wealth of available predictive data, it requires special new tools to extract relevant concepts in the presence of distracting noise, non-unique scattering profiles, and redundant informa-

tion. While powerful, analysis methods from other fields—*e.g.* X-ray diffraction theory, machine learning algorithms, adaptive classifiers, and computer vision technology—have not yet been adapted for use in the domain of two-dimensional cellular scattering.

As such, this dissertation presents new simulation tools, biomedical image analysis methods, and pattern interpretation methods for use in understanding wide-angle cellular light scattering data. It also explores the fundamental predictive limitations of wide-angle cytometry, and describes a number of clinically relevant relationships between light scattering and cell structure. Taken as a whole, this opens up two-dimensional scattering signatures as a valuable resource for clinical and academic cellular investigation, and paves the way for similar research in related domains where complex multi-dimensional biomedical data must be related to relevant physical phenomena.

### **1.2.1 Publication and Presentation of Results**

Core work from this project resulted in a number of first-author papers in leading biomedical optics journals and conferences, detailing new ways to interpret, analyze, and simulate the light scattering signatures of human cells<sup>33–37</sup>. These papers form a basis for the following chapters. This work and the related experiments also lead to a patent pending on new methods to interpret wide-angle cytometry data for medical use.

The methods and tools developed as part of this dissertation project were also found to be highly transferrable to other biomedical image analysis problems. Most notably, new scattering image segmentation methods were extended and successfully applied to the automated interpretation of chromosomal probe images for multiple myeloma (MM) cancer testing—*i.e.* a fully automated LOC implementation for fluorescent in situ hybridization (FISH)<sup>38–41</sup>. This new FISH method and the related image analysis software is protected under a patent pending.

During the course of this project, additional collaborative work was performed using the developed tools to assist in image-based data processing for microfluidic FISH implementations<sup>38</sup>, and in temperature control validation for LOC genetic analysis systems. Collaborative work was also done on a novel swarm-based image segmentation system<sup>42</sup>, and on data analysis and modeling for patient-related MM and Waldenström’s Macroglobulinemia (WM) studies<sup>43–45</sup>. Work performed during the course of this project also resulted in a first-author publication on adaptable microvalving control and automation for LOC devices<sup>46</sup>.

### 1.3 Overview of the Remaining Chapters

Chapter 2 presents background material on fields relating to this dissertation, and also summarizes key aspects of previous work. This is followed by a breakdown of the core problem into its component parts. Chs. 3 and 4 describe a rapid method to simulate light scattering from human cells. This is tied to a discussion about the fundamental limits and constraints of a wide-angle scattering pattern analysis—*i.e.* what kinds of structural information can and cannot be observed using current imaging methods.

Using this simulation method as a base, Ch. 5 presents a novel parametric approach for interpreting scattering patterns; this is done through a large-scale simulation study, paired with methods for biomedical image analysis, attribute selection, and pattern classification. Ch. 6 follows with an alternate parameterization scheme and image segmentation method that extends the amount of information available for extraction using wide-angle light scattering signatures.

Linking the new simulation method of Chs. 3 and 4 with the parameterization approach of Chs. 5 and 6, Ch. 7 explores an iterative method able to specifically examine the internal structure of a cellular scattering distribution. Ch. 8 closes this dissertation with a set of concluding remarks, a look at future directions, and a summary of social and academic contributions made by this work.

In summary, this dissertation presents a new simulation method and validates it against experimental and theoretical results. This simulation method and the corresponding validation approaches are then put to use in an image analysis context, leading to a set of new pattern interpretation techniques for use in rapid light-scattering-based medical diagnostics.

### References

- [1] K. Singh, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, “Analysis of cellular structure by light scattering measurements in a new cytometer design based on a liquid-core waveguide,” *IEE Proc.-Nanobiotechnology* **151**(1), 10–16 (2004).
- [2] K. Singh, X. Su, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, “A miniaturised wide-angle 2-D cytometer,” *Cytometry A* **69A**, 307–315 (2006).
- [3] Z. Wang, J. El-Ali, M. Englund, T. Gotsaed, I. R. Perch-Nielsen, K. B. Mogensen, D. Snakenborg, J. P. Kutter, and A. Wolff, “Measurements of scattered light on a microchip flow cytometer with integrated polymer based optical elements,” *Lab on a Chip* **4**(4), 372–377 (2004).
- [4] D. Watson, N. Hagen, J. Diver, P. Marchand, and M. Chachisvilis, “Elastic light scattering

- from single cells: orientational dynamics in optical trap,” *Biophysical Journal* **87**(2), 1298–1306 (2004).
- [5] H. B. Steen, “Flow cytometer for measurement of the light scattering of viral and other sub-microscopic particles,” *Cytometry Part A* **57A**(2), 94–99 (2004).
- [6] N. Kasai and M. Kakudo, *X-Ray Diffraction by Macromolecules*, Springer series in chemical physics, 80 0172-6218 (Springer, New York, 2005).
- [7] A. Katz, A. Alimova, M. Xu, E. Rudolph, M. K. Shah, H. E. Savage, R. B. Rosen, S. A. McCormick, and R. R. Alfano, “Bacteria size determination by elastic light scattering,” *IEEE Journal of Selected Topics in Quantum Electronics* **9**(2), 277–287 (2003).
- [8] K. Sem’yanov and V. P. Maltsev, “Analysis of sub-micron spherical particles using scanning flow cytometry,” *Particle & Particle Systems Characterization* **17**(5-6), 225–229 (2000).
- [9] P. L. Gourley, “Biocavity laser for high-speed cell and tumour biology,” *Journal of Physics D: Applied Physics* **36**(14), R228–R239 (2003).
- [10] P. L. Gourley, J. K. Hendricks, A. E. McDonald, R. G. Copeland, K. E. Barrett, C. R. Gourley, K. K. Singh, and R. K. Naviaux, “Mitochondrial correlation microscopy and nanolaser spectroscopy - new tools for biophotonic detection of cancer in single cells,” *Technology in Cancer Research & Treatment* **4**(6), 585–592 (2005).
- [11] P. L. Gourley, J. K. Hendricks, A. E. McDonald, R. G. Copeland, K. E. Barrett, C. R. Gourley, and R. K. Naviaux, “Ultrafast nanolaser flow device for detecting cancer in single cells,” *Biomedical Microdevices* **7**(4), 331–339 (2005).
- [12] P. L. Gourley and R. K. Naviaux, “Optical phenotyping of human mitochondria in a biocavity laser,” *IEEE Journal of Selected Topics in Quantum Electronics* **11**(4), 818–826 (2005).
- [13] J. Kruger, K. Singh, A. O’Neill, C. Jackson, A. Morrison, and P. O’Brien, “Development of a microfluidic device for fluorescence activated cell sorting,” *Journal of Micromechanics and Microengineering* **12**(4), 486–494 (2002).
- [14] J. R. Mourant, T. M. Johnson, S. Carpenter, A. Guerra, T. Aida, and J. P. Freyer, “Polarized angular dependent spectroscopy of epithelial cells and epithelial cell nuclei to determine the size scale of scattering structures,” *Journal of Biomedical Optics* **7**(3), 378–387 (2002).
- [15] L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, “Observation of periodic fine structure in reflectance from biological tissue: a new technique for measuring nuclear size distribution,” *Physical Review Letters* **80**(3), 627–630 (1998).
- [16] L. Perelman, “Optical diagnostic technology based on light scattering spectroscopy for early cancer detection,” *Expert Review of Medical Devices* **3**(6), 787–803 (2006).

- [17] K. A. Sem'yanov, P. A. Tarasov, J. T. Soini, A. K. Petrov, and V. P. Maltsev, "Calibration-free method to determine the size and hemoglobin concentration of individual red blood cells from light scattering," *Applied Optics* **39**(31), 5884–5889 (2000).
- [18] V. P. Maltsev, "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [19] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, "2D light scattering patterns of mitochondria in single cells," *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [20] R. Drezek, A. Dunn, and R. Richards-Kortum, "A pulsed finite-difference time-domain (FDTD) method for calculating light scattering from biological cells over broad wavelength ranges," *Optics Express* **6**(7), 147–157 (2000).
- [21] R. Drezek, A. Dunn, and R. Richards-Kortum, "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements," *Applied Optics* **38**(16), 3651–3661 (1999).
- [22] J. D. Wilson, C. E. Bigelow, D. J. Calkins, and T. H. Foster, "Light scattering from intact cells reports oxidative-stress-induced mitochondrial swelling," *Biophysical Journal* **88**(4), 2929–2938 (2005).
- [23] J. S. Modica-Napolitano and K. Singh, "Mitochondria as targets for detection and treatment of cancer," *Expert Reviews in Molecular Medicine* **4**(9), 1–19 (2004).
- [24] V. J. Sieben, "Self-assembly on microfluidic platforms," Ph.D. thesis, University of Alberta (2009).
- [25] D. C. Wallace, "Mitochondrial diseases in man and mouse," *Science* **283**(5407), 1482–1488 (1999).
- [26] D. C. Wallace, "Diseases of the mitochondrial dna," *Annual Review of Biochemistry* **61**, 1175–1212 (1992).
- [27] M. Brandon, P. Baldi, and D. C. Wallace, "Mitochondrial mutations in cancer," *Oncogene* **25**(34), 4647–4662 (2006).
- [28] G. M. Enns, "The contribution of mitochondria to common disorders," *Molecular Genetics and Metabolism* **80**(1-2), 11–26 (2003).
- [29] S. Sikder, J. M. G. Reyes, C. S. Moon, O. Suwan-apichon, J. H. Elisseeff, and R. S. Chuck, "Noninvasive mitochondrial imaging in live cell culture," *Photochemistry and Photobiology* **81**(6), 1569–1571 (2005).
- [30] A. M. Villa and S. M. Doglia, "Mitochondria in tumor cells studied by laser scanning confocal microscopy," *Journal of Biomedical Optics* **9**(2), 385–394 (2004). URL <http://link.aip.org/link/?JBO/9/385/1>.

- [31] U. Sinha, A. Bui, R. Taira, J. Dionisio, C. Morioka, D. Johnson, and H. Kangarloo, “A review of medical imaging informatics,” *Techniques in Bioinformatics and Medical Informatics* **980**, 168–197 (2002).
- [32] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1998).
- [33] P. M. Pilarski and C. J. Backhouse, “A method for cytometric image parameterization,” *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.
- [34] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Rapid simulation of wide-angle scattering from mitochondria in single cells,” *Optics Express* **16**(17), 12,819–12,834 (2008). URL <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-17-12819>.
- [35] P. M. Pilarski and C. J. Backhouse, “Theoretical foundations for robust wide-angle scattering pattern analysis,” *Optics Express* **submitted**, 5 pages (2009).
- [36] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Computational analysis of mitochondrial placement and aggregation effects on wide-angle cell scattering patterns,” in *Proceedings of SPIE*, A. Wax and V. Backman, eds., vol. 7187, p. 71870J (12 pages) (2009). URL <http://dx.doi.org/10.1117/12.809730>.
- [37] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Robust pattern analysis techniques for the interpretation of wide-angle light scattering images from single cells,” *Journal of Biomedical Optics* **submitted**, 15 pages (2009).
- [38] V. Sieben, C. Debes-Marun, P. Pilarski, G. Kaigala, L. Pilarski, and C. Backhouse, “FISH and chips: chromosomal analysis on microfluidic platforms,” *IET Nanobiotechnology* **1**(3), 27–35 (2007).
- [39] V. J. Sieben, C. Debes-Marun, P. M. Pilarski, G. K. Kaigala, L. M. Pilarski, and C. J. Backhouse, “Microchips for optimized fish screening in myeloma,” *Haematologica—The Hematology Journal* **92**(6), 64–65 (2007).
- [40] P. M. Pilarski, V. J. Sieben, C. D. Marun, and C. J. Backhouse, “Computer vision for fish screening in myeloma,” *Haematologica—The Hematology Journal* **92**(6), 108–108 (2007).
- [41] P. Pilarski, V. J. Sieben, C. Debes Marun, and C. Backhouse, “An artificial intelligence system for detecting abnormal chromosomes in malignant lymphocytes,” in *Canadian Society for Immunology, Annual Conference*, p. 126 (Halifax, N.S., Canada, 2006).
- [42] T. Mirzayans, N. Parimi, P. Pilarski, C. Backhouse, L. Wyard-Scott, and P. Musilek, “A swarm-based system for object recognition,” *Neural Network World* **15**(3), 243–255 (2005).
- [43] S. Adamia, A. A. Reichert, H. Kuppusamy, J. Kriangkum, A. Ghosh, J. J. Hodges, P. M. Pilarski, S. P. Treon, M. J. Mant, T. Reiman, A. R. Belch, and L. M. Pilarski, “Inherited and

acquired variations in the hyaluronan synthase 1 (HAS1) gene may contribute to disease progression in multiple myeloma and Waldenstrom macroglobulinemia,” *Blood* **112**(13), 5111–5121 (2008).

- [44] S. Adamia, P. Pilarski, A. Belch, and L. Pilarski, “Genetic abnormalities in Waldenstrom’s macroglobulinemia,” *Clinical Lymphoma & Myeloma* **9**(1), 30–32 (2009). URL <http://dx.doi.org/10.3816/CLM.2009.n.006>.
- [45] L. Pilarski, E. Baigorri, M. Mant, P. Pilarski, P. Adamson, H. Zola, and A. Belch, “Multiple myeloma includes CD20+ B and plasma cells that persist in patients treated with rituximab,” *Clinical Medicine: Oncology* **2**, 275–287 (2008). URL [http://la-press.com/article.php?article\\_id=659](http://la-press.com/article.php?article_id=659).
- [46] P. Pilarski, S. Adamia, and C. J. Backhouse, “An adaptable microvalving system for on-chip polymerase chain reactions,” *Journal of Immunological Methods* **305**(1), 48–58 (2005).

# Chapter 2

## Background

Work on the interpretation of light scattering patterns falls at the boundary between a number of related but theoretically diverse disciplines. The problem includes aspects of classical optics, physical optical devices, numerical simulation methods, biomedical pattern analysis and image processing, machine learning, and data mining. While wide-angle pattern interpretation is a relatively new sub-problem, the case of scattering pattern analysis itself is widely explored and is the subject of ongoing research. This is hardly surprising—the problem is multi-facteted and in many cases application specific, and each aspect comes with its own set of challenges and areas for further study. As such, this chapter presents a concise background discussion that describes the current state of the art as it relates to this project, and lists a set of key references for further reading. It then examines the actual problem of inverse scattering pattern analysis, describes current and past work, and isolates the core components of the scattering problem.

### 2.1 Background on Related Areas of Study

This section is a short synopsis of the fields relating to this dissertation. Where possible, reference has been made to comprehensive survey papers instead of re-examining individual sources. Where necessary, the introduction to each of the following chapters will thoroughly examine the background literature pertaining to its contents.

#### 2.1.1 Cellular Light Scattering Theory

Cells are optically complex entities, containing a heterogenous mixture of scattering objects ranging in size from a few nanometers to many microns<sup>1-3</sup>. As would be expected, this optical variability leads to equally complex scattering behaviour<sup>1</sup>—*e.g.* for incident laser light, particles such as proteins and very small mitochondria will uniformly scatter light,

while large structures such as the cell nucleus will transmit light in a geometric fashion<sup>2</sup>.

For visible light, large structures, such as the nucleus and cell wall, are primarily responsible for intensity observed in the front and back scatter regions (*i.e.* in the direction of incident light)<sup>1,2,4-9</sup>. The scattering intensity contributions of these structures may appear as large intensity bands in a 2D scattering signature<sup>10</sup>, or as regularly spaced peaks in 1D scattering samples<sup>4,11</sup>. Because of this relationship, the spacing of peaks and bands in scattering patterns has been used by a number of groups to predict cellular and nuclear size<sup>4,11</sup>.

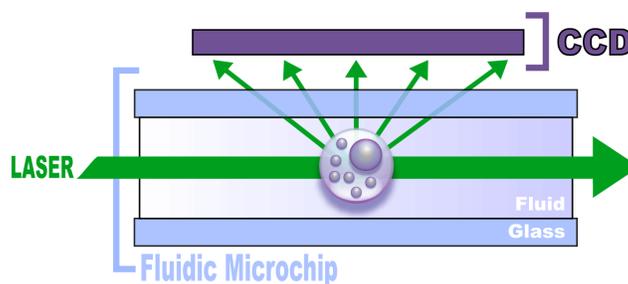
It has also been shown that changes to the arrangement of intracellular organelles, such as the mitochondria, can dramatically affect portions of recorded scattering signatures<sup>1,3,4,10-15</sup>, specifically the light scattered at large angles (*i.e.* the side scatter region, normal to the direction of incident light)<sup>1,2,12,16</sup>. Mitochondria have in fact been shown to be the dominant source of light scattering in this region, contributing up to 70–90% of recorded side-scatter intensity<sup>2</sup>. Interesting new results also indicate that mitochondria-sized scatterers may also have a significant effect on light scattering in the 5–90° range<sup>17</sup>. Melanin is also thought to be a major source of large-angle light scattering<sup>1</sup>. The main result of this is that the presence of organelles causes noticeable changes to scattering patterns beyond the effect of a simple change in the cell's effective index of refraction.

For greater detail, Gourelly *et al.*<sup>2,18</sup> and Dunn<sup>1</sup> present excellent overviews of cellular optics and the optical characteristics of intracellular components. Hecht<sup>19</sup>, Bohren and Huffman<sup>20</sup>, van de Hulst<sup>21</sup>, and Gaskill<sup>22</sup> all provide comprehensive detail on optical propagation and diffraction theory, light scattering from small particles, and Fourier optics.

### **2.1.2 Fixed-Angle and Wide-Angle Cytometry**

The analysis of cellular light scattering signatures is a crucial part of many modern medical diagnostic methods<sup>4,7,13</sup>. These include label-free methods that capture scattered light at one or more angles<sup>4,7,10,11,23</sup>, and are closely related to techniques that analyze the laser power and optical properties of cellular bodies<sup>5,18</sup> or the spectral components of reflected/refracted light<sup>4</sup>. For this work, the focus will be placed on label-free cytometers, as opposed to those that use chemical agents to generate excitation at fixed frequencies.

One of the most common label-free cytometry methods for patient cell analysis is the standard one-dimensional or fixed-angle flow cytometer<sup>4,11</sup>. These systems enjoy widespread use in medical institutions, but operate only on a small fraction of a full wide-angle scattering pattern<sup>4</sup>. In most cases, light is propagated through a cell or population of cells in sequence, and the scattered intensity is captured at two angles: perpendicular to the direc-



**Figure 2.1:** Schematic diagram of a wide-angle cytometer. It includes a fluidic channel, a laser source, and a two-dimensional charge-coupled device (CCD).

tion of incident light (side scatter) and parallel to the direction of incident light (forward scatter)<sup>4</sup>. This is used to generate representative scatter plots, where the position of points can be used to infer the number of organelles in a cell (position along the side scatter axis) and the size of the cell (position along the forward scatter axis)<sup>4</sup>. Taken as a whole, this data can then be used to determine cell health, for instance the progression of disease in MM and WM patients<sup>24</sup>.

In more detailed systems, an entire angular intensity slice (called an “indicatrix”) is captured, often using a moving detector<sup>4</sup>. The intensity features in this profile—*e.g.* the spacing and pitch of intensity peaks—can then be related to cellular properties, such as size, content, or optical density<sup>4</sup>. While some systems operate on single cells, many capture only light refracted through whole populations of cells<sup>4,11,23,25,26</sup>.

Recently, researchers have extended traditional cytometry methods to capture a wide-angle two-dimensional scattering pattern from a single living cell<sup>10,13,15,27</sup>. Through the integration of LOC technology and streamlined image-capture devices, wide-angle cytometry systems are able to measure intensity over a large scattering region, theoretically approaching a full hemisphere around the sample cell<sup>13</sup>. A graphic depiction of this system is shown in Fig. 2.1. This provides pattern recognition systems with a rich set of conceptually valuable input information, generating a corresponding increase in available diagnostic information<sup>10,13,28</sup>; wide-angle scattering patterns take the form of two-dimensional grey-scale images, and thus lend themselves to analysis with classical image processing and pattern recognition methods<sup>28</sup>.

### 2.1.3 Optical Simulation Methods

Of the wide range of optical simulation methods, the methods that directly impact this work include those based on Mie theory, the Rayleigh-Gans approximation (R-G), and Finite Difference Time Domain (FDTD) calculations<sup>3,6,13–15</sup>.

Mie scattering calculates the result of a planar incident wave that is scattered by a homogeneous sphere<sup>3,20,21</sup>. Using spherical Bessel functions, Mie theory combines incident and scattered waves to achieve a final solution describing the resulting energy patterns. It can also be extended to determine the scattering intensity for a population of spheres of varying radius. Books by van de Hulst<sup>21</sup> and Bohren and Huffman<sup>20</sup> provide mathematical derivations for a number of different Mie scattering scenarios. However, while Mie solutions give accurate (and relatively rapid) assessments of light propagation both inside and outside a scattering particle, they are inherently limited by constraints in terms of geometry and optical complexity<sup>13,26</sup>.

As such, R-G formulations have been proposed to tackle the problem of arbitrary shapes in cells. The R-G approximation assumes isotropic Rayleigh scattering by all of a scatterer's volume elements, with no interaction effects or multi-scattering by neighbouring elements. By summing the complex amplitude components from each volume element, it is possible to efficiently calculate the scattering from inhomogeneous, arbitrarily shaped particles<sup>21</sup>. As described by van de Hulst, the solution can be further simplified for specific geometric objects and distributions<sup>21</sup>. In practical use, R-G simulation approaches have been shown to accurately approximate experimental scattering from non-spherical biological objects (*e.g.* bacteria<sup>29</sup>). However, the R-G approximation is still limited to particles with refractive indices close to that of the surrounding medium, and small phase-shift values throughout the particle<sup>21</sup>. These constraints are needed for the model's assumptions to remain valid—*i.e.* they ensure that each volume element in the actual particle receives and transmits light in a way that is largely independent of all other elements.

Many aspects of cellular structure are not spherical or within narrow optical ranges, leading to limitations in the viability of Mie or R-G simulation for cellular scattering pattern analysis<sup>3,13,15</sup>. FDTD simulations provide a way to simulate optically complex structures comprised of any number of inhomogeneous, arbitrarily shaped scattering objects<sup>1,3,6,13–15</sup>. Using a discretized version of Maxwell's equations (Yee's algorithm<sup>1</sup>) and appropriate boundary constraints/transformations, FDTD methods give an extremely accurate representation of light scattering through complex media<sup>1,3</sup>. In addition, and unlike R-G, they also take into account interactions between particles and multi-scattering effects<sup>1,3</sup>. Recently, FDTD methods have been extended to 3D scattering objects, providing a new way to examine scattering from complex cells<sup>3,13</sup>. However, all FDTD methods come with a large computational cost—depending on the desired resolution and discretization level, simulations require massive amounts of memory, and could take days or weeks to run on large computational clusters (*i.e.* multi-processor supercomputers)<sup>13,15</sup>. This limits their effectiveness for large-scale data generation and real-time comparison.

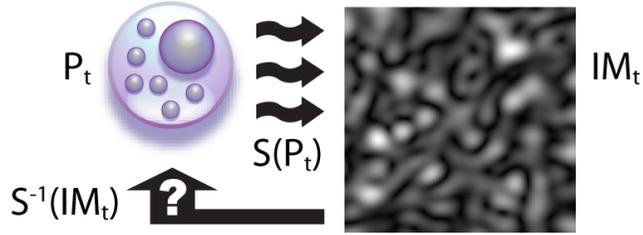
#### 2.1.4 Biomedical Image Analysis

The interpretation of wide-angle cellular light scattering patterns can also be viewed in terms of a biomedical imaging problem. Biomedical image analysis encompasses a wide range of image processing and pattern analysis techniques used for interpreting and classifying medically related image data<sup>30,31</sup>. Examples include the determination of tumor boundaries from Magnetic Resonance Imaging (MRI) or mammography data, skeletal analysis using X-ray images, and ultrasound imaging of soft tissue<sup>30</sup>. Imaging methods also include thermal analysis, light microscopy, electron microscopy, spectrography, and tomography<sup>30</sup>.

This is a challenging problem domain. Not only do the biological objects under observation have a high degree of variability, even within similar classes of sample, but observations are subject to many different forms of noise, occlusion, and data corruption due to the constraints of working with patients and the limitations of imaging hardware<sup>30</sup>. Data formats can range from optical images or 3D models that are interpretable by a human expert to complex multi-dimensional arrays that require a large degree of computational processing<sup>30</sup>.

Primary problems addressed by the field include: the preprocessing and optimization of image data; the extraction of regions of interest; the characterization of data using shape, texture, and orientation metrics; statistical image categorization; classification of samples; and the reconstruction of volumetric models from sequential data<sup>30</sup>. As such, biomedical image analysis encompasses work from computer vision, pattern analysis, and machine learning literature. It also overlaps a great deal with data mining and content-based image retrieval problems, where large bodies of data must be efficiently stored, searched, and classified<sup>31</sup>; this is described in a review by Sinha *et al.*<sup>31</sup>. As wide-angle light scattering analysis involves the collection and interpretation of vast amounts of complex image data, it is evident that techniques from other biomedical image analysis problems may transfer well to the current problem domain.

Rangayyan's book provides a detailed introduction to biomedical image analysis, and describes a number of image analysis algorithms that are potentially applicable to wide-angle light scattering<sup>30</sup>. Specific techniques and algorithms related to this dissertation are discussed at length in the following chapters. More specific information on computer vision and general scene interpretation is presented by Shapiro and Stockman<sup>32</sup>, while fundamental methods for data mining, patterns analysis, and applied machine learning can be found in the work of Duda *et al.*<sup>33</sup>, Witten and Frank<sup>34</sup>, Alpaydin<sup>35</sup>, and Engelbrecht<sup>36</sup>.



**Figure 2.2:** A graphic representation of the cellular inverse scattering problem: “given a recorded two-dimensional laser light scattering pattern ( $IM_t$ ), generated from the unknown structure  $P_t$  via some scattering process  $S(P_t)$ , what is the unknown cell structure responsible for the pattern (*i.e.* what is  $P_t$ )?”

## 2.2 Previous Work in Cellular Scattering Pattern Analysis

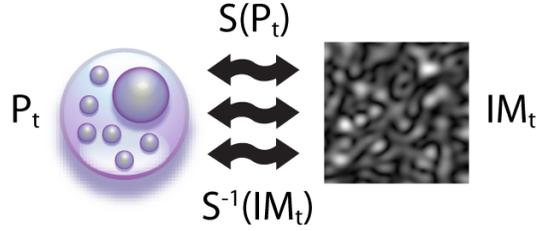
This section gives a description of the central problem in cellular light scattering: *the inverse scattering problem for light through a biological cell*. It presents an overview of the problem itself, followed by a discussion of prior art. Previous approaches are grouped into three categories, and each is explored in depth.

### 2.2.1 The Inverse Scattering Problem

The inverse scattering problem is the task of relating a scattering intensity signature to the complex 3D structure of the scattering object<sup>20</sup>. For the problem of light through a biological cell, there is no direct analytical solution<sup>3,4,25,37,38</sup>. It is not possible to simply apply a formula to an intensity pattern and receive the complete 3D geometry and/or density distribution of the scatterer<sup>37</sup>. The problem is also plagued by non-uniqueness—it is theoretically possible for more than one distribution to yield extremely similar, or in some cases identical, scattering patterns<sup>20,37</sup>. Conversely, it is possible for rotated copies of the same distribution to produce dramatically different scattering patterns<sup>3</sup>.

In its simplest form, the wide-angle cellular inverse scattering problem may be framed as follows: given a recorded two-dimensional laser light scattering pattern ( $IM_t$ ), generated from the unknown structure  $P_t$  via some scattering process  $S(P_t)$ , what is the unknown cell structure responsible for the pattern (*i.e.* what is  $P_t$ )? A graphical representation of this problem structure is shown in Fig. 2.2.

A solution to this kind of problem can take two forms: general or specific (hard<sup>20</sup>). The weaker of these two cases, the *general case*, only requires that the solution be able to produce general information relating to the scatterer’s structure (such as cell size, shape, or number



**Figure 2.3:** *The direct solution:* invert the light-scattering mechanism to recreate the detailed 3D structure of the scatterer  $P_t$ . This requires that  $IM_t$  contain enough information to perform a realistic inverse  $S^{-1}(IM_t)$ ; in experimental optics situations this is rarely the case<sup>20</sup>.

of organelles; the “parameters” of  $P_t$ ). It may neglect exact relationships between cellular components. The stronger *specific case* requires the solution to produce an accurate three-dimensional model, density distribution, or approximate geometry of the scattering body, inclusive of the relationships between scatterer components (*i.e.* fully reconstruct  $P_t$ )<sup>20</sup>.

The required strength of a solution depends heavily on the application; for medical patient sample classification, a general solution may be enough to determine the difference between healthy and sick patients, while for exploratory research it may be necessary to know the exact intracellular geometry of a sample. As such, a number of different approaches have emerged in the literature, directed toward finding general or specific solutions to the cellular inverse problem. For the purpose of this dissertation, this prior art has been divided into three major categories. Each category will be examined below with regard to advantages, disadvantages, scope, and recent progress.

## 2.2.2 Direct Approaches

**Specific Solution.** Many groups have pursued the direct inversion approach: taking an intensity image or set of scattering intensity points and, using a numerical process, attempting to derive a 2D or 3D representation of the interrogated cell’s structure<sup>4,37</sup>. This is equivalent to determining the target distribution as follows:  $P_t = S^{-1}(IM_t)$ , where  $S^{-1}(IM_t)$  is the mathematical inverse of the scattering process  $S(IM_t)$ . A graphical representation of the direct approach is shown in Fig. 2.3. While elegant and potentially powerful, a cellular scattering inversion of this kind has currently not been solved in a closed analytical form<sup>4</sup>.

One major difficulty in experimental situations is the loss of amplitude and phase information that occurs when scattered light is captured by a detection apparatus (*e.g.* a CCD camera)<sup>20</sup>. To fully reconstruct a scattering distribution from its detected energy field, it is necessary to know the phase and amplitude of all incoming waves, along with the field

inside all scattering particles<sup>20</sup>; this level of recording is rarely, if ever, achievable in practice<sup>20</sup>. In experimental (especially cytometry) situations, only intensity values are recorded. While X-ray diffraction and optical literature has shown that a direct inversion is theoretically possible by performing an Fourier transform on the scattering signature<sup>39,40</sup>, this still relies on knowing the full set of wave components at every point on the scattering field—things that are extremely challenging or impossible to obtain in practice.

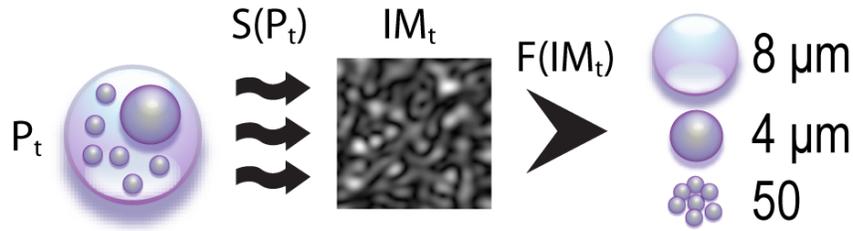
Despite the challenges, Thibault *et al.* show the use of a difference map algorithm to iteratively estimate phase parameters in a related problem domain—the extraction of reproducible 2D yeast cell images from their X-ray diffraction patterns<sup>40</sup>. However, as noted by Thibault *et al.*, there are still a host of issues preventing the use of this technique to infer the full 3D makeup of a cell<sup>40</sup>. Shapiro *et al.* also managed to perform 3D imaging of a whole cell, at resolution values approaching 10nm, using an iterative phasing algorithm; however, their optical setup also relies on X-ray diffraction microscopy as opposed to visible light<sup>41</sup>.

Most phase reconstruction algorithms in the literature focus on crystal or macromolecule reconstruction, as opposed to biological cell imaging (*e.g.* the imaging of carbon nanotubes<sup>42</sup>). They typically use iterative local search and maximum entropy methods in both structure and scattering space, and are limited to a select set of scattering models<sup>42,43</sup>. Some of the most successful phase reconstruction tactics rely on recursive comparisons to the target scattering signature, and thus fall more appropriately under the category of iterative “generate-and-test” solutions, described below<sup>43</sup>.

**Pros:** The primary advantage of the direct approach is that it theoretically results in a full 3D representation of the scatterer. This is the best case scenario. It is also potentially quite rapid, as only a single model  $P_t$  need be considered and mathematically refined.

**Cons:** The direct method has been shown to be intractable for case of experimental light scattering through cells<sup>4,20</sup>; this is due to the nature of recording devices<sup>20,43</sup> / the lack of phase information<sup>4,20,40,43</sup>, the complex structure and optical properties of cells and organelles<sup>2,4,5,18,20,21,37</sup>, and the non-uniqueness of recorded patterns, due in part to rotation, translation, and scaling<sup>3,12,25</sup>. In practice,  $IM_t$  does not contain enough information to effect  $P_t = S^{-1}(IM_t)$ .

**State of the art:** There are currently no fully realized direct or numerical solutions to the inverse scattering problem of wide-angle laser light scattering from biological cells<sup>4</sup>.



**Figure 2.4:** *The parametric solution:* employ some function or series of functions  $F(IM_t)$  to extract a vector of key feature values from a scattering signature  $IM_t$  and relate these features to physical properties of the scatterer (*e.g.* cell/nucleus size, number of organelles).

### 2.2.3 Parametric Methods

**General Solution.** Possibly the most widely explored and successful approach to date, parametric inversion is a branch of methods that infer a set of particle or cell characteristics from selected indicatrices or scattering image features<sup>4</sup>. Instead of attempting to derive a full 3D model of the scatterer, patterns are instead related to a subset of cellular parameters. This is shown in Fig. 2.4, where parameters are derived from a target image using some function or set of procedures  $F(IM_t)$ . Scattering pattern features are chosen so as to be sensitive to specific aspects of scatterer structure<sup>4</sup>, and are mapped—via rulesets, mathematics, or machine learning<sup>4,11,25</sup>—to a parameter vector that describes key aspects of the scatterer (*e.g.* cell size, nuclear size, cell shape, number of component organelles, concentration, or optical properties such as effective index of refraction).

Typically, large bodies of simulated data (or in some cases carefully labeled experimental data) are used to develop a set of inference equations or “rules of thumb” that are then applied to new experimental data<sup>4,11,16,23,26</sup>. These inference equations usually extract features from scattering intensity samples taken at discrete intervals over a fixed 1D angular range—*i.e.* from an indicatrix<sup>4,11</sup>.

The parametric approach has been successfully demonstrated by a number of groups for the case of scattering indicatrices from cells<sup>4,11,16,25,26,44</sup>. For example, Maltsev has shown how a *flying light scattering indicatrix* (FLSI) parametric solution can determine the size and refractive index of a spherically modeled scatterer from scanning flow cytometry data, for particles ranging in diameter from 0.9–15  $\mu\text{m}$  with refractive index values between 1.37–1.60<sup>4</sup>.

Sem’yanov *et al.* have demonstrated a similar method to infer particle volume and hemoglobin concentration in red blood cells (RBC) via a 1D scattering indicatrix over a 15–55° range<sup>11,44</sup>. They use the location and magnitude of scattering intensity maxima and minima

to infer particle characteristics. Ghosh *et al.* also show a method to determine the size and hemoglobin content of RBC for scanning flow cytometry data, though they instead use features from the scattering signature's fast Fourier transform (FFT) to determine particle characteristics<sup>23</sup>.

Wilson and Foster further show how, using a light scattering sample, size distribution and scattering cross section, a Mie-theory-based particle sizing routine can extract information regarding the size of dominant scattering objects—in this case intracellular organelles—in a heterogeneous scatterer<sup>16</sup>. Fang *et al.* demonstrate an alternate Mie/Rayleigh-Gans approach that is able to determine the size distribution of organelles in a bulk solution well below the diffraction limit ( $< 130nm$ ) from light scattering spectroscopy data<sup>26</sup>.

Machine learning solutions have also been applied to effect parametric inverse solutions, with great success. Ulanowski *et al.* have developed a radial basis function neural network capable of determining size and index of refraction values for distributions of small spherical scatterers<sup>25</sup>. Their approach is theoretically expandable to any number of scattering reference points, and any number of particle characteristics, and was robust to missing or corrupted scattering intensity data<sup>25</sup>. With the generalization power and noise rejection properties of modern machine learning methods<sup>33</sup>, this approach could hold great promise.

As evident from the above examples, recent attempts to parametrically analyze the information contained in scattering patterns have focused on extracting parameters and mathematical relations from scattering intensity samples recorded at a set of fixed angles or over a 1D angular slice. In addition, it is important to note that many current methods<sup>4,11,23,25</sup> rely on scattering from a bulk population (*e.g.* a solution of mitochondria), as opposed to being able to ascertain the structure of an individual cell. While powerful, the derived relations only take into account a fraction of the available scattering information, and are therefore limited in their detection ability. With the advent of new 2D Finite Difference Time Domain (FDTD) simulation methods<sup>3,13</sup> and wide-angle cytometry schemes<sup>10,13</sup>, we now have the potential to extend previous parametric solutions to information-rich two-dimensional scattering signatures.

**Pros:** In the absence of a tractable direct inverse, the parametric approach still allows the inference of particle characteristics from scattered light. Most examples are very rapid, and have the added advantage of generalization: only certain relevant cellular features are extracted, as opposed to a full knowledge of the unique scatterer structure. This is advantageous in clinical situations—3D structure is not always needed for diagnosis, and most tests are performed on large cell populations. Once created, inference “rules of thumb” could be easily ported to LOC medical device firmware. Parameters may also be invariant to cell rotation and translation, which could alleviate some of the difficulties found in other 3D structure prediction schemes<sup>3</sup>.

**Cons:** One major disadvantage of the parametric approach is also its advantage—generality. It does not present insights into the exact structure of the scatterer, and is thus less useful for exploratory research. A large amount of labeled training data (usually simulated) is also needed before it is possible to form inference rules or linking equations, and training data must be carefully chosen so as not to bias the solution<sup>33</sup>. Lastly, depending on the inference method, the rules connecting indicatrix values and cell parameters might not be observable—*e.g.* the case of an ANN<sup>25</sup>—limiting the exploratory potential of the technique.

**State of the art:** Numerous groups have demonstrated successful parametric solutions, resulting in the characterization of one or more scatterer properties. However, there have been no examples in the literature of a parametric solution that takes into account a full 2D scattering signature (noting that Ulanowski *et al.*'s approach could be expanded to 2D).

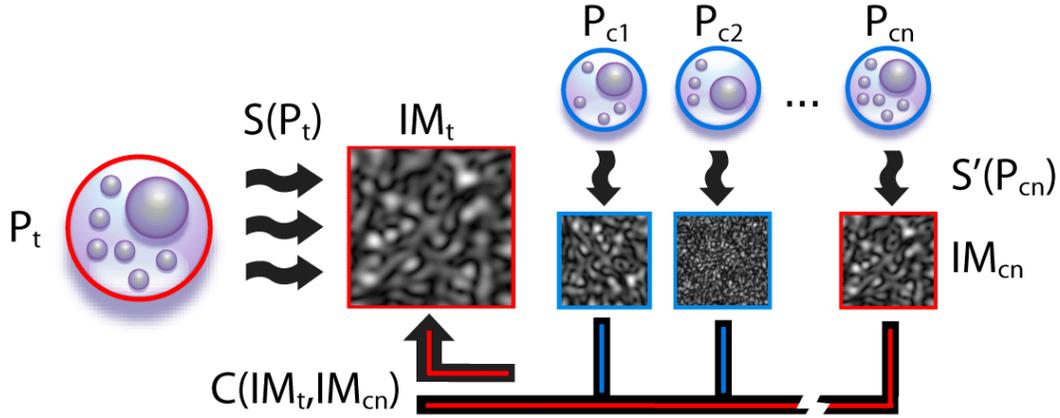
## 2.2.4 Iterative Methods

**Specific Solution.** As a successful alternative to direct or parametric inversion, several groups have shown the use of iterative “generate-and-test” methods to infer 3D structures from scattering measurements<sup>4,37,38,45</sup>. These methods hinge on the creation of candidate scattering profiles (*e.g.* via Mie/Rayleigh-Gans<sup>16,20,21,26</sup> or FDTD methods<sup>3,6,9,13,46</sup>) and iteratively comparing them to an experimental target scattering signature; the difference between target and candidate patterns is used to alter the test model until a match is made.

Fig. 2.5 outlines this method. A number of candidate signatures  $IM_{cn}$  are generated from a population of scattering distributions  $P_{cn}$ , via some simulation process  $S'(P_{cn})$ . These new signatures are compared to the target scattering pattern  $IM_t$  using some function  $C(IM_t, IM_{cn})$ ; a close match between signatures (either in terms of direct pixel comparison or low error between image parameters) indicates that  $P_{cn}$  may be a viable approximation of  $P_t$ , and the models are refined and re-compared until one or more are within some success threshold.

For the one-dimensional case, Maltsev describes how Mie theory scattering simulations have been used to generate potential angular scattering slices (indicatrices)<sup>4</sup>. These indicatrices are compared to the indicatrix of an experimental sample, and the test model is altered until a match is found<sup>4</sup>. The final model is used as an indication of the unknown experimental sample.

Javidi *et al.* show an image-based approach, where pattern matching is applied to microbial morphology<sup>45</sup>. In their system, holography images from a target sample with unknown structure are divided into a set of image features via a segmentation algorithm. By comparing the feature vector to a reference database via a Rigid Graph Matching (RGM) algorithm, they predict the structure of the unknown sample<sup>45</sup>.



**Figure 2.5:** *The generate-and-test (iterative) solution:* create a series of candidate cell models  $P_{cn}, n \in N$  and simulate their scattering signatures  $IM_{cn}$  via a function  $S'(P_{cn})$  that approximates the true scattering process  $S(P_t)$ ; compare each candidate pattern to the experimental target signature via a comparison metric  $C(IM_t, IM_{cn})$ . Once  $C(IM_t, IM_{cn})$  is below a given threshold, accept  $P_{cn}$  as one possible model for the target  $P_t$ .

While there are few examples of generate-and-test solutions for the case of cellular light scattering, X-ray diffraction literature regularly makes use of similar iterative methods. Notably, Chacon *et al.* demonstrate an elegant evolutionary computing approach to derive 3D protein structures from 2D X-ray scattering data. They use a genetic algorithm to model a scattering structure as a set of spheres, where the difference between the protein's actual scattering signature and Debye scattering simulations for each evolved model forms the comparison metric<sup>37,38</sup>. Their work demonstrated the accurate prediction of protein geometry, but incurred great computational cost. This appears to be the case for most generate-and-test approaches, precluding their use for real-time analysis, but this may change with future optimizations and advances in parallel computing.

Other examples of generate-and-test solutions include the use of evolutionary algorithms for crystal structure prediction<sup>43,47</sup> and Reverse Monte Carlo (RMC) algorithms to determine the structure of crystals, liquids, polymers, and other particle systems<sup>48–50</sup>. It is evident that iterative model refinement, coupled with rapid simulation and pattern matching, can produce excellent (though not always optimal<sup>49</sup>) solutions to inverse scattering problems approaching the complexity of those found in biological cells.

The examples above outline a key divide in iterative methods: *online systems* where candidate images are generated at runtime, with the candidate model being perturbed online (*e.g.* Chacon *et al.*<sup>37</sup>), and *offline systems* where candidate images and structures are generated prior to a run and stored for later reference (*e.g.* Javidi *et al.*<sup>45</sup>). The first instance,

online, requires a very rapid simulation method to generate images from candidate structures, but has minimal search, retrieval, and storage requirements. The second instance, database lookup/matching, may make use of highly accurate but computationally laborious simulation routines offline (*e.g.* FDTD<sup>3</sup>), but requires a large amount of storage and intensive database infrastructure, such as a content-based image retrieval (CBIR) system<sup>51</sup>. The offline case also has the limitation that new examples cannot be computed during a run—it cannot refine existing models.

In practice, these two approaches could be combined in a method similar to Schaeffer's recent solution to the game of checkers—a large database of pre-computed endgames is coupled with an online search heuristic to rapidly solve a sequence of optimal moves for feasible starting positions<sup>52</sup>. In a similar fashion, it may be possible to seed a generate-and-test inverse method with a sampling from a database of known candidate solutions; however, this has not yet been demonstrated in the scattering literature and is a viable topic for further exploration.

**Pros:** The key advantage of a generate-and-test solution is its ability to produce a specific 3D scatterer structure. The method is very flexible, simple to apply, and can be extended to include all relevant structural features. New example data and specification changes do not require the regeneration of processing rules or retraining of the system. It is also interpretable and verifiable: unlike some parametric approaches (*e.g.* the neural network of Ulanowski *et al.*<sup>25</sup>), it is possible to see visually and numerically how target and candidate images compare, and how this relates to structure. Image filters could also be easily applied to candidate data during comparison to help match different experimental setups (*e.g.* simulate experimental noise or integration effects).

**Cons:** One major disadvantage of iterative solutions is that they are potentially much slower and more computationally demanding than a parametric system. Unless implemented on massively parallel firmware, this could preclude their use in mainstream medical firmware or routine clinical diagnostics (though they would still be applicable in exploratory clinical research). The flexibility of the solution forces a tradeoff with its versatility of implementation. Output data, being a complex 3D structure, is also much harder to interpret for quick medical classifications. Additionally, there is no implicit guarantee that the generated solution will be an optimal representation of the target structure<sup>49</sup>. Another disadvantage is the reliance on accurate simulation methods, which must be repetitively applied and for cellular scattering problems could incur significant computational costs<sup>13</sup>.

**State of the art:** A generate-and-test inverse solution has not been successfully shown for wide-angle cellular scattering signatures, but has had great success on similar problems in X-ray diffraction crystallography and fault detection.

## 2.3 Key Components of the Problem

As presented in Ch. 1, if we hope to identify relationships between wide-angle scattering patterns and cellular structure, it is essential to develop methods that are able to extract meaningful associations from empirical evidence. Armed with the knowledge that direct inversion methods are for the most part intractable for the cellular light scattering problem<sup>4,20</sup>, it is valuable to explore the core components (and challenges) shared by both parametric and iterative solutions. As evidenced by the previous approaches described in the literature, both parametric and iterative methods do in fact share a number of key features. In the most general sense, they both require:

- Access to a large repository of labeled scattering signatures and/or a method to obtain new scattering patterns—this includes experimental data and simulated data.
- A way to characterize the complex intensity patterns contained in each signature.
- A way to accurately relate these patterns back to cellular structure.

These three requirements are explored in the following sub-sections, as they relate to the case of wide-angle cellular light scattering. While the focus is placed on 2D scattering images, the described requirements also impact—and stem from—indicatrix-based (1D) or fixed-angle scattering problems.

### 2.3.1 Experimental and Simulated Data

The accuracy of any solution to the cellular inverse scattering problem depends on how well it represents and conveys knowledge of the *ground truth*—the actual relationship between a scattering pattern and the scattering distribution. For a direct solution, this ground truth is built directly into the scattering equation and its inverse. However, for other solutions (*i.e.* parametric & generate-and-test), the ground truth needs to be somehow reconstructed from available empirical evidence. For a light scattering problem with no direct inverse, this empirical evidence can only be extracted from a large quantity of scattering examples (*e.g.*  $IM_c$ ) labeled with one or more quantitative properties—or some qualitative approximation—of their generating structure (*e.g.*  $P_c$ ).

As in data mining problems<sup>33,34</sup>, labeled image data can be gathered either through simulation techniques or experimental collection. The actual labeling of images can be done in a number of ways—*e.g.* by a human expert, automatically from the parameters of a simulator, from domain knowledge, or through some combination of these methods<sup>31,33,51</sup>.

Experimental cytometry images (*e.g.* from a LOC wide-angle cytometer<sup>10,13,27</sup> or fixed-angle / scanning flow cytometer<sup>4,11</sup>) can capture the light scattered by intact cells. However, due to the complex inner structure of cellular samples, its is often difficult, expensive, or in some cases impossible to obtain a sample's true internal structure (*e.g.*  $P_t$ ). At best, an image can be labeled with general information about cell type, approximate organelle content and distribution, and aspects of a cell's shape and behaviour. This is usually done by a human expert—a time consuming process. For medical classification, this could be sufficient, but it does little to aid in producing a *specific* solution.

While simulated images contain far less structural information than their experimental counterparts, they have the advantage that all images come with a comprehensive set of structural labels—the cell model and scattering system parameters used in image creation. As described above, this ability to generate labeled data has been a key component of many previous scattering analysis methods. General scattering models, such as Mie scattering or Rayleigh-Gans, have proved effective at rapidly generating approximate scattering signatures for cell-like structures, and come with a set of more general feature labels such as the number, spacing, and radius of scatters.<sup>3,4,20,21,26</sup> However, their detail is limited by structural approximations needed to make computation tractable (*e.g.* spherical scatter structure minimal phase shift<sup>3,20,21</sup>). Alternatively, FDTD simulations have been shown to produce very accurate scattering patterns for complex 2D and 3D cellular models<sup>1,3,3,6,13,14,46</sup>. They can be labeled with any number of parameters extracted from the internal optical structure of the scatterer, or even the full 3D scattering distribution. However, for any level of relevant cell complexity these methods are very slow, taking days even on multi-processor high-performance computing networks<sup>13</sup>. A detailed assessment of the simulation and data generation problem is presented in the introduction to Ch. 3, and new methods for acquiring labeled images are explored in Chs. 3 and 4.

**State of the art:** While essential to an inverse solution, prior to this work there was no effective way to both rapidly and accurately generate a large labeled dataset of 2D wide-angle scattering signatures that still transmit key medically relevant aspects of cellular structure.

### 2.3.2 Feature Extraction

Feature extraction is a critical aspect of both iterative and parametric solutions to the wide-angle inverse scattering problem—to be useful, the complex image topography of a scattering signature must be formatted in a understandable (and/or processable) fashion.

As discussed above, the goal of feature extraction in biomedical image processing is to take image data and extract from it a set of parameters that indicate the nature of the sample<sup>30,32</sup>.

The parameters to be extracted by this processing are dependent on the nature of the source image and on their intended use—parameters can include, but are not limited to, the location of edges or region boundaries (both in color or texture), the shape and size of regions in a scene, the histogram breakdown of image intensity and color, image frequency information, image regularity, and image texture data<sup>30,32,53</sup>.

The central problem in wide-angle scattering image analysis is extracting a set of image features that summarize the information-rich aspects of a scattering pattern. As indicated by previous parametric solutions, some image features contain more predictive value than others<sup>4,11</sup>. For instance, it has been demonstrated that things like the spacing and location of intensity bands in an image relate to the size of the cell and its nucleus<sup>4,11,23</sup>. Conversely, the number, size, and shape of intensity regions in the side-scatter region may correspond to the number and distribution of small organelles like the mitochondria<sup>13,54</sup>. Extracting numerical or semantic representations of these features from a scattering image is a challenging task, as noise and overlapping intensity data can obfuscate important patterns.

Feature extraction is also important for any form of image comparison, such as the similarity metrics used in generate-and-test solutions. When comparing images for similarity, a vector of image features can be used to judge the absolute difference between between a target and candidate image (for instance, their city-block or Euclidean distance)<sup>32,51</sup>. This can in some cases be more effective than raw pixel-by-pixel comparison (*e.g.* the  $\chi^2$  similarity metric used in Reverse Monte Carlo methods<sup>49</sup>), as a well-selected feature representation can be resilient to noise and redundancy in the analyzed data, while actively selecting for characteristics important to the application domain<sup>4,30,34</sup>.

The use of feature extraction in biomedical imaging applications has been shown for a number of different situations<sup>30–32,55–58</sup>. However, the task of applying feature extraction to wide-angle scattering images comes with its own set of unique objectives and challenges. Once information-rich features have been identified, they must be extracted under conditions where different image intensity regions blend together almost seamlessly, vary widely in intensity over their span, and deviate dramatically for different cellular distributions and arrangements<sup>3</sup>. These challenges and their image-processing background are described in greater detail in the introduction to Ch. 6; the process of feature extraction itself is addressed in detail in Chs. 4, 5, and 6.

**State of the art:** While there are methods to extract relevant information from 1D scattering indicatrices, prior to the present work there were no image processing and/or feature extraction methods tailored for use on the 2D scattering problem.

### 2.3.3 Linking Methods

Numerical relations, pattern analysis systems, classifiers, and rule generators are perhaps the most important component of any potential solution to the inverse problem. In most cases, these “linking methods” are heralded as the final methodological approach to solving an inverse problem, despite their strong dependence on both feature extraction and accurately labeled sample data. This emphasis is reasonable—the goal of a linking method is to take an image or its feature-based representation and correlate it with key aspects of a scatterer’s structure. Put differently, linking methods encapsulate or reference knowledge extracted from a database of labeled scattering data, giving them predictive power.

In most biomedical implementations, linking methods take as their input a vector of image features extracted from the target experimental image ( $IM_t$ )<sup>30</sup>. Their output is a general or specific breakdown of the target’s structure—an approximation of  $P_t$ . In its simplest form—*i.e.* the direct solution—a linking method could be regarded as the inverse scattering equation  $S^{-1}(IM_t)$ . An example of a this kind of mathematical linking method is the inverse Fourier transform, which links scattering amplitude and phase values to the shape and density of a scatterer<sup>22,39</sup>. However, as described earlier, a direct mathematical inverse is not tractable for the case of cellular light scattering. This puts the focus on linking methods that complement parametric or iterative approaches.

For parametric approaches, linking methods are typically based on rules or empirically generated inference equations (*e.g.* cell size determination via the location and pitch of intensity maxima<sup>4,11</sup>). However, the most successful approaches use machine learning algorithms<sup>25</sup>—a database of extracted parameters is used to train (or automatically generate) a classifier with predictive power. There are currently a number of machine learning techniques that could be used in a parametric setting to relate scattering features with structural parameters; these include artificial neural networks, support vector machines, bayesian networks, and tree classifiers<sup>33–36</sup>. Methods may also use a collection of algorithms designed to complement (or in some cases compete<sup>33,59</sup>) with each other to improve their predictive ability. Most notably in previous work, Ulanowski *et al.* demonstrated an artificial neural network that was able to distinguish scattering profiles using one-dimensional scattering samples (indicatrix values)<sup>25</sup>; their system proved effective for identifying the size and index-of-refraction for homogeneous spherical particles.

Generate-and-test solutions implement a different kind of linking method, where a system of image comparisons and/or database lookups (for instance evolutionary algorithms or content based image retrieval<sup>37,38,51</sup>) connects the target image to its structure via a set of candidate images and structures<sup>31,37,38,51</sup>. Each linking method has its own advantages and

disadvantages, and there is no universally regarded “best approach” that can be applied to all problems<sup>33,59</sup>. A comprehensive demonstration of linking methods as applied to the wide-angle scattering problem is presented in Chs. 5 and 7.

**State of the art:** As with feature extraction, while there have been demonstrations of linking methods for use with 1D scattering samples, prior to this work there were no successful applications of linking methods to the 2D scattering image case.

## 2.4 Summary

As described in this chapter, work on the wide-angle cellular inverse scattering problem draws on (and impacts) a number of other fields, from biomedical image analysis to crystallography and optical device design. Each of these fields comes with its own background knowledge and associated challenges. Taken as a whole, they give a solid basis for work presented in the following chapters, and serve as a resource of excellent tools to explore the challenging new domain of 2D scattering signatures. In a similar fashion, the interdisciplinary methods and techniques presented in this dissertation are expected to transfer new insights back to these fields, and give a starting point to engage other complex problems.

This dissertation builds on research done by the international light-scattering community. Previous approaches to the cellular inverse scattering problem can be broadly categorized into direct, parametric, and iterative methods, and can predict general or highly specific information about a scatterer. While direct mathematical inverse solutions are for the most part infeasible due to the constraints imposed by optical detection methods, work on parametric and iterative methods has been highly successful for indicatrix-based cellular scattering analysis and X-ray diffraction crystallography. Prior to this work, however, no similar approaches have been successfully developed for the case of wide-angle cellular scattering.

Previous approaches are tied together by a common set of requirements—the need for a source of labeled data, feature extraction routines, and powerful relationship-based “linking methods”. Each of these building blocks will be addressed in the following chapters, and connected to form several viable approaches to the interpretation of wide-angle scattering patterns. Chs. 3 and 4 address the problem of labeled scattering data by presenting a new simulation method. This method is then used as the source of labeled sample data for the remainder of the dissertation. Chs. 4, 5, and 6 show several approaches for feature extraction and the intelligent parameterization of scattering images. Lastly, Chs. 5 and 7 tie these developments together into a set of general and specific linking methods capable of identifying medically relevant aspects of wide-angle cellular signatures.

## References

- [1] A. K. Dunn, "Light scattering properties of cells," Ph.D. thesis, University of Texas at Austin (1997).
- [2] P. L. Gourley, J. K. Hendricks, A. E. McDonald, R. G. Copeland, K. E. Barrett, C. R. Gourley, and R. K. Naviaux, "Ultrafast nanolaser flow device for detecting cancer in single cells," *Biomedical Microdevices* **7**(4), 331–339 (2005).
- [3] C. Liu, C. E. Capjack, and W. Rozmus, "3-D simulation of light scattering from biological cells and cell differentiation," *Journal of Biomedical Optics* **10**(1), 014007 (12 pages) (2005).
- [4] V. P. Maltsev, "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [5] P. L. Gourley and R. K. Naviaux, "Optical phenotyping of human mitochondria in a biocavity laser," *IEEE Journal of Selected Topics in Quantum Electronics* **11**(4), 818–826 (2005).
- [6] R. Drezek, A. Dunn, and R. Richards-Kortum, "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements," *Applied Optics* **38**(16), 3651–3661 (1999).
- [7] L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, "Observation of periodic fine structure in reflectance from biological tissue: a new technique for measuring nuclear size distribution," *Physical Review Letters* **80**(3), 627–630 (1998).
- [8] R. Meyer, "Light scattering from biological cells: dependence of backscatter radiation on membrane thickness and refractive index," *Applied Optics* **18**, 585–588 (1979).
- [9] A. Karlsson, J. P. He, J. Swartling, and S. Andersson-Engels, "Numerical simulations of light scattering by red blood cells," *IEEE Transactions on Biomedical Engineering* **52**(1), 13–18 (2005).
- [10] K. Singh, X. Su, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, "A miniaturised wide-angle 2-D cytometer," *Cytometry A* **69A**, 307–315 (2006).
- [11] K. A. Sem'yanov, P. A. Tarasov, J. T. Soini, A. K. Petrov, and V. P. Maltsev, "Calibration-free method to determine the size and hemoglobin concentration of individual red blood cells from light scattering," *Applied Optics* **39**(31), 5884–5889 (2000).
- [12] D. Watson, N. Hagen, J. Diver, P. Marchand, and M. Chachisvilis, "Elastic light scattering from single cells: orientational dynamics in optical trap," *Biophysical Journal* **87**(2), 1298–1306 (2004).
- [13] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, "2D light scattering patterns of mitochondria in single cells," *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.

- [14] X.-T. Su, K. Singh, W. Rozmus, C. Backhouse, and C. Capjack, "Light scattering characterization of mitochondrial aggregation in single cells," *Optics Express* **17**(16), 13,381–13,388 (2009). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-17-16-13381>.
- [15] X.-T. Su, "Light scattering in an integrated microfluidic waveguide cytometer," Ph.D. thesis, University of Alberta (2008).
- [16] J. D. Wilson, C. E. Bigelow, D. J. Calkins, and T. H. Foster, "Light scattering from intact cells reports oxidative-stress-induced mitochondrial swelling," *Biophysical Journal* **88**(4), 2929–2938 (2005).
- [17] J. D. Wilson and T. H. Foster, "Mie theory interpretations of light scattering from intact cells," *Optics Letters* **30**(18), 2442–2444 (2005).
- [18] P. L. Gourley, "Biocavity laser for high-speed cell and tumour biology," *Journal of Physics D: Applied Physics* **36**(14), R228–R239 (2003).
- [19] E. Hecht, *Optics*, 4th ed. (Addison Wesley, San Francisco, 2002).
- [20] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1998).
- [21] H. van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1981).
- [22] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics* (John Wiley & Sons, New York, 1978).
- [23] N. Ghosh, P. Buddhiwant, A. Uppal, K. Majumder, H. S. Patel, and P. K. Gupta, "Simultaneous determination of size and refractive index of red blood cells by light scattering measurements," *Applied Physics Letters* **88**, 084,101 (3 pages) (2006).
- [24] L. Pilarski, E. Baigorri, M. Mant, P. Pilarski, P. Adamson, H. Zola, and A. Belch, "Multiple myeloma includes CD20+ B and plasma cells that persist in patients treated with rituximab," *Clinical Medicine: Oncology* **2**, 275–287 (2008). URL [http://la-press.com/article.php?article\\_id=659](http://la-press.com/article.php?article_id=659).
- [25] Z. Ulanowski, Z. Wang, P. H. Kaye, and I. K. Ludlow, "Application of neural networks to the inverse scattering problem for spheres," *Applied Optics* **37**(18), 4027–4033 (1998).
- [26] H. Fang, M. Ollero, E. Vitkin, L. M. Kimerer, P. B. Cipolloni, M. M. Zaman, S. D. Freedman, I. J. Bigio, I. Itzkan, E. B. Hanlon, and L. T. Perelman, "Noninvasive sizing of subcellular organelles with light scattering spectroscopy," *IEEE Journal of Selected Topics in Quantum Electronics* **9**(2), 267–276 (2003).
- [27] K. Singh, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, "Analysis of cellular structure by light scattering measurements in a new cytometer design based on a liquid-core waveguide," *IEE Proc.-Nanobiotechnology* **151**(1), 10–16 (2004).

- [28] P. M. Pilarski and C. J. Backhouse, "A method for cytometric image parameterization," *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.
- [29] A. Katz, A. Alimova, M. Xu, E. Rudolph, M. K. Shah, H. E. Savage, R. B. Rosen, S. A. McCormick, and R. R. Alfano, "Bacteria size determination by elastic light scattering," *IEEE Journal of Selected Topics in Quantum Electronics* **9**(2), 277–287 (2003).
- [30] R. M. Rangayyan, *Biomedical Image Analysis* (CRC Press, Boca Raton, 2004).
- [31] U. Sinha, A. Bui, R. Taira, J. Dionisio, C. Morioka, D. Johnson, and H. Kangarloo, "A review of medical imaging informatics," *Techniques in Bioinformatics and Medical Informatics* **980**, 168–197 (2002).
- [32] L. G. Shapiro and G. C. Stockman, *Computer Vision* (Prentice Hall, Upper Saddle River, N.J., 2001).
- [33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification: Second Edition* (Wiley Interscience, New York, 2001).
- [34] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. (Morgan Kaufmann, New York, 2005).
- [35] E. Alpaydin, *Introduction to Machine Learning* (MIT Press, Cambridge, 2004).
- [36] A. P. Engelbrecht, *Computational Intelligence: An Introduction* (John Wiley & Sons, New York, 2002).
- [37] P. Chacon, F. Moran, J. F. Diaz, E. Pantos, and J. M. Andreu, "Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm," *Biophysical Journal* **74**(6), 2760–2775 (1998).
- [38] P. Chacon, J. F. Diaz, F. Moran, and J. M. Andreu, "Reconstruction of protein form with X-ray solution scattering and a genetic algorithm," *Journal of Molecular Biology* **299**(5), 1289–1302 (2000).
- [39] N. Kasai and M. Kakudo, *X-Ray Diffraction by Macromolecules*, Springer series in chemical physics, 80 0172-6218 (Springer, New York, 2005).
- [40] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre, "Reconstruction of a yeast cell from X-ray diffraction data," *Acta Crystallographica, Section A: Foundations of Crystallography* **62**, 248–261 (2006).
- [41] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre, "Biological imaging by soft X-ray diffraction microscopy," *Proceedings of the National Academy of Sciences of the United States of America* **102**(43), 15,343–15,346 (2005).

- [42] J. M. Zuo, I. Vartanyants, M. Gao, R. Zhang, and L. A. Nagahara, "Atomic resolution imaging of a carbon nanotube from diffraction intensities," *Science* **300**(5624), 1419–1421 (2003).
- [43] G. Webster and R. Hilgenfeld, "An evolutionary computational approach to the phase problem in macromolecular X-ray crystallography," *Acta Crystallographica, Section A: Foundations of Crystallography* **57**, 351–358 (2001).
- [44] K. Sem'yanov and V. P. Maltsev, "Analysis of sub-micron spherical particles using scanning flow cytometry," *Particle & Particle Systems Characterization* **17**(5-6), 225–229 (2000).
- [45] B. Javidi, I. Moon, S. Yeom, and E. Carapezza, "Three-dimensional imaging and recognition of microorganism using single-exposure on-line (SEOL) digital holography," *Optics Express* **13**(12), 4492–4506 (2005).
- [46] R. Drezek, A. Dunn, and R. Richards-Kortum, "A pulsed finite-difference time-domain (FDTD) method for calculating light scattering from biological cells over broad wavelength ranges," *Optics Express* **6**(7), 147–157 (2000).
- [47] C. W. Glass, A. R. Oganov, and N. Hansen, "USPEX - Evolutionary crystal structure prediction," *Computer Physics Communications* **175**(11-12), 713–720 (2006).
- [48] T. Proffen and R. B. Neder, "DISCUS: A program for diffuse scattering and defect-structure simulation," *Journal of Applied Crystallography* **30**, 171–175 (1997).
- [49] R. L. McGreevy, "Reverse Monte Carlo modelling," *Journal of Physics: Condensed Matter* **13**(46), R877–R913 (2001).
- [50] V. FitzGerald, D. M. Pickup, D. Greenspan, G. Sarkar, J. J. Fitzgerald, K. M. Wetherall, R. M. Moss, J. R. Jones, and R. J. Newport, "A neutron and X-ray diffraction study of bioglass with reverse Monte Carlo modelling," *Advanced Functional Materials* **17**(18), 3746–3753 (2007).
- [51] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications - clinical benefits and future directions," *International Journal of Medical Informatics* **73**(1), 1–23 (2004).
- [52] J. Schaeffer, N. Burch, Y. Bjoernsson, A. Kishimoto, M. Mueller, R. Lake, P. Lu, and S. Sutphen, "Checkers is solved," *Science* **317**(5844), 1518–1522 (2007).
- [53] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognition* **26**, 1277–1294 (1993).
- [54] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, "Rapid simulation of wide-angle scattering from mitochondria in single cells," *Optics Express* **16**(17), 12,819–12,834 (2008). URL <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-17-12819>.
- [55] E. Glory, G. Derocle, N. Ollivier, V. Meas-Yedid, G. Stamon, C. Pinset, and J. C. Olivo-Marin, "An automatic image analysis approach to quantify stained cell cultures," *Cellular and Molecular Biology* **53**(2), 44–50 (2007).

- [56] N. Richard, M. Dojat, and C. Garbay, "Automated segmentation of human brain MR images using a multi-agent approach," *Artificial Intelligence in Medicine* **30**(2), 153–175 (2004).
- [57] B. Prasad, S. Du, W. Badawy, and K. Kaler, "A real-time multiple-cell tracking platform for dielectrophoresis (DEP)-based cellular analysis," *Measurement Science & Technology* **16**(4), 909–924 (2005).
- [58] M. Schmidt, "Automated brain tumor segmentation," Master's thesis, University of Alberta (2005).
- [59] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Natural Computing Series, 2nd ed. (Springer, New York, 2007).

## Chapter 3

# Rapid Simulation of Wide-Angle Scattering from Mitochondria in Single Cells

### 3.1 Introduction

This chapter shows that it is possible to produce a good approximation to the scattering pattern of small, isotropically scattering mitochondria by adapting classical X-ray diffraction theory (XRD). We also show how this theory may be extended to deal with the scatter patterns arising from collections of mitochondria whose individual scatter patterns are no longer isotropic, and a more general case: heterogeneous combinations of cellular micro- and nano-structures. The XRD framework provides a wealth of tools that could be applied to this problem. Although XRD is commonly thought of in conjunction with large ('infinite') crystalline structures, it is also applicable in cases of finite and non-crystalline structures—much like the situation of scattering from a single cell.

Our approach allows for the rapid simulation of mitochondrial scattering patterns, enabling a number of powerful linking methods. By quickly simulating large libraries of images that approximate the results of mitochondrial scattering, parametric computer methods could be used to identify key correlations between scattering patterns and scatterer structure. These correlations may be used in an experimental situation to rapidly (*e.g.* real-time) predict the nano-structural makeup of a single cell from scattering patterns captured by a wide-angle cytometer (*e.g.* the work of Singh *et al.*<sup>1,2</sup>). Access to a rapid simulation routine also facilitates iterative reverse methods; approaches such as Reverse Monte Carlo (RMC) analysis<sup>3</sup> can use simulation results to iteratively adapt a scattering model and to fit the profile of an experimentally observed scatterer. Both parametric and iterative approaches

---

*A version of this chapter has been published. Pilarski et al. 2008. Optics Express. 16(17): 12819–12834.*

depend on the availability of an effective and rapid method of calculating a realistic scatter pattern. The present work gives a theoretical basis for a new simulation routine that fits these requirements, and demonstrates its effectiveness in simulating organelle-related light scattering.

## 3.2 Background

Previous work has shown that the light scattered by biological cells can be used to infer some aspects of internal cellular structure<sup>4–8</sup>. This is pertinent to emerging wide-angle cytometry systems currently under development, *e.g.* Singh *et al.*<sup>2</sup>, where the rich structure of two-dimensional scattering signatures from single cells may be used to explore the micro- and nano-structural makeup of the scattering source<sup>2,8</sup>. Structural information on the intracellular components (such as the organization and number of mitochondria) has great clinical relevance, as it may be used to detect and characterize certain diseases, such as cancer<sup>9–12</sup> and cardiomyopathy<sup>10</sup>. There is a great need for analysis and simulation methods to describe the relationship between experimental scattering patterns and cell structure.

A number of groups have recently worked on the assessment of scatter structure from light intensity measurements or predictions, and these approaches may be categorized into *forward methods* and *reverse methods* depending on how they relate to the simulation process. Forward methods rely on the prediction of many feasible scattering patterns based on a knowledge of the scattering structure and a model of light propagation (*e.g.* Finite Difference Time Domain (FDTD) simulations<sup>13–15</sup>), while reverse methods attempt to deduce some aspect of scatterer geometry from a pattern of scattered light (again using a model or algorithm to relate pattern to structure). In practice, many iterative and parametric approaches are a hybrid of forward and reverse methods.

A true reverse method (*i.e.* a direct solution to the inverse scattering problem) would give the exact and detailed geometric structure of a cellular scatterer from a pattern of scattered light. Such a direct reverse solution has been shown to be computationally intractable for the problem of biological cells<sup>5</sup>. However, some advances in the reverse domain have been able to extract one or two physical aspects of the scatterer. Of note, an ‘indicatrix’ has been used on one-dimensional scattering slices and collections of angular slices to determine cell size and hemoglobin content<sup>5–7</sup>. Scattering has also been used to determine red blood cell size and refractive index via a Fourier transform<sup>16</sup>. Other groups have worked on predicting the properties of bulk solutions of multiple scattering bodies using Mie or Rayleigh-Gans (R-G) theory fitting and Light Scattering Spectroscopy (LSS)<sup>17,18</sup> or elastic/angularly resolved light scattering<sup>19–21</sup>.

Due to the complexity of the forward scattering problem, scattering from individual organelles such as mitochondria is usually described in the literature by Mie and R-G theory<sup>19,20,22</sup>. While mitochondria have been described (and treated) as ellipsoids or spheroids, with diameters in the 400–800nm range and lengths in the 800–3000nm range<sup>23–25</sup>, there is in fact a startlingly wide variety of shapes and sizes documented for these organelles<sup>26</sup>. Nevertheless, it has been demonstrated that mitochondria are the primary large-angle scatterers in eukaryotic cells<sup>27</sup>.

Simulation methods such as those based on the Mie and R-G theories give some analytical basis for the intensity patterns observed from cytometry experiments (*e.g.* intensity banding<sup>2,28</sup>), but have significant restrictions in their use (*e.g.* for Mie theory, treating scattering bodies as spheroids). More recently, FDTD code—a discretized solution to Maxwell’s equation—has been used to predict scattering from cells<sup>15</sup>. While extremely detailed and true to experimental wide-angle cytometry results<sup>8</sup>, FDTD is computationally intensive for any wide-angle 2D cell simulation with realistic parameters<sup>8</sup>; such simulations must be run on large super-computer arrays and may take hours or days to generate a simulated pattern<sup>8</sup>.

Though there are many feasible simulation methods, as described above, to date the FDTD method appears to be one of the most effective methods capable of generating realistic wide-angle scattering patterns from a three-dimensional cell model<sup>8,15</sup>. However, this leads to a computational bottleneck when generating large numbers of simulations, and does not provide a computationally tractable complement to the reverse problem of determining scatterer geometry from an experimental scattering pattern. If we hope to characterize and rapidly classify the effect of nano-structural cell components on scattering, another simulation method must be developed.

## **3.3 Methods**

### **3.3.1 A Theoretical Approach**

As described above, the analysis of nanostructural contributions to scattering in complete cells has so far proved quite difficult. However, it is known that organelles in general, and mitochondria in particular, are the dominant cause of large-angle scattering (*i.e.* light scattered perpendicular to the path of the incident light, commonly called side-scatter or large-angle scattering). As we demonstrate in this work, it is therefore possible to rapidly simulate and analyze the important aspects of large-angle cellular scattering by examining the scattering behaviour of mitochondria.

It has been shown that approximately 90% of side-scattered visible light from human cells is due to the presence of mitochondria<sup>9,23</sup>. Small structures, such as the mitochondria, readily scatter at large angles and are in fact the dominant cause of intensity artifacts in this angular region<sup>9,23,29</sup>. This is due in part to the complex internal structure of mitochondria and their numerous index-of-refraction changes<sup>4,9</sup>. As seen from recent experimental and simulation work<sup>8</sup>, side-scattered light from the mitochondria in human cells (*e.g.* immune system cells such as the Raji cell-line<sup>8</sup>) typically takes the form of small asymmetrical ‘blobs’ in a two-dimensional scattering pattern, where blob size and spacing is related to the distribution of small scattering bodies within the cell<sup>8,30</sup>. Conversely, larger cell structures—such as the nucleus and the cell wall—lead to high intensities of forward and back scatter (*i.e.* angles approaching the path of the incident light)<sup>9,18,23,29</sup> and broad intensity banding in scattering images<sup>8</sup>.

This is easily understood by examining the scattering regimes present: cellular components much smaller than the incident wavelength will scatter uniformly in all directions (*i.e.* isotropically) or near-isotropically via the Rayleigh ( $size \ll \lambda$ ) and Mie regimes ( $size \leq \lambda$ ), while larger bodies will scatter along the light path according to geometric transmission ( $size \gg \lambda$ )<sup>31</sup>. For more information, a detailed description of light scattering by human cells is presented by Gourley *et al.*<sup>9</sup>.

Given the dominance of mitochondrial scattering in the large-angle domain, as verified by the recent experiments of Su *et al.*<sup>8</sup>, we show that it is possible to model isotropic and anisotropic scattering of mitochondria by adapting classical XRD theory. Although this theory can be extended to larger (anisotropic scattering) mitochondria, in the limit of small mitochondria we find that they can be treated in the same way as isotropically radiating electrons in standard XRD. This greatly reduces the computational burden and provides a means of partially solving the inverse problem.

Another way to view this problem is using R-G theory—the assumed independence and isotropic radiation of all scattering volume elements in a particle or scatterer. Many volume elements would be effectively “empty”, and could be passed over to speed computation. There are a number of similarities between such a modified R-G formulation and our approach described below. However, as XRD theory has a number of other important theoretical advantages—for instance, a mathematical (Fourier) framework to interpret the interaction of different scattering shapes and structures<sup>32</sup>—we approach the problem from this vantage point.

To simplify the discussion, from this point on we will use and refer to the standard XRD terminology and notation presented by Kasai and Kakudo<sup>32</sup>; relevant terms will be redefined as needed to ensure clarity for the reader.

### 3.3.2 A Basis in X-ray Diffraction Theory

It can be shown that large-angle cellular light scattering from small mitochondria fulfills the fundamental constraints and assumptions for XRD-like analysis<sup>32</sup>—specifically: *Thompson scattering*—the incident and scattered wavelengths are the same; *Scatterers radiate isotropically*; *the Fraunhofer approximation*—energy arrives as a plane wave at the receptive field (*i.e.* the detection plane, likely a charge-coupled device (CCD)); and *the kinematical theory of diffraction*—secondary and tertiary scattering interactions (*i.e.* multi-scatter) should be negligible compared to primary scattering and wave interference<sup>32</sup>. First, like XRD, the incident and scattered wavelengths are the same (*i.e.* elastic scattering). Secondly, in the limit of very small mitochondria (*e.g.* sub-wavelength nanostructures, *i.e.* on the order of a 100nm in size), side-scatter will be very close to isotropic, especially when observed via a small solid angle in the side-scatter region<sup>9</sup>. Thirdly, for the geometric arrangement of a wide-angle cytometer<sup>2,8</sup> the distance to the receptive plane (millimetres) is much greater than the distance between scatterers (micron scale), thus fulfilling the Fraunhofer approximation. Finally, for a standard cytometry system, the impact of multi-scatter, especially for a single-cell situation, is expected to be minimal—first order scattering is barely detectable; second order effects will not be detectable.

Given this, we can begin to examine the mitochondrial scattering problem from an XRD perspective. Neglecting a common prefactor, the amplitude measured at a point  $\mathbf{s}'$  (on a detecting plane or receptive field, *e.g.* a CCD) of the collective scattering of a set of isotropic scatterers at locations  $\mathbf{r}_n$  is given by the ‘explicit’ formulation:

$$A(\mathbf{s}') = \sum_{n=1}^N \{ \cos[2\pi \cdot (|\mathbf{s}' - \mathbf{r}_n| + d_n^o)] - i \sin[2\pi \cdot (|\mathbf{s}' - \mathbf{r}_n| + d_n^o)] \} \quad (3.1)$$

Where path length  $d_n^o$  indicates the distance from a scattering point  $\mathbf{r}_n$  to the light source or incident reference plane, and all distances are measured in units of the wavelength,  $\lambda$ . The intensity on the detecting plane can then be described by the magnitude of the recorded amplitude values (*i.e.* the multiplication of complex conjugates<sup>32</sup>:  $I(\mathbf{s}') = |A(\mathbf{s}')|^2$ ). As detailed by Kasai and Kakudo,<sup>32</sup> the intensities,  $I(\mathbf{s}')$  calculated using Eq. (3.1) are equivalent to those from another ‘explicit’ calculation of  $A(\mathbf{S})$ :

$$A(\mathbf{S}) = \sum_{n=1}^N \exp\{-2\pi i(\mathbf{S} \cdot \mathbf{r}_n)\} \quad (3.2)$$

Where  $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$ , and  $\mathbf{s}_0$  is a unit vector that represents the direction of the illuminating beam,  $\mathbf{s}$  is a unit vector that represents the scatter direction of interest and the product  $\mathbf{S} \cdot \mathbf{r}_n$  represents the phase shift (or “path difference”<sup>32</sup>) of the detected light along the observation angle<sup>32</sup>. The path difference in the distance term  $|\mathbf{s}' - \mathbf{r}_n| + d_n^o$  is the real-space equivalent of Kasai and Kakudo’s reciprocal path difference  $\mathbf{S} \cdot \mathbf{r}_n$ <sup>32</sup>; both formulations can be shown to have the same amplitude behaviour at any point in the far field.

With  $\rho(\mathbf{r}) = \sum_{n=1}^N \delta(\mathbf{r} - \mathbf{r}_n)$ , Kasai and Kakudo showed that this can also be written as a volume integral, over  $v_r$  (their Eq. (2.13)<sup>32</sup>):

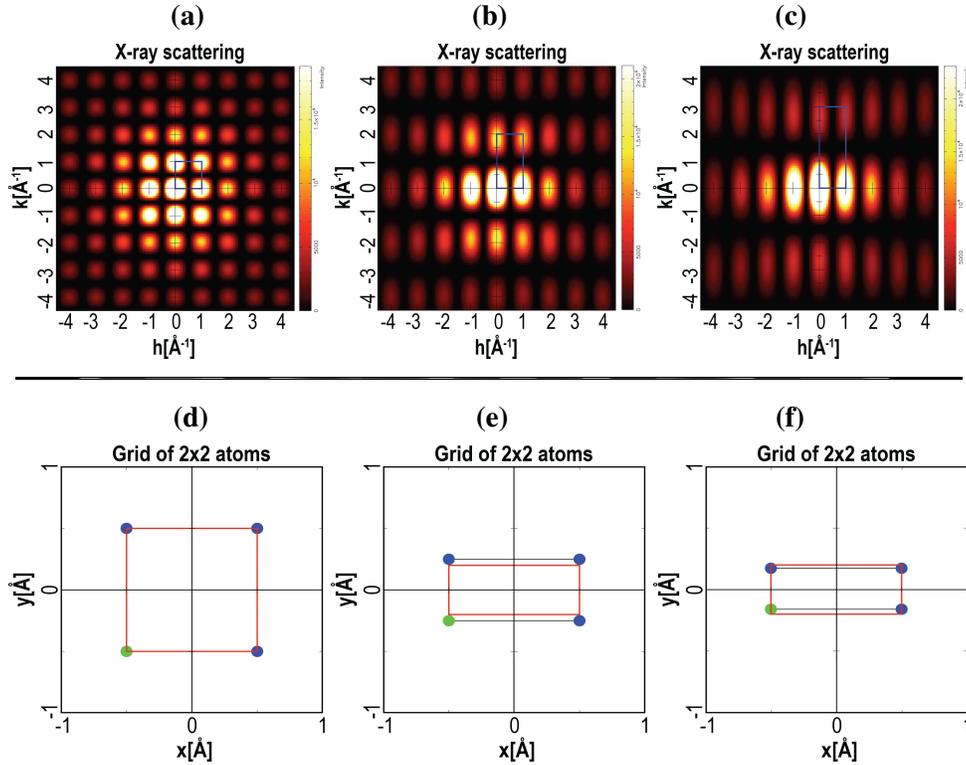
$$A(\mathbf{S}) = \int_0^\infty \rho(\mathbf{r}) \exp\{-2\pi i(\mathbf{S} \cdot \mathbf{r})\} dv_r \quad (3.3)$$

We see that, in Eq. (3.3),  $A(\mathbf{S})$  is the ‘forward’ Fourier transform of the scattering distribution  $\rho(\mathbf{r})$ , and relates directly to the amplitudes  $A(\mathbf{S})$  on the receptive field. If we knew  $A(\mathbf{S})$ , we could take an inverse Fourier transform (in the reciprocal space) to determine the exact position of the scatterers, *i.e.* (Kasai and Kakudo’s Eq. (2.14)<sup>32</sup>):

$$\rho(\mathbf{r}) = \int_0^\infty A(\mathbf{S}) \exp\{2\pi i(\mathbf{S} \cdot \mathbf{r})\} dv_s \quad (3.4)$$

The central challenge of XRD methods is that we cannot record  $A(\mathbf{S})$ , but only  $I(\mathbf{S})$ . As such, we lose all phase information and the inverse transform cannot be made readily. Nevertheless, the forward transform is useful to allow patterns to be compared and potential structures investigated<sup>3</sup>, the discovery of the helical structure of DNA being a famous example. In addition, although this is not our focus here, XRD methods exist to use  $I(\mathbf{S})$  to infer  $\rho(\mathbf{r})$ . This is commonly used to predict macro-molecule and protein structure from scattering and diffraction patterns with the help of phase reconstruction algorithms<sup>33–35</sup> or methods such as the Reverse Monte Carlo analysis<sup>3</sup>. With such a wealth of tools, it is clear that an XRD-like approach to cytometry would have significant benefits.

As pointed out by Proffen and Neder<sup>3</sup>, there is often a significant advantage to using one of the above ‘explicit’ formulations rather than relying on other methods of performing Fourier transforms—in many circumstances these explicit methods are very much less computationally challenging. This is especially the case when dealing with small numbers of scatterers—a situation that readily arises with microfluidics or “lab on a chip” (LOC) approaches that allow the manipulation and interrogation of a single cell. In a single cell, we deal with hundreds (or fewer) scattering centres (*i.e.* mitochondria) and the explicit methods can be readily performed on any type of calculation device (rather than needing a super-computer as for FDTD methods).



**Figure 3.1:** Example of how changes to the spatial distribution,  $\rho(\mathbf{r})$ , (d) through (f), impact the spacing of features within two-dimensional scattering patterns,  $I(\mathbf{S})$ , (a) through (c), in X-ray scattering. As the y-axis spacing between atoms *decreases* by a factor of two and three in panels (e) and (f), y-axis spacing in the reciprocal plot *increases* by a factor of two and three in panels (b) and (c). Physical and Fourier dimensions are listed in  $\text{\AA}$  and  $\text{\AA}^{-1}$  on the plots above. Data was generated using DISCUS, the Fourier-transform based scattering simulator of Proffen and Neder<sup>3,36,37</sup>.

### 3.3.3 Scattering Examples

As a simple visual example of the relationship between scattering intensities (*e.g.* from Eq. (3.3)) and the distribution of scatterers, Fig. 3.1 shows how changes to spatial distribution ( $\rho(\mathbf{r})$ ; Fig. 3.1, bottom row) affect the spacing of features in a two-dimensional X-ray scattering pattern ( $I(\mathbf{S})$ ; Fig. 3.1, top row) of a small  $2 \times 2$  grid of atoms (*i.e.* 4 atoms in total) with horizontal spacings of  $1\text{\AA}$  and vertical spacings of  $1\text{\AA}$  (a),  $0.5\text{\AA}$  (b), and  $0.33\text{\AA}$  (c). This figure was generated using DISCUS, the widely used X-ray scattering simulator of Proffen and Neder<sup>3,36,37</sup>. DISCUS is a XRD teaching and simulation tool that applies explicit (discrete) Fourier transforms to simulate the scattering from complex bounded (*i.e.* non-infinite and constrained) crystals and collections of atoms<sup>3,36,37</sup>. In this example (Fig. 3.1), as the y-axis spacing between atoms *decreases* by a factor of two and three, from  $1\text{\AA}$  to  $0.33\text{\AA}$ , panels d) through f), vertical spacing between features in the reciprocal plot (*e.g.* the scattering or Fourier domain) *increases* by a factor of two and three, from  $1\text{\AA}^{-1}$  to  $3\text{\AA}^{-1}$ , panels a) through c). This calculation is equivalent to that of Eq. (3.3).

**Table 3.1:** The mtPatterns algorithm

---

1	Create an array $\rho(\mathbf{r})$ of $0 < n \leq \#mt$ scattering points in $\mathcal{R}^3$ space within $[R_i, R_o]$
2	Create a receptive field $\mathbf{I}$ as an array of size $U \times V$ , normal to the $z$ -axis
3	Position receptive field $\mathbf{I}$ distance $d$ above population center $\mathbf{r}_o$ along $z$ -axis
4	FOR all $\mathbf{s}' \in \mathbf{I}$ :
5	FOR $n = [1, \#mt]$ :
6	CALCULATE: $A_e(\mathbf{s}', \mathbf{r}_n), A_m(\mathbf{s}', \mathbf{r}_n)$
7	IF non-uniform scattering: $A(\mathbf{s}', \mathbf{r}_n) = A(\mathbf{s}', \mathbf{r}_n) \cdot M(\theta, \phi, \theta_o, \phi_o)$
8	UPDATE: $A_e(\mathbf{s}') += A_e(\mathbf{s}', \mathbf{r}_n); A_m(\mathbf{s}') += A_m(\mathbf{s}', \mathbf{r}_n)$
9	$\mathbf{I}(\mathbf{s}') = A_e(\mathbf{s}')^2 + A_m(\mathbf{s}')^2$
10	Return: $\mathbf{I}$

---

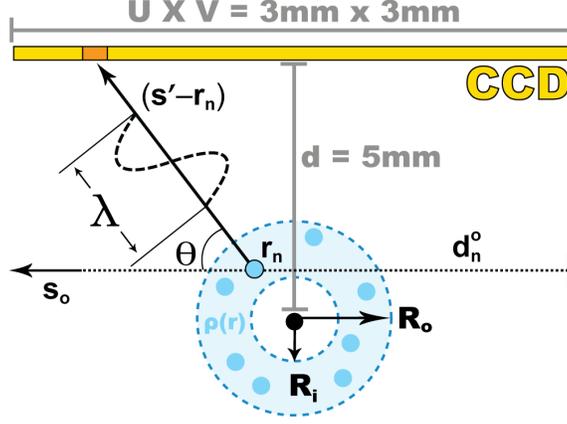
Takes:  $\{\#mt, R_i, R_o, U, V, d, d^o, M\}$ , Returns:  $\{\mathbf{I}\}$

### 3.3.4 The mtPatterns Algorithm

With this theoretical background in mind, we now present a novel method to quickly and inexpensively simulate the large-angle scattering from a series of mitochondria-like scatterers. As discussed above, a LOC approach allows us to measure the scatter pattern of the hundreds of mitochondria from a single immobilized cell. Given the finite number of scatterers, it is most appropriate to use an explicit formulation to calculate the scatter, specifically Eq. (3.1)—*i.e.* we simply sum the radiation from each source within an arbitrary distribution. In essence, it is advantageous to treat the cell as being a disordered crystal comprised of mitochondria. As indicated in the previous section, the basic form of this algorithm assumes that point sources radiate isotropically, that incident and radiated light is of the same frequency, that the interaction between scatterers (multi-scatter) is negligible, and that the spacing between scatterers is much smaller than the distance to the receptive field (*i.e.* the Fraunhofer approximation).

**Overview.** The mtPatterns algorithm (Tab. 3.1) is a procedural implementation of Eq. (3.1), presented and derived in Sec. 3.3.2. It takes a user-specified distribution of scatterers and generates a two-dimensional scattering plot as its output. Scattering calculation is done by independently summing the isotropic scattering contributions of each scatterer for every point on the receptive field. For the case of anisotropic scattering, a point-spread function may be applied to extend the algorithm to the case of larger spherical or variably shaped mitochondria.

**Algorithm Flow.** To begin, the algorithm uses user input to create the scatter distribution and receptive field (steps 1–3). For each point on the receptive field, the algorithm next collects and sums the individual real and imaginary amplitude contributions described in Eq. (3.1) (steps 4–8). At this point, the amplitude of each scatterer may also be scaled by a



**Figure 3.2:** Diagram of the scattering simulation process. The ray  $\mathbf{s}_o$  indicates the direction of incident light, while the segment  $(\mathbf{s}' - \mathbf{r}_n)$  describes the vector between scatterer  $\mathbf{r}_n$  and point  $\mathbf{s}'$  on the receptive field.  $R_o$  and  $R_i$  indicate the inner and outer radius of the scattering distribution  $\rho(\mathbf{r})$ .

known point-spread function to compensate for anisotropic scattering (step 7). Finally, the algorithm computes the total intensity value at each pixel, and returns the collected intensity values to the user in image and array format (steps 9–10).

**Definitions.** The following definitions are used in Tab. 3.1 and the remainder of this document.  $\rho(\mathbf{r})$  is a three-dimensional array of isotropic scatterers (*mt*), created from a user-specified spatial distribution of scattering points  $\mathbf{r}_n$ ,  $n \in [1, \#mt]$ . (*n.b.* ‘*mt*’ is shorthand for ‘mitochondria’.) For this work we assume that the bounds of this distribution are a spherical shell of inside radius  $R_i$  and outside radius  $R_o$  (*i.e.* the mitochondria are located outside the nucleus of the cell). However, any arbitrary volume may be specified.  $\mathbf{I}(u, v)$  is a two-dimensional receptive field created with size  $U \times V$  and at a specified distance  $d$  along the  $z$ -axis away from the scattering population’s centroid  $\mathbf{r}_o$ . A schematic diagram of this setup is shown in Fig. 3.2. From this point on, we use the vector  $\mathbf{s}'$  to represent a point on the receptive field with coordinates  $(u, v, d)$  from the origin. For scattering computation,  $A_e$  and  $A_m$  are the real and imaginary amplitude contributions described in Eq. (3.1), and  $M(\theta, \phi, \theta_o, \phi_o)$  is a point-spread function which may be used to scale amplitudes by a known angular distribution. As noted in Sec. 3.3.2, amplitude calculation is based on  $|\mathbf{s}' - \mathbf{r}_n| + d_n^o$ , the path difference between the observing point  $\mathbf{s}'$  and the scattering point  $\mathbf{r}_n$ . Light source distance  $d_n^o$  for each scatterer is calculated based on a light source origin located a fixed distance from  $\mathbf{r}_o$  along the  $x$ -axis.  $I(\mathbf{s}') = |A(\mathbf{s}')|^2 = A_e(\mathbf{s}')^2 + A_m(\mathbf{s}')^2$  is the summation of real and imaginary amplitude components for each pixel, which form the output intensity profile of the scattering image. To compress the dynamic range of the observed images, we plot the square root of  $I(\mathbf{s}')$ .

From this point on we use spherical coordinates to specify location and spacing in the scattering image and related point-spread functions. This allows us to describe patterns in terms of the angle between the path of the incident light and the ray connecting the scattering object to a point on the receptive plane. We define  $\theta$  as the angle between the scattered light wave vector ( $\mathbf{s}' - \mathbf{r}_n$ ) and the incident light wave vector  $\mathbf{s}'_o$  (Fig. 3.2), and  $\phi$  is the angle between the projection of the scattered light wave vector in xy-plane and the x-axis. For reference,  $\theta = 0^\circ$  is pure forward scattering,  $\theta = 180^\circ$  is pure back scattering. Similarly,  $\phi = 0^\circ$  is scattering towards the centre of the receptive field. In the case of spherical scatterers there is no breaking of the symmetry of the system and so there is no  $\phi$  dependence.

For isotropic scatter the above algorithm suffices to calculate  $I(\mathbf{s}')$ . However, as will be discussed below, a number of circumstances can give rise to anisotropic scatter. To allow the calculation of anisotropic scatter we include, as an option, a point-spread function  $M(\theta, \phi, \theta_o, \phi_o)$  to allow for scatter in the direction specified by  $\theta$  and  $\phi$  from a scatterer aligned in a direction specified by  $\theta_o$  and  $\phi_o$ . The function,  $M(\theta, \phi, \theta_o, \phi_o)$ , may be applied to scale the isotropic radiation of each scatterer by a known set of angular intensity values (which may be generated numerically<sup>8,15</sup> or empirically). As will be described below, we have not yet found the need to extend this analysis beyond its application to spherical mitochondria (whether scattering isotropically or anisotropically). As such, we need only consider  $M(\theta)$ , to allow the simulation of anisotropic scatter from larger, spherical mitochondria. In this work,  $M(\theta)$  was a look-up table of real values from 0.0 to 1.0, indexed by the angle  $\theta$ . Continuous point-spread functions may also be used. Polarisation effects could also serve to break the symmetry to the extent of requiring a  $M(\theta, \phi)$  for the simulation of scatter from spherical mitochondria. However, since the present work bases its analysis upon the point-spread functions provided by Gourley *et al.* and these do not contain a  $\phi$ -dependence, we do not consider such effects here, although the algorithm could easily do so if the more detailed point-spread functions were available (and needed).

### 3.3.5 Isotropic Scattering from Mitochondria in the Mie Regime

As discussed by Gourley *et al.*<sup>23</sup>, many mammalian mitochondria have diameters in the range of 400–800nm. To fall within the realm of true Rayleigh scattering, a structure must be significantly smaller than  $\lambda$  (in this case less than 100nm). Thus, we expect the majority of human mitochondria to scatter in the Mie regime (*i.e.* anisotropically). However, recent work by Gourley *et al.* has shown that mitochondria-like biospheres do in fact exhibit isotropic behaviour up to approximately 500nm in diameter<sup>23</sup>.

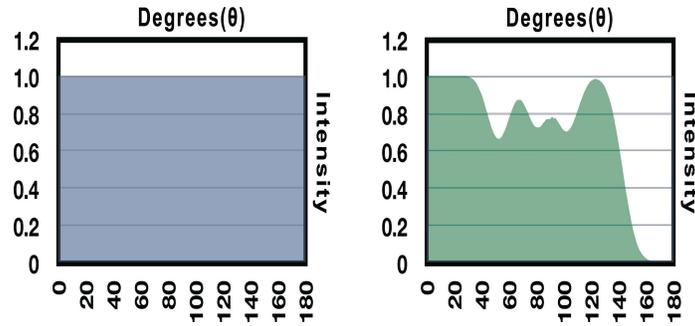
In their Fig. 1, Gourley *et al.*<sup>23</sup> present their results using FDTD code to assess the degree of anisotropy that will be produced by spherical mitochondria 100nm, 500nm and 1000nm in diameter. Although we extract quantitative data from their graphs, the uncertainties in doing so are large enough that the results should be interpreted qualitatively. From Gourley *et al.*'s graphs, we see no discernible anisotropy for the 100nm diameter scatterers. For a 500nm mitochondrion, we estimate from that graph that the anisotropic variation seen across a 30-degree-wide cone will be approximately 30%, whereas for the 1000nm diameter mitochondria we estimate a 300% variation over the same range. This small variation for 500nm mitochondria indicates that the scattering behaviour of the mitochondria in human cells is likely to be well modeled by a simulation based on isotropic scatter.

### 3.3.6 Test Images

Using `mtPatterns`, we created an extensive set of test images and compared them (via qualitative observations and quantitative methods such as comparing characteristic blob spacing) to experimental images generated by a miniaturized cytometry device. These results and comparisons are presented in Sec. 3.4. In addition to a bank of test images with cell size and nuclear size values  $R_o = 0.1\mu m - 20.0\mu m$  and  $R_i = 0.1\mu m - 20.0\mu m$  (with the number of simulated mitochondrial scatters,  $\#mt$ , varying from 5–1000), additional tests were performed with the specific cell parameters for the Raji human cell line given by Su *et al.* ( $R_o \approx 8.0\mu m$ ,  $R_i \approx 4.0\mu m$ ,  $\#mt \approx 83-677^8$ ). This allowed us to explore the behaviour of our algorithm over a wide range of parameters, and also compare simulations to actual experimental cytometry results for Raji cells. Each test was performed at least three times with different random mitochondrial placements (with similar results obtained for each).

The following parameters were used in all `mtPatterns` simulations: an incident light wavelength of  $\lambda = 632nm$ , a CCD receptive field area of  $U \times V = 3mm \times 3mm$ , with the scatter centroid centered  $d = 5mm$  below the CCD plane. These dimensions are much as for the experimental apparatus used by Su *et al.*<sup>8</sup>, and can be seen to fulfill the XRD constraints presented in the previous section. Shown in Fig. 3.2, this setup gives a viewable side-scatter region between  $77.3^\circ$  and  $106.7^\circ$  in both the  $\theta$  and  $\phi$  axes (a solid angle corresponding to a cone of  $\simeq 30^\circ$ ).

Two point-spread functions  $M(\theta)$  were used: pure isotropic radiation (characteristic of mitochondria and other biospheres smaller than  $\lambda$ <sup>23</sup>) and a theoretical anisotropic point-spread function for a single spherical scatterer (characteristic of larger mitochondria in the Mie scattering regime, as described by the inset of Fig. 1 from Gourley *et al.*<sup>23</sup>, or from the polar plots of Fig. 2 of Gourley *et al.*<sup>9</sup>). Although the latter were easier to interpret, these



**Figure 3.3:** Two angular point-spread functions  $M(\theta)$  for individual scatterers: pure isotropic radiation (left, characteristic of  $100nm$  diameter mitochondria (much smaller than  $\lambda$ )), and anisotropic scatter (right, characteristic of  $500\text{--}1000nm$  diameter mitochondria)

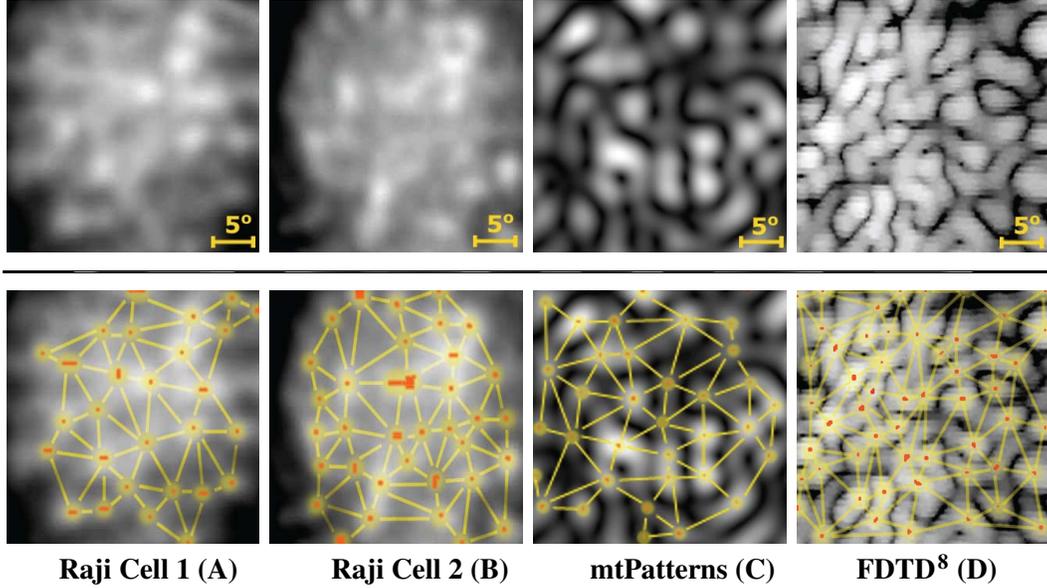
are difficult to extract data from and the behaviours should be taken as qualitatively correct rather than exact. The anisotropic case is from scattering for spherical scatterers in the Mie regime (size  $\geq \lambda$ , or  $\sim 500\text{--}1000nm$  in diameter), while the isotropic case is for scatterers in the Rayleigh or near-Rayleigh regime (size  $< \lambda$ , and up to  $\sim 500nm$ <sup>23</sup>). For the anisotropic cases,  $M(\theta)$  was used. These point-spread functions are shown in Fig. 3.3.

### 3.4 Results and Discussion

The following sections present both a qualitative and a quantitative analysis of the results of applying the `mtPatterns` algorithm, and show that a scattering simulation based only on the distribution of small mitochondria can effectively simulate wide-angle experimental images (recalling that mitochondria are the dominant sources of wide-angle scattering<sup>23</sup>). In XRD, analysis is often done by calculating the Fourier transform of potential crystal structures and comparing them to experimental scattering patterns. We show that a similar process is possible for scattering from mitochondria in a single cell; the following results show the possibility for predictive assessment of cell structure, in this case mitochondrial number and distribution, based on comparison with experimental scatter patterns.

#### 3.4.1 Comparison of `mtPatterns` and Experimental Cytometry Images

To evaluate the `mtPatterns` algorithm, `mtPatterns` results were compared to experimental wide-angle cytometry patterns captured using the method described by Su *et al.*<sup>8</sup> from the laser scattering of human Raji cells. Experimental Raji data was generated by one of the first published demonstrations of a microchip-based wide-angle cytometer<sup>8</sup>. All experimental data was scaled and cropped to the same side-scatter angular range as the



**Figure 3.4:** A comparison of feature spacing between the experimental cytometry data for two Raji cells (A and B,  $R_o \cong 8.0\mu m$ ,  $R_i \cong 4.0\mu m$ , each with an estimated hundreds of  $mt$ ), the `mtPatterns` algorithm (C,  $R_o = 8.0\mu m$ ,  $R_i = 4.0\mu m$ ,  $\#mt = 300$ ), and a re-processed version of the FDTD results of Su *et al.*'s Fig. 5(d)<sup>8</sup> (D,  $R_o = 8.0\mu m$ ,  $R_i = 4.0\mu m$ ,  $\#mt = 300$ ) over a  $30^\circ$  range in  $\phi$  and  $\theta$ , located in the side-scatter region. The point / line overlays in the bottom row indicate blob center points / spacing gaps used in blob spacing calculations. The horizontal axis corresponds to changes in  $\theta$ , vertical to changes in  $\phi$ .

`mtPatterns` data (from  $77.3^\circ$  to  $106.7^\circ$  in  $\theta$  and  $\phi$ , Sec. 3.3.6, selected to best show the side-scatter effects of the scatter distribution) and normalized to the same intensity and contrast levels. We also performed comparisons with a reprocessed version of the FDTD data published by Su *et al.*, describing a  $2mm$  by  $2mm$  CCD region with an angular range between  $\sim 60^\circ$  to  $90^\circ$  in  $\theta$  and  $\sim 77^\circ$  to  $107^\circ$  in  $\phi$  (their Fig. 5(d),<sup>8</sup> generated from a mitochondrial distribution with many of the same parameters as the `mtPatterns` test, but with a different random seed for organelle placement). FDTD data was cropped and normalized as described above. Estimated cell parameters from the experimental and FDTD data (as per Su *et al.*<sup>8</sup>) were used as input to `mtPatterns` to generate the simulations. This process allowed for direct visual and numerical comparison.

As we did not have access to the exact random placement of organelles used in these FDTD simulation studies, and precise organelle placement data cannot be obtained for experimental samples, a pixel-by-pixel comparison of scattering images was not feasible. Instead, the similarity between images may be visually judged by examining the shape and image structure of scattering images obtained from `mtPatterns` and from real experimental cytometers. Fig. 3.4, top, shows the experimental scattering signatures for two human Raji cells, taken using a miniaturized cytometer, with estimated cellular parameters of  $R_o \approx 8.0\mu m$ ,

$R_i \approx 4.0\mu m$ , and about 100 to 700  $mt$ , as per the experiments of Su *et al.*<sup>8</sup>. Qualitatively, these patterns are structurally similar—in terms of blob spacing and size (angular size, Fig. 3.4, top)—to the pattern generated by `mtPatterns` when initialized with the average parameters of a Raji cell estimated by Su *et al.* (Fig. 3.4(c), top;  $R_o = 8.0\mu m$ ,  $R_i = 4.0\mu m$ ,  $\#mt = 300$ ). The `mtPatterns` blob size and spacing also agreed with the re-processed version of Su *et al.*'s<sup>8</sup> FDTD pattern for a random distribution of 300 $mt$  ( $R_o = 8.0\mu m$ ,  $R_i = 4.0\mu m$ , no nuclear or cellular scattering, Fig. 3.4(d), top), though FDTD showed on average a slightly lower characteristic spacing and blob size than both experimental samples and `mtPatterns`.

Given the lack of organelle placement data for experimental samples, and the inherent complexity of wide-angle patterns, it is difficult to objectively compare scattering images in quantitative terms<sup>30</sup>. However, since the spacing of the scattering distribution relates to scattering pattern blob spacing (Eqs. (3.2), (3.3)), one effective way to numerically compare scattering images is to evaluate their characteristic angular spacing, defined as the average angular distance (in degrees) between neighboring scattering intensity maxima. This metric also can help guide the process of inferring experimental scattering structure from simulated images.

For the numerical comparisons presented here, angular spacing (in image space) between neighbouring maximum intensity regions (*i.e.* blob peaks) was measured as shown by the mesh in Fig. 3.4 and normalized to the angular range of each image to give a set of angular blob spacing values for each image. These spacing values were then averaged to compute each image's characteristic spacing value and its variability.

Using this metric, we compared the experimental images of Fig. 3.4 to the `mtPatterns` simulation generated using the corresponding parameters. This comparison can be seen qualitatively in the bottom panels of Fig. 3.4, and quantitatively as follows. As described above, `mtPatterns` was initialized with the estimated parameters of a Raji cell ( $R_o = 8.0\mu m$ ,  $R_i = 4.0\mu m$ ,  $\#mt = 300$ , as given by Su *et al.*<sup>8</sup>). With these parameters, we observed an average angular spacing (and standard deviation) between the maxima of intensity regions (*i.e.* blob centers) of  $5.51 \pm 1.51^\circ$  (73 samples; Fig. 3.4(c)). This compared well to the spacing values from the experimental cytometry images:  $5.12 \pm 1.47^\circ$  (85 samples; *Raji Cell 1*—Fig. 3.4(a)) and  $6.04 \pm 1.46^\circ$  (62 samples; *Raji Cell 2*—Fig. 3.4(b)). It also was in agreement with the re-processed FDTD results of Su *et al.*:  $4.98 \pm 1.63^\circ$  (139 samples; Fig. 3.4(d)).

As such, the `mtPatterns` characteristic spacing value was observed to be within the range of values from the two experimental cells and the FDTD simulation: spacing FDTD < *Raji Cell 1* < `mtPatterns` < *Raji Cell 2*. The slight variations in average characteristic spacing

between the two experimental samples are likely due to changes in internal cell structure (such as mitochondrial number and placement) and are well within one standard deviation. The smaller spacing in FDTD may also be due to modeling differences and the  $\theta - 17^\circ$  shift in the observed angular range<sup>8</sup>.

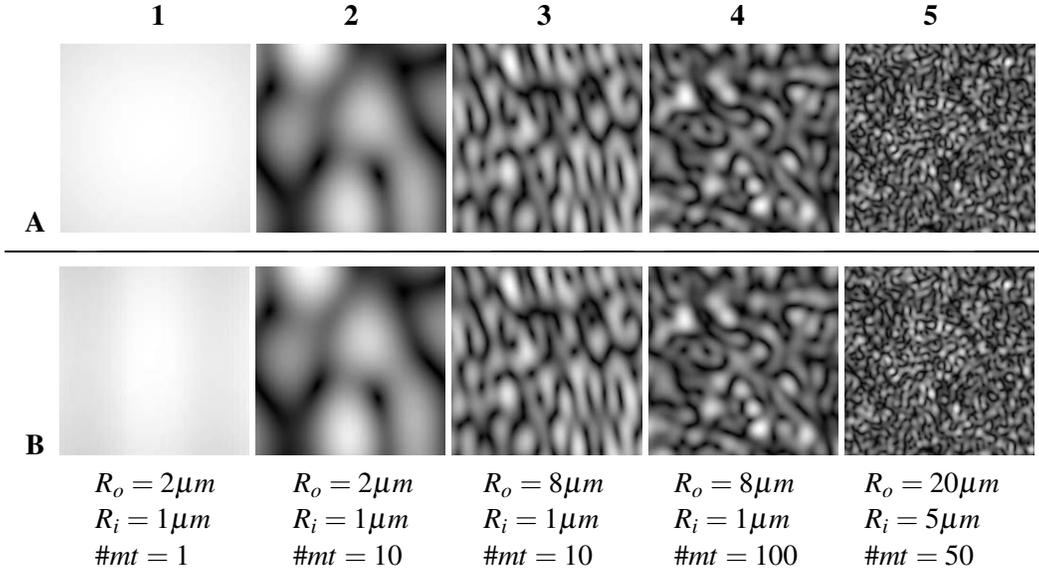
Lastly, we found that the experimental results (Fig. 3.4(a,b)) appeared to have more intensity toward the centre of the image. This background intensity could indicate the contributions of larger cell components and microstructures (*e.g.* the nucleus), light from the experimental setup, and/or the washed-out superposition of any non-uniform mitochondrial scattering (discussed below). As this background consists primarily of broad low-frequency features, it could potentially be separated from, and analysed separately from, the blob-like features. Feature extraction methods (*e.g.* the Cythe algorithm<sup>30</sup>, presented in Ch. 6) may be used to facilitate this.

The combination of the ability to calculate blob spacing and the ability to rapidly generate realistic simulated scattering plots with the `mtPatterns` algorithm facilitates the prediction of mitochondrial spacing directly from an observed scattering image. Once it is possible to quickly simulate realistic mitochondrial scattering patterns, simple visual and numerical comparisons could be used to perform real-time classification of cytometrically interrogated cells. As shown by XRD, great leaps in structural assessment can be made by comparing experimental and simulated scattering patterns.

### 3.4.2 Extension to Scattering from Larger Spherical Mitochondria

As discussed and shown above, at least for smaller mitochondria it appears that the scatter can be well-approximated by a model of isotropic scatter. However, the degree of scatter anisotropy becomes far more pronounced for larger mitochondria. We will now show a theoretical basis for extending our simulation approach to deal with arbitrarily large degrees of anisotropy from spherical mitochondria (*i.e.* much larger than  $500nm$ , when their dimensions exceed the wavelength of the incident light).

Up to now we have dealt with patterns with scaling  $M(\theta)$  due to  $\phi$ -invariant scatter from spherical mitochondria. More generally, the scatter of a mitochondrion specified by the vector  $\mathbf{S}$  in the reciprocal domain can be written as  $M(\mathbf{S})$ . Such a scattering pattern from an isolated (larger) mitochondrion,  $M(\mathbf{S})$ , may be calculated using numerical methods such as FDTD (as done by<sup>23</sup>), R-G, or captured in an experimental situation. This pattern could be the result of energy scattering from a complex sub-structure in the spatial domain, which we will denote  $m(\mathbf{r})$ .



**Figure 3.5:** A series of sample images generated using the `mtPatterns` algorithm. Images are shown for both uniform scattering (row A) and scattering using a complex point-spread function (row B, as per Fig. 3.3, right). Increased scatterer spacing can be seen to lead to a decrease in scattering pattern blob spacing (*e.g.* 2A, 3A). In terms of scattering blob placement and spacing, there is little difference between uniform (A) and non-uniform (B) mitochondrial scattering. To allow visual comparison, all images have been normalized with respect to the same minimum and maximum intensity values (*i.e.* 0–255), and cover the same solid angle ( $77.3^\circ$ – $106.7^\circ$  in  $\theta$  and  $\phi$ ). Forward scatter is toward the right.

From XRD, we know that the scattering pattern of an array of an identical and aligned sub-structures can be related to the scatter pattern of the sub-structure alone,  $m(\mathbf{r})$ , and that of the array alone,  $\rho(\mathbf{r})$ , through the convolution theorem. In the present case  $m(\mathbf{r})$  is the sub-structure within a single mitochondrion, while in XRD it is the arrangement of electrons in a single scattering body. This convolution takes the form:  $\rho(\mathbf{r}) * m(\mathbf{r})$ . The scattering pattern  $A_{true}(\mathbf{S})$  of the distribution of mitochondria with complex sub-structure will therefore be the Fourier transform of this convolution, Eq. (3.5), or, equivalently, the product of the Fourier transform of the array multiplied by the Fourier transform of the sub-structure, Eq. (3.6). (Although there are methods in XRD analysis for application to disordered crystals with misaligned substructures<sup>32</sup>, the XRD theory presented in this section is not able to account for misalignments (*i.e.*  $\phi_o$  and  $\theta_o$ ) in the sub-structure—this type of analysis will be discussed below.)

$$A_{true}(\mathbf{S}) = F\{\rho(\mathbf{r}) * m(\mathbf{r})\} \quad (3.5)$$

Thus, if  $A(\mathbf{S})$  is the transform of the distribution of mitochondria,  $\rho(\mathbf{r})$  (assumed to be isotropic scatterers), and  $M(\mathbf{S})$  is the transform of isotropic-scattering sub-structure,  $m(\mathbf{r})$ , within the mitochondria that gives rise to the observed anisotropic behaviour, then the true scatter pattern,  $A_{true}(\mathbf{S})$  is:

$$A_{true}(\mathbf{S}) = A(\mathbf{S})M(\mathbf{S}) \quad (3.6)$$

In the context of XRD, this is discussed in more detail in standard references (such as<sup>32,37</sup>). This approach may be applied to mitochondrial light scattering by noting that an array of mitochondrial positions  $\rho(\mathbf{r})$  may therefore be convolved with some internal sub-structure  $m(\mathbf{r})$  to account for the anisotropic scatter from the internal sub-structure. However, it is not important to know or guess at  $m(\mathbf{r})$ ; we need only obtain  $M(\mathbf{S})$ —the scatter pattern of the individual mitochondrion—from FDTD, R-G, or experiments. Since we know the final scattering pattern will be the multiplication of two Fourier transforms, Eq. (3.6), we can simply use Eq. (3.2) or (3.3) to compute the scattering pattern  $A(\mathbf{S})$  for a given distribution of isotropically scattering mitochondria  $\rho(\mathbf{r})$  and multiply the result by  $M(\mathbf{S})$ , to obtain the real mitochondrial scattering pattern,  $A_{true}(\mathbf{S})$ . This shows that the analysis of anisotropic scatter from larger spherical mitochondria can be done in a way that remains consistent with XRD theory (and hence toolsets).

As shown in Fig. 3.5, there was little difference between `mtPatterns` scattering images created using isotropic scatterers (Fig. 3.5(a)) and scatterers with the more complex point-spread functions characteristic of larger mitochondria (Fig. 3.5(b)). Our experiments showed little structural variation (in terms of scattering blob size, spacing, and intensity) between large-angle scattering patterns generated using a uniform point-spread function (Fig. 3.3, left) and those generated by radiators using the point-spread function described by Gourley *et al.* (Fig. 3.3, right). This was expected, as the theoretical point-spread function of a non-isotropic mitochondrion is almost identical to an isotropic point-spread function when observed over a solid angle equivalent to a cone approximately 30° across, centered on the side-scatter region. As such, the use of isotropic scattering points to simulate mitochondria proved experimentally to be a valid first-order approximation in the large-angle scattering domain. In other words, it appears that small variations in scattering intensity with angle do not affect the overall pattern—they are readily ‘washed out’. However, if the degree of anisotropy from a spherical scatterer were large enough that we needed to account for it, the above approach would suffice (through the use of our algorithm or by means of other XRD-like methods).

### 3.4.3 Extension to Scattering from Non-Spherical Mitochondria

In the above, anisotropic scatter was considered to originate from larger spherical mitochondria. Another source of anisotropy is from non-spherical mitochondria. Although much of the work in the literature has considered the mitochondrion to be an ellipsoid or a sphere, the structure of the mitochondrion is not yet firmly established. Assuming that the mitochondria are not physically correlated in part of a larger structure, we would expect orientations that are random. As such, we would expect less of an overall effect than seen in the case of the larger spherical mitochondria where the anisotropy was a function of  $\theta$  (and hence correlated). Gourley *et al.*<sup>23</sup> have made a similar suggestion. On the other hand, recent reports suggest that the mitochondria do form an interconnected network, depending on the needs, type and state of the cell<sup>38,39</sup>. A further exploration of this is underway, and will require extensive simulation and an examination of real cells in a variety of states. Much as above, if the degree of anisotropy from a non-spherical scatterer were large enough (or correlated enough) that we needed to account for it, our algorithm would be able to do so. Since doing so would introduce a large number of new variables ( $\phi_o$  and  $\theta_o$  for each mitochondrion), it would seem unwise to attempt to do this until the need has clearly been demonstrated.

### 3.4.4 Observations on the Effect of Scatter Distribution and Interaction

Fig. 3.5 also shows that an increase in the number of scatterers lead to increased image complexity (*i.e.* smaller blobs and smaller characteristic spacing). In addition, an increased scatterer distribution radius  $R_o$  also led to increased image complexity. A tightly arranged mitochondrial distribution (*e.g.* smaller mitochondrial volume, such as Fig. 3.5(a2)) resulted in larger homogeneous intensity regions with greater characteristic spacing, while wider distributions (such as Fig. 3.5(a5)) generated a number of small, tightly spaced intensity regions.

This behaviour is expected from Fourier theory and X-ray diffraction literature, as described above. As in Fourier theory, the spacing between, and size of, intensity regions on the simulated CCD region were inversely proportional to the spacing of the original organelle distribution. The more dense the scattering distribution, the more distance between blobs in the scattering plot (Fig. 3.5). These trends will be useful in future feature-based classification systems, as image feature complexity can help deduce initial cell structure<sup>30</sup>, whether by direct comparison of features or phase reconstruction techniques.

Lastly, when the distribution of scattering points was tightly packed within a volume much smaller than  $\lambda$ , we found that the resulting scattering pattern closely matched that of a

single scatterer. Under these conditions, scattering contributions from individual scatterers washed out to form a single bright intensity region. We found this to be true for any number of scatterers. However, the un-normalized intensity of the region increased in proportion to the number of scatters ( $\#mt$ ). This result is expected from scattering theory, and has been experimentally corroborated by Gourley *et al.*<sup>23</sup>; as  $R_o \ll \lambda$ , scattered phases will align and all scattering contributions will be uniformly constructive regardless of scatterer number. This is also analogous to the scattering for a random array of apertures, shown in Fig. 10.33 of Hecht<sup>31</sup>. As such, the un-normalized average intensity of a signature might prove to be a useful predictor of the raw number of mitochondria within a experimental sample; this would allow the classification of both the mitochondrial spacing (through blob spacing, described above) and number (through un-normalized blob intensity) of scattering mitochondria, all from a single scattering plot.

### 3.5 Conclusions

In the limit of small mitochondria, we show that large-angle light scattering of mitochondria may be treated using an adaptation of classical X-ray diffraction theory, giving rise to a computationally efficient solution to the scattering simulation problem. In addition, we show that this approach may be expanded to the general case of larger mitochondria which may or may not scatter in an isotropic fashion. This result facilitates the rapid simulation of realistic mitochondrial scattering patterns—without the need for computationally challenging or restrictive routines—and allows the determination of some aspects of cell structure directly from experimental scattering patterns. Large-angle mitochondrial scattering signatures can be calculated in seconds, allowing the rapid creation of large predictive image databases and facilitating great advances in real-time patient care, cell structure prediction, and cell morphology research. In the present work, the computation time of the average `mtPatterns` simulation was between a fraction of a second and a minute on a standard personal laptop computer running the Python interpreter.

Through rapid simulation, the `mtPatterns` algorithm allows the prediction mitochondrial spacing (via scattering pattern blob spacing) and number (via scattering pattern intensity), directly from an observed scattering image. In addition, given the dramatically different scattering contributions of cellular micro- and nano-structures, we show that it is possible to isolate (using, for example, feature extraction methods<sup>30</sup>) just the high-frequency scattering contributions of nanostructures by removing—or separately analyzing for useful diagnostic information—the broad band-like scattering of larger microstructures.

This analogy to XRD and ability to inexpensively generate plots that contain the key structurally relevant features of real scattering signatures has not previously been demonstrated in the literature, and makes `mtPatterns` an effective tool for future cytometry research. Given that defects associated with mitochondria are implicated not only in carcinogenesis, but also with a number of neurodegenerative disorders<sup>38</sup>, the use of this cytometric tool has the potential to improve our understanding and detection of these diseases.

The `mtPatterns` simulation algorithm has been validated against experimental data and Fourier theory. Given its similarity to experimental patterns, and its ability to rapidly generate scattering signatures, we will use it as the primary source of data for the remainder of this dissertation—it is applied to generate the image data used in the following chapters.

## References

- [1] K. Singh, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, “Analysis of cellular structure by light scattering measurements in a new cytometer design based on a liquid-core waveguide,” *IEE Proc.-Nanobiotechnology* **151**(1), 10–16 (2004).
- [2] K. Singh, X. Su, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, “A miniaturised wide-angle 2-D cytometer,” *Cytometry A* **69A**, 307–315 (2006).
- [3] T. Proffen and R. B. Neder, “DISCUS: A program for diffuse scattering and defect-structure simulation,” *Journal of Applied Crystallography* **30**, 171–175 (1997).
- [4] P. L. Gourley, “Biocavity laser for high-speed cell and tumour biology,” *Journal of Physics D: Applied Physics* **36**(14), R228–R239 (2003).
- [5] V. P. Maltsev, “Scanning flow cytometry for individual particle analysis,” *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [6] K. A. Sem’yanov, P. A. Tarasov, J. T. Soini, A. K. Petrov, and V. P. Maltsev, “Calibration-free method to determine the size and hemoglobin concentration of individual red blood cells from light scattering,” *Applied Optics* **39**(31), 5884–5889 (2000).
- [7] Z. Ulanowski, Z. Wang, P. H. Kaye, and I. K. Ludlow, “Application of neural networks to the inverse scattering problem for spheres,” *Applied Optics* **37**(18), 4027–4033 (1998).
- [8] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, “2D light scattering patterns of mitochondria in single cells,” *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [9] P. L. Gourley and R. K. Naviaux, “Optical phenotyping of human mitochondria in a biocavity laser,” *IEEE Journal of Selected Topics in Quantum Electronics* **11**(4), 818–826 (2005).

- [10] D. C. Wallace, "Mitochondrial diseases in man and mouse," *Science* **283**(5407), 1482–1488 (1999).
- [11] D. C. Wallace, "Diseases of the mitochondrial dna," *Annual Review of Biochemistry* **61**, 1175–1212 (1992).
- [12] M. Brandon, P. Baldi, and D. C. Wallace, "Mitochondrial mutations in cancer," *Oncogene* **25**(34), 4647–4662 (2006).
- [13] R. Drezek, A. Dunn, and R. Richards-Kortum, "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements," *Applied Optics* **38**(16), 3651–3661 (1999).
- [14] R. Drezek, A. Dunn, and R. Richards-Kortum, "A pulsed finite-difference time-domain (FDTD) method for calculating light scattering from biological cells over broad wavelength ranges," *Optics Express* **6**(7), 147–157 (2000).
- [15] C. Liu, C. E. Capjack, and W. Rozmus, "3-D simulation of light scattering from biological cells and cell differentiation," *Journal of Biomedical Optics* **10**(1), 014007 (12 pages) (2005).
- [16] N. Ghosh, P. Buddhiwant, A. Uppal, K. Majumder, H. S. Patel, and P. K. Gupta, "Simultaneous determination of size and refractive index of red blood cells by light scattering measurements," *Applied Physics Letters* **88**, 084,101 (3 pages) (2006).
- [17] Y. L. Kim, Y. Liu, R. K. Wali, H. K. Roy, M. J. Goldberg, A. K. Kromin, K. Chen, and V. Backman, "Simultaneous measurement of angular and spectral properties of light scattering for characterization of tissue microarchitecture and its alteration in early precancer," *IEEE Journal of Selected Topics in Quantum Electronics* **9**(2), 243–256 (2003).
- [18] H. Fang, M. Ollero, E. Vitkin, L. M. Kimerer, P. B. Cipolloni, M. M. Zaman, S. D. Freedman, I. J. Bigio, I. Itzkan, E. B. Hanlon, and L. T. Perelman, "Noninvasive sizing of subcellular organelles with light scattering spectroscopy," *IEEE Journal of Selected Topics in Quantum Electronics* **9**(2), 267–276 (2003).
- [19] A. Katz, A. Alimova, M. Xu, E. Rudolph, M. K. Shah, H. E. Savage, R. B. Rosen, S. A. McCormick, and R. R. Alfano, "Bacteria size determination by elastic light scattering," *IEEE Journal of Selected Topics in Quantum Electronics* **9**(2), 277–287 (2003).
- [20] J. D. Wilson, C. E. Bigelow, D. J. Calkins, and T. H. Foster, "Light scattering from intact cells reports oxidative-stress-induced mitochondrial swelling," *Biophysical Journal* **88**(4), 2929–2938 (2005).
- [21] J. D. Wilson, W. J. Cottrell, and T. H. Foster, "Index-of-refraction-dependent subcellular light scattering observed with organelle-specific dyes," *Journal of Biomedical Optics* **12**(1), 014010 (10 pages) (2007).
- [22] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1998).

- [23] P. L. Gourley, J. K. Hendricks, A. E. McDonald, R. G. Copeland, K. E. Barrett, C. R. Gourley, and R. K. Naviaux, “Ultrafast nanolaser flow device for detecting cancer in single cells,” *Biomedical Microdevices* **7**(4), 331–339 (2005).
- [24] J. S. Modica-Napolitano and K. Singh, “Mitochondria as targets for detection and treatment of cancer,” *Expert Reviews in Molecular Medicine* **4**(9), 1–19 (2004).
- [25] L. J. Garcia-Rodriguez, *Mitochondria*, vol. 80 of *Methods in Cell Biology*, chap. Appendix 1. Basic properties of mitochondria., pp. 809–812, 2nd ed. (Academic Press, San Diego, 2007).
- [26] I. E. Scheffler, *Mitochondria* (John Wiley & Sons, New York, 1999).
- [27] 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,” *Applied Optics* **37**, 3586–3593 (1998).
- [28] B. Shao, J. S. Jaffe, M. Chachisvilis, and S. C. Esener, “Angular resolved light scattering for discriminating among marine picoplankton: modeling and experimental measurements,” *Optics Express* **14**(25), 12,473–12,484 (2006).
- [29] A. K. Dunn, “Light scattering properties of cells,” Ph.D. thesis, University of Texas at Austin (1997).
- [30] P. M. Pilarski and C. J. Backhouse, “A method for cytometric image parameterization,” *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.
- [31] E. Hecht, *Optics*, 4th ed. (Addison Wesley, San Francisco, 2002).
- [32] N. Kasai and M. Kakudo, *X-Ray Diffraction by Macromolecules*, Springer series in chemical physics, 80 0172-6218 (Springer, New York, 2005).
- [33] P. Thibault, V. Elser, C. Jacobsen, D. Shapiro, and D. Sayre, “Reconstruction of a yeast cell from X-ray diffraction data,” *Acta Crystallographica, Section A: Foundations of Crystallography* **62**, 248–261 (2006).
- [34] J. M. Zuo, I. Vartanyants, M. Gao, R. Zhang, and L. A. Nagahara, “Atomic resolution imaging of a carbon nanotube from diffraction intensities,” *Science* **300**(5624), 1419–1421 (2003).
- [35] D. Shapiro, P. Thibault, T. Beetz, V. Elser, M. Howells, C. Jacobsen, J. Kirz, E. Lima, H. Miao, A. M. Neiman, and D. Sayre, “Biological imaging by soft X-ray diffraction microscopy,” *Proceedings of the National Academy of Sciences of the United States of America* **102**(43), 15,343–15,346 (2005).
- [36] T. Proffen and R. B. Neder, “DISCUS: A program for diffuse scattering and defect-structure simulation - update,” *Journal of Applied Crystallography* **32**, 838–839 (1999).
- [37] R. B. Neder and T. Proffen, “Teaching diffraction with the aid of computer simulations,” *Journal of Applied Crystallography* **29**, 727–735 (1996).

- [38] S. Hoppins, L. Lackner, and J. Nunnari, “The machines that divide and fuse mitochondria,” *Annual Review of Biochemistry* **76**, 751–780 (2007).
- [39] R. Rizzuto, P. Pinton, W. Carrington, F. S. Fay, K. E. Fogarty, L. M. Lifshitz, R. A. Tuft, and T. Pozzan, “Close contacts with the endoplasmic reticulum as determinants of mitochondrial Ca<sup>2+</sup> responses,” *Science* **280**, 1763–1766 (1998).

## Chapter 4

# Theoretical Foundations for Robust Wide-Angle Scattering Pattern Analysis

### 4.1 Introduction

Studies have shown that important structural information is contained within measurements of angular light scattering intensity from a single cell. This information includes the size, shape, and distribution of cellular micro- and nano-structures<sup>1-4</sup>. Optical simulation is an important tool for understanding these scattering relationships, and several techniques are available to simulate the scattering of light through single cells—*e.g.* Mie theory<sup>1,5</sup>, Finite Difference Time Domain (FDTD) simulation<sup>1,3,6</sup>, and Rayleigh-Gans theory<sup>5</sup>. Each method is governed by different computational and physical constraints<sup>3</sup>.

In the previous chapter, we presented a computationally inexpensive new simulation method, the `mtPatterns` algorithm<sup>7</sup>, to generate realistic two-dimensional (2D) scattering signatures for organelle populations within a single cell. We found that the patterns generated by this method were in agreement with published experimental and simulated (FDTD) results. Validations were performed using standard image comparison metrics—numerical measures for the size, spacing, and shape of intensity regions within images<sup>4,7</sup>.

We now use this simulation method to examine the effect of phase shifts and optical changes on the scattering patterns of organelle populations. In particular, we compare two instances of the `mtPatterns` algorithm: one with the correct representation of the distance  $d_n^o$  from a scatterer to the light source (“corrected”, Eq. 4.1), and one where all organelles start with exactly the same phase value (“uncorrected”, Eq. 4.2).

---

<sup>1</sup>A version of this chapter has been submitted for publication. Pilarski and Backhouse. 2009. *Optics Express*.

$$A(\mathbf{s}') = \sum_{n=1}^N \{ \cos[2\pi \cdot (|\mathbf{s}' - \mathbf{r}_n| + d_n^o)] - i \sin[2\pi \cdot (|\mathbf{s}' - \mathbf{r}_n| + d_n^o)] \} \quad (4.1)$$

$$A(\mathbf{s}') = \sum_{n=1}^N \{ \cos[2\pi \cdot (|\mathbf{s}' - \mathbf{r}_n| + 0)] - i \sin[2\pi \cdot (|\mathbf{s}' - \mathbf{r}_n| + 0)] \} \quad (4.2)$$

Comparison of the uncorrected and corrected forms of the `mtPatterns` algorithm presented an interesting result: while a direct pixel-by-pixel comparison of images showed differences between corrected and uncorrected patterns, the effect of the non-uniform phase shift was largely invisible to our visual validation methods, and also to a number of standard image processing metrics used for biomedical data analysis and comparison<sup>4,8</sup>. Texture-based validation and processing methods—*e.g.* methods that characterize intensity region size, spacing, orientation, and shape—were sensitive to medically relevant changes in intracellular structure, but robust to random organelle shifts and noise (such as the uncorrected phase term).

In this chapter, we present a discussion of the underlying optical constraints that govern the robust interpretation of 2D cellular scattering patterns, and examine how these constraints affect the diagnostic potential of wide-angle cytometry systems. The result is an important observation about the effects of organelle distribution shape and size on key properties of cellular scattering patterns.

## 4.2 Background

Robust scattering image characterization has important ramifications for the fields of wide-angle scattering pattern analysis and cytometry, and has clinical relevance<sup>4</sup>. For example, methods to consistently identify the distribution of mitochondria (metabolically related organelles) within a cell could help predict a number of disorders, including tumor development and the chemotherapy response of breast and lung cancer patients<sup>9–11</sup>.

Unfortunately, diagnostic methods that base their predictions on angular light-scattering data are susceptible to small variations in cellular geometry. Similar scattering bodies may generate different scattering signatures due to slight rotations, shifts in their optical properties, and/or changes to the exact placement of their internal components<sup>5,6</sup>. The exact intensity at any given point in two scattering signatures (from nearly identical scatterers) can be significantly different. As such, many diagnostic methods that examine direct comparisons between images are not robust to normal levels of experimental variation, and

finding suitable analysis techniques with both predictive power and suitable invariance is a challenging problem<sup>6</sup>.

This image comparison problem is not unique to light scattering—chronic variability between two instances of the same, or a similar, biological model (*e.g.* two mammograms captured at different times or angles) is one of the reasons shape and texture based image analysis metrics are employed in many biomedical analysis systems, as opposed to direct pixel-by-pixel image comparison<sup>8</sup>. While direct image comparisons are often a poor basis for meaningful predictions and classifications, other image metrics have been used to successfully relate and compare samples even in the presence of experimental noise and/or using data from different patients<sup>8</sup>. This fact is especially significant for 2D scattering pattern analysis, as cellular geometry, and thus scattering, is unlikely to be constant between different cells. The interpretation and comparison of scattering patterns is primarily a question of image analysis and pattern recognition, and therefore hinges on standard tools and metrics developed in these fields.

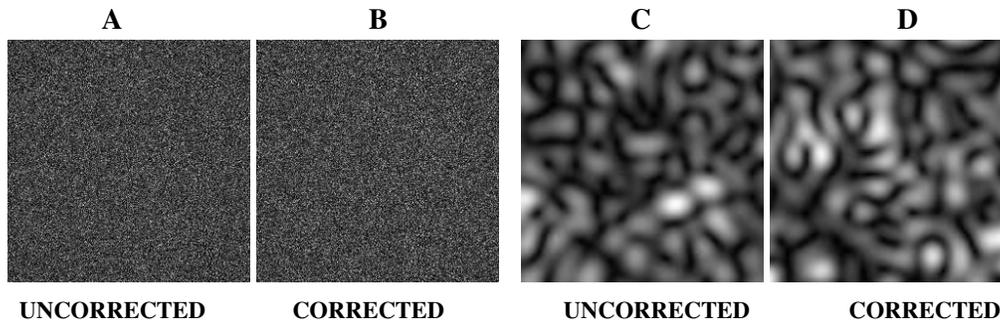
### 4.3 Comparison of Corrected and Uncorrected Patterns

To characterize the difference between corrected and uncorrected simulation results, we note that the modified path difference in the uncorrected formulation is equivalent to introducing an additional phase shift  $\tau(\mathbf{r}_n)$  that is dependent on the exact position of each scatterer. In terms of the Fourier transform, scattering can be represented as:

$$F(\mathbf{S}) = \mathcal{F}\{\rho(\mathbf{r}_n)\exp[2\pi i\mathbf{S} \cdot \tau(\mathbf{r}_n)]\} \quad (4.3)$$

where  $\rho(\mathbf{r}_n)$  defines the scatterer placement and  $\tau(\mathbf{r}_n)$  is a phase shift from  $(-\pi, \pi]$  that is a function of the distance between a scattering point  $\mathbf{r}_n$  and the light source or reference plane. For all but the simplest cases, it is extremely difficult (or impossible) to find an analytical solution to the Fourier transform of non-uniform phase shifts<sup>12</sup>. However, more general relationships have been identified. As shown by McClain and Gregory, a *phase-only shift* to points in the scatterer distribution relates to a corresponding *positional shift* to intensity peaks in the Fourier (scattering) plane<sup>12</sup>. An example of this can be seen experimentally in Fig. 4.1. While the location of intensity peaks in the image changes between the phase-shifted and non-phase-shifted populations, peak breadth and spacing appear to remain relatively static.

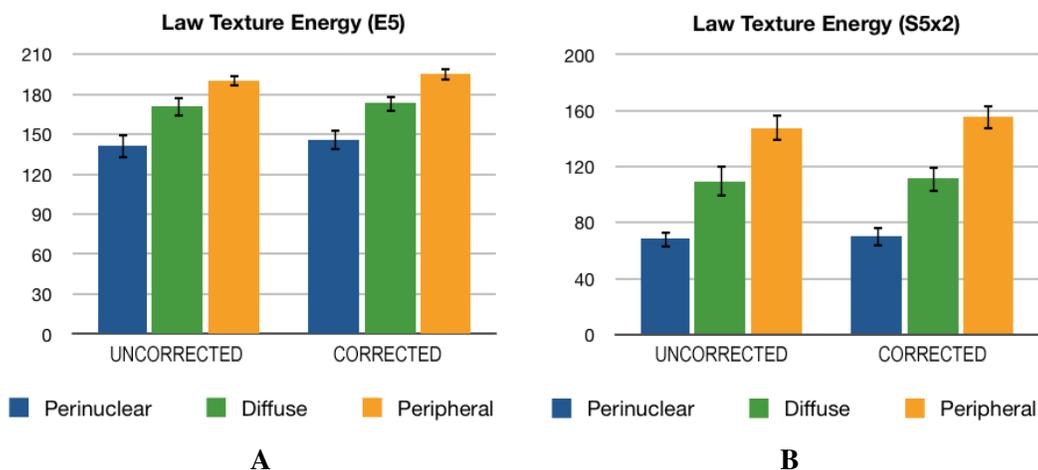
With this information as a background, we performed a detailed comparison of scattering images generated using both the corrected and uncorrected forms of the `mtPatterns` algo-



**Figure 4.1:** LEFT: The effect of a phase perturbation on a population’s scattering pattern, using a very wide distribution aperture (minimal shape effects): scattering using uncorrected phase values (A), and scattering using corrected phase values (B) for 1000 scatterers, placed randomly in a  $1500\mu\text{m}$  radius sphere, using the simulation conditions described in Pilarski *et al.*<sup>4</sup>. Peak breadth can be seen to remain small, and does not change noticeably due to phase perturbation. RIGHT: Comparison of the wide-angle scattering patterns for a mitochondrial distribution simulated by the `mtPatterns` algorithm, using the uncorrected equation (C) and corrected equation (D) for 300 scatterers, placed randomly in a diffuse cellular distribution as described by Pilarski *et al.*<sup>4</sup>.

rithm. As noted in Sec. 4.2, the comparison of scattering images is primarily a pattern analysis problem, best treated with standard tools for differentiating and classifying biomedical images. As such, we compared corrected and uncorrected `mtPatterns` scattering images in terms of twenty-one representative image descriptors—statistical metrics, Haralick texture measures, and Law texture energy metrics<sup>4,8</sup>. As described by Rangayyan, these descriptors quantify a wide range of statistical, local, and regional image characteristics<sup>8</sup>. Together they provide a unique “fingerprint” for each image, characterizing intensity peak breadth, shape, arrangement, and size, as well as general properties like image contrast, entropy, and homogeneity.

Comparing the corrected and uncorrected `mtPatterns` images in Fig. 4.1 (C and D), we found seventeen of the twenty-one image descriptor values were less than one standard deviation apart (*i.e.* no statistical difference between images). Experiments were then done on groups of fifteen scattering images from three different organelle distributions (perinuclear, diffuse, and peripheral, as in Pilarski *et al.*<sup>4</sup>), using both Eqs. 4.1 and 4.2. No statistical difference was found between corrected and uncorrected image groups for the majority of the twenty-one image metrics. An example of this is shown in Fig. 4.2 for two of the image metrics used in this study; these particular features represent the levels of edge and spot texture in an image. In summary, we found that the exact position of intensity peaks within a scattering pattern changed as a result of an arbitrary phase shift or organelle population perturbation—*e.g.* the random repositioning of organelles or use of the uncorrected Eq. 4.2—but the average breadth, spacing, and shape of intensity peaks did not. This is shown in qualitatively in Fig. 4.1, and numerically in Fig. 4.2.



**Figure 4.2:** Comparison of scattering pattern databases in terms of standard image descriptors representing the edge (A) and spot (B) content in scattering images. Tests were performed for three different distribution types, with fifteen samples per distribution, per trial.

## 4.4 Discussion of Robust Behaviour

While surprising, the results in Fig. 4.1 and Fig. 4.2 agree with expectations from X-ray diffraction theory and Fourier optics. As described by Kasai and Kakudo, when the size of a scattering distribution is relatively small with respect to the spacing between scatterers, the breadth and shape of intensity peaks in the scattering plane is dominated by the size and shape of the distribution (its *shape factor*)<sup>13</sup>. Should the spacing between scatterers be much smaller than the size of the distribution (for instance, the case of a very large or infinitely wide crystal), peak breadth is instead determined by the relationship between scatterers<sup>13</sup>. This fact is used in crystallography to help determine the size and shape of crystals, paracrystals, and amorphous scattering bodies<sup>13</sup>. A detailed treatment is provided in Kasai and Kakudo, Ch. 13<sup>13</sup>.

For the case of scattering from organelle populations similar to the mitochondria in a human cell, we have the former case: the spacing between organelles is only one to two orders of magnitude less than the width of the cell itself. As such, we expect the breadth and shape of intensity peaks in the scattering pattern to be determined by the size of the distribution, as opposed to the spacing and arrangement of scatterers (*i.e.* be “diffraction limited” by the shape and size of the aperture function or *shape factor*). As phase perturbations in the scatterer domain lead to positional shifts in the Fourier domain (as per McClain and Gregory<sup>12</sup>), we expect their impact on peak size and shape to be negligible compared to the broadening effect of the shape factor. Not surprisingly then, image analysis techniques based on texture will be largely unaffected by these changes.

These observations, paired with the results presented in Sec. 4.3 (in particular Fig. 4.2), also suggest that changes to the number/concentration of scatterers will have minimal impact on peak breadth and shape. The shift in scatterer spacing due to realistic concentration changes should not be large enough to overcome the broadening due to distribution shape. This is consistent with recently published experimental results, where organelle concentration was found to be poorly correlated to scattering peak size and shape<sup>4</sup>. It also explains why the non-uniform phase shift in the uncorrected Eq. 4.2 does not have a noticeable impact on validation metrics or texture-based processing routines. These measures are based on the size and spacing of intensity regions—things that, for realistically constrained examples of the cellular scattering case, appear to be determined by the size and shape of the scattering distribution and not the exact placement or optical alignment of scatterers.

## 4.5 Conclusions

As demonstrated in this chapter, the breadth and shape of intensity peaks in wide-angle scattering patterns is dominated by a scattering distribution's *shape factor* (*i.e.* 3D distribution shape and size), as opposed to any broadening due to phase variation or scatterer placement.

These observations are important for cellular scattering pattern analysis. It appears that—at least in terms of scattering peak geometry—the shape factor of a scattering distribution will overwhelm changes due to organelle placement and concentration. As such, metrics based on image texture, specifically the shape and size of intensity peaks, seem resilient to randomness and variability that challenged previous methods. At the same time, they remain sensitive to physical changes of medical interest—*e.g.* the distribution of organelles like the mitochondria.

This points to a robust set of tools for analyzing and characterizing organelle distributions based on wide-angle cellular scattering patterns. However, it also indicates that it may be difficult or impossible to determine exact scatterer position and concentration from a wide-angle scattering signature using only the shape, size, and spacing of intensity peaks. Future work will describe methods to determine both distribution shape and scatterer concentration from a single scattering image.

## References

- [1] A. K. Dunn, "Light scattering properties of cells," Ph.D. thesis, University of Texas at Austin (1997).
- [2] V. P. Maltsev, "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [3] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, "2D light scattering patterns of mitochondria in single cells," *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [4] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, "Computational analysis of mitochondrial placement and aggregation effects on wide-angle cell scattering patterns," in *Proceedings of SPIE*, A. Wax and V. Backman, eds., vol. 7187, p. 71870J (12 pages) (2009). URL <http://dx.doi.org/10.1117/12.809730>.
- [5] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1998).
- [6] C. Liu, C. E. Capjack, and W. Rozmus, "3-D simulation of light scattering from biological cells and cell differentiation," *Journal of Biomedical Optics* **10**(1), 014,007 (12 pages) (2005).
- [7] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, "Rapid simulation of wide-angle scattering from mitochondria in single cells," *Optics Express* **16**(17), 12,819–12,834 (2008). URL <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-17-12819>.
- [8] R. M. Rangayyan, *Biomedical Image Analysis* (CRC Press, Boca Raton, 2004).
- [9] S. Sikder, J. M. G. Reyes, C. S. Moon, O. Suwan-apichon, J. H. Elisseeff, and R. S. Chuck, "Noninvasive mitochondrial imaging in live cell culture," *Photochemistry and Photobiology* **81**(6), 1569–1571 (2005).
- [10] A. M. Villa and S. M. Doglia, "Mitochondria in tumor cells studied by laser scanning confocal microscopy," *Journal of Biomedical Optics* **9**(2), 385–394 (2004). URL <http://link.aip.org/link/?JB0/9/385/1>.
- [11] D. C. Wallace, "Mitochondrial diseases in man and mouse," *Science* **283**(5407), 1482–1488 (1999).
- [12] J. L. McClain Jr. and D. A. Gregory, "Fourier transforms of phase objects and implications for optical correlators," *Microwave and Optical Technology Letters* **25**(6), 406–412 (2000).
- [13] N. Kasai and M. Kakudo, *X-Ray Diffraction by Macromolecules*, Springer series in chemical physics, 80 0172-6218 (Springer, New York, 2005).

## Chapter 5

# Pattern Analysis Techniques for Wide-Angle Light Scattering Images

### 5.1 Introduction

Image analysis and pattern recognition are key elements of biomedical analysis schemes<sup>1</sup>. In a number of cases, the decomposition of an experimental image into a set of representative features and attributes allows biomedical data to be quickly analyzed, classified, and used to supplement clinical diagnostic practices<sup>1,2</sup>. One relevant new domain for this kind of analysis is the prediction of organelle arrangement within a cell from a sample of its wide-angle light scattering pattern—a clinically important problem<sup>3,4</sup>. The development of robust techniques to identify the shape and size of organelle distributions from light scattering measurements would facilitate the rapid identification and treatment of diseases such as cancer and neurodegenerative disorders.

A number of groups have demonstrated that pattern analysis methods can be used to identify relationships between light scattering patterns and aspects of cellular structure. However, the exact relationship between wide-angle patterns and cellular organelle distributions has yet to be explored in detail, and prior to this work there were no techniques to consistently relate 2D light scattering to organelle arrangement within a cell.

In this chapter, we demonstrate clear correlations between the arrangement (distribution) of organelles within a cell and the composition of wide-angle scattering patterns, identifying scattering trends that are independent of cellular rotation, organelle concentration, and the inherent randomness of organelle position within a cell. In addition, we show qualitative and quantitative descriptions of the identified scattering trends, and present an example of their potential for use in automated cell classification.

---

<sup>1</sup>A version of this chapter has been submitted for publication. Pilarski et al. 2009. *Journal of Biomedical Optics*. A version of this chapter has been published. Pilarski et al. 2009. *Proc. of SPIE*. 7187: 71870J–12.

## 5.2 Background

Analysis of 2D wide-angle scattering patterns is a challenging problem. Patterns take the form of complex intensity plots, and small changes to the organization, orientation, shape, and optical properties of scatterers and scattering populations can significantly alter their resulting patterns<sup>5,6</sup>. It is therefore difficult to find methods that can identify the scattering contributions of medically relevant cellular properties (*e.g.* organelle distribution) while remaining robust to experimental noise and sample to sample differences.

This is a common problem in many biomedical analysis problems, where images from different samples—or even two images taken from the same sample population—can be very different in terms of image composition, even after image alignment (*i.e.* registration)<sup>1</sup>. Pixel-by-pixel comparison of images often fails to give an effective metric of the similarity or difference between samples<sup>1</sup>. Given this, it is important to find features or attributes of an image that are common to all examples of a class while staying robust to normal variations between samples.

Like other biomedical problems, the solution lies in more advanced image analysis and pattern recognition methods. In domains where direct image comparison often leads to poor predictions and classifications, image attributes have been identified and successfully used to provide a more robust numerical representation of image content<sup>1,7</sup>. Routines have been developed to extract shape, texture, and statistics-based attributes from experimental images. Taken as a group, these attributes provide a unique signature or “fingerprint” for each image<sup>1,7</sup>. Attributes and groups of attributes are then evaluated for medical significance, usually through the use of additional attribute selection and classification routines<sup>1</sup>.

Within the set of possible image attributes, texture attributes have the advantage that they are largely objective with respect to the determination of region boundaries and ground truth values<sup>1</sup>; as such, they provide a viable platform for examining the composition of complex light scattering images. Some examples of widely used texture metrics include *Law texture energy*, *Haralick texture measures*, and *statistical texture measures*<sup>1,7</sup>. These metrics provide a compact and unbiased representation of both global and local intensity features in a sample image.

Each texture metric uses a different method to interpret image data. Statistical texture measures evaluate the distribution of an image’s intensity histogram in terms of standard statistical features: variance, skewness, and kurtosis; they do not take into account any spatial information<sup>1</sup>. By contrast, Haralick texture metrics (*e.g.* H1–H11) are based on local image relationships stored in a relational table known as a *grey-level correlation matrix* (GLCM)<sup>1</sup>.

This matrix is a measure of how likely a pixel of a given intensity level will be next to—at a given distance and angle—a pixel of another intensity level (computed for all different intensity levels). After forming a GLCM for each image, a number of Haralick measures are computed to describe the level of homogeneity, correlation, variance, consistency, information content, and entropy in a scattering image.

Law texture measures give a more qualitative view of image texture. They are based on the convolution of a filter kernel (usually  $5 \times 5$ ) with the target image, and are computationally less demanding than Haralick metrics. The summation of intensity (energy) in the filtered image can be used to evaluate the relative strength of different feature types within an image—*i.e.* a texture's edge content (E5), spot content (S5), wave content (W5), and ripple content (R5). A more detailed discussion of statistical, Haralick, and Law texture measures and their meaning can be found in Rangayyan 2004<sup>1</sup>.

Data processing of this type is usually followed by attribute selection<sup>2</sup>. Attribute selection is the process of determining which of a number of attributes (in this case texture measures) contain the most information about the class or label of a data point. There are a number of different attribute selection techniques, each with their own areas of specialty and modes of operation<sup>8</sup>. As such, it is common to use a suite of different attribute selection techniques to allow for better identification of relevant features—attributes indicated by the majority of selection techniques are likely to hold predictive information about the class of a sample. Machine learning and classification methods can then be used on a subset of these selected attributes to rapidly and accurately classify new experimental data<sup>1</sup>.

Like most problems where the link between classification label and data is not analytically defined, this approach relies on a comprehensive database of labeled images from which correlations and relationships can be identified<sup>1,2</sup>. Without a large repository of sample data, attribute selection and classification may not accurately reflect relationships present in experimental samples<sup>2</sup>. Unfortunately, a database of labeled scattering patterns was previously very difficult or impossible to acquire for the case of cellular light scattering, due to the large computational cost of simulation (*e.g.* using Finite Difference Time Domain methods<sup>5,9</sup> (FDTD)), the lack of well-labeled experimental testing data, and restrictive simulation models<sup>9,10</sup>.

To alleviate this issue, we recently demonstrated a computationally inexpensive simulation method—the `mtPatterns` algorithm—that is capable of rapidly simulating wide-angle light scattering from organelles in a single cell<sup>10,11</sup>. This method builds on Fourier theory and Rayleigh-Gans theory to approximate the scattering contributions from different 3D organelle distributions. The `mtPatterns` scattering simulations have been shown to agree well with both experimental scattering patterns and FDTD simulations<sup>10</sup>, providing labeled

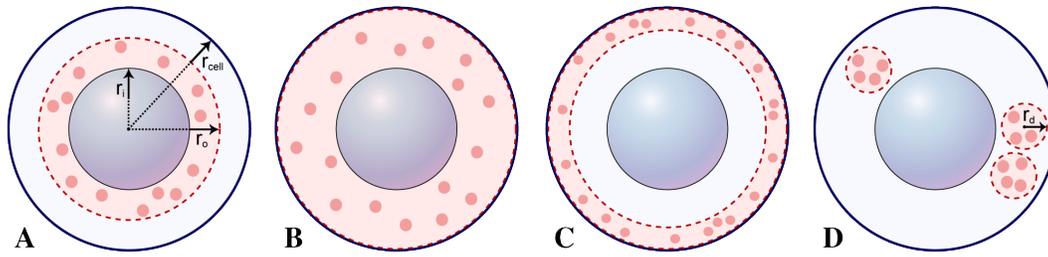
scattering data without the high computational cost of FDTD—each pattern can be generated in seconds to minutes on a desktop computer<sup>10</sup>. This makes it possible to create a large and varied database of labeled wide-angle scattering images.

### 5.2.1 Clinical Relevance and Cell Morphology

Recent work has shown that the distribution and aggregation of organelles within a cell can have great predictive value<sup>12,13</sup>. For this work we focus on one specific intracellular organelle: the mitochondrion. Not only are mitochondria responsible for the majority of wide-angle light scattering<sup>14</sup>, making them a primary determinant of scattering pattern composition<sup>10</sup>, but their form and function have been found to relate to the progression of a number of ailments<sup>12–18</sup>. Specifically, mitochondria are critical to cellular energy production. Mutations that affect mitochondria and related proteins have been identified in disorders such as cancer<sup>14,15,17</sup>, cardiomyopathy<sup>17</sup>, diabetes<sup>15</sup>, common neurodegenerative disorders like Parkinson and Alzheimer Diseases<sup>9,15</sup>, and can also affect a patient's response to treatments such as chemotherapy<sup>12,13</sup>.

Despite a debate surrounding the exact shape, size, and distribution of mitochondria within human (and other mammalian) cells, it appears their morphology is highly cell-dependent<sup>14,19</sup>, and they are commonly held to be able to alter their shape, size, aggregation, and location within a cell based on external stimulus<sup>12,13,18–21</sup>. As an example, Robin and Wong have calculated that the number of mitochondria in different mammalian cells can vary from approximately 83–677 organelles, with human lung fibroblasts averaging  $308 \pm 47$  mitochondria<sup>22</sup>. Frey *et al.* and Modica-Napolitano *et al.* illustrate that mitochondrial length along an axis commonly ranges from  $0.1\text{--}2.0\mu\text{m}$ <sup>18,21</sup>, while mitochondrial size values for different diseased cells (deficiencies in cytochrome oxidase) have been estimated at  $0.75\text{--}0.83\mu\text{m}$  (<sup>14,23</sup>, and even as high as  $1.5\text{--}2.7\mu\text{m}$  for some breast cancer (MCF7) and cervical carcinoma (Siha) cells<sup>24</sup>. The cellular area (*i.e.* cytoplasm) available for use by mitochondrial distributions also varies with cell type. In carcinomas of interest, average cell size is approximately  $15\text{--}16\mu\text{m}$  in diameter, with a nuclear diameter of  $7.5\text{--}9.5\mu\text{m}$ <sup>24,25</sup>. For example, Siha cells have an outer diameter of  $15.5 \pm 2.4\mu\text{m}$  and a nuclear diameter of  $8.9 \pm 1.5\mu\text{m}$ , while MCF7 cells have an outer diameter of  $15.0 \pm 2.1\mu\text{m}$  and a nuclear diameter of  $9.2 \pm 1.3\mu\text{m}$ <sup>24</sup>. In previous simulation work, Su *et al.* used an outer diameter of  $16.0\mu\text{m}$  and a nuclear diameter of  $8.0\mu\text{m}$  for Raji (Burkitt's lymphoma) cells<sup>9</sup>, and this holds with a study by Chignola *et al.*<sup>25</sup>.

Mitochondrial clustering within the cytoplasm can take many forms, including *perinuclear* (distributed around the nucleus), *peripheral* (distributed around the outside of the cell), *diffuse* (evenly distributed within a cell), and *aggregate* (in tight clumps within a



**Figure 5.1:** Schematic of four different organelle distributions: (A) a *perinuclear distribution*<sup>12</sup>, where organelles are distributed close to the nuclear wall ( $r_i = r_{nucleus}$ ) but still inside an outer sphere of radius  $r_o$ ; (B) a *diffuse distribution*<sup>12</sup> where organelles are randomly placed throughout the entire volume between the cell wall and the nuclear wall; (C) a *peripheral distribution*<sup>12</sup>, where organelles are placed in a spherical shell of inside radius  $r_i$  and outer radius  $r_o = r_{cellwall}$ ; and (D) an *aggregate distribution*<sup>12,14</sup>, where organelles are distributed in one or more tightly packed spherical regions of radius  $r_d$  (e.g. a highly organized collection of mitochondria<sup>14</sup>).

cell)<sup>12–14,19,26</sup>. A schematic representation of these distributions is shown in Fig. 5.1 (not to scale). The exact geometry values for each distribution class varies with medical condition and cell type, and are generally defined by human experts on a case by case basis. For this study, distribution dimensions are based on a sampling of MitoTracker cell stains, organelle distribution histograms, and cell imaging results published by Stojkovic *et al.*<sup>27</sup>, Sikder *et al.*<sup>12</sup>, and Hallmann *et al.*<sup>26</sup>

For many disorders, there is a noticeable genetic or structural difference between the mitochondrial populations of healthy and damaged cells<sup>9,12,14,26</sup>. In particular, the distribution and aggregation of mitochondria within a cell has been shown to relate to different disease-related metabolic states, and can help predict a damaged cell's response to treatment<sup>12</sup>. The presence of *perinuclear* mitochondrial clustering in cultured cancer cells has been shown to indicate a sensitivity to chemotherapy, while a *peripheral distribution* in the same culture indicates potential chemotherapy resistance<sup>12,13</sup>. In a similar fashion, a differentiation between *diffuse* and *perinuclear* clusters can help distinguish between normal and transformed cells<sup>12</sup>, and the detection of *aggregates* can help in cell classification<sup>12,14</sup>.

If it were possible to rapidly assess mitochondrial distribution, it could be used as a powerful indicator of cellular health. Unfortunately, there are few—if any—current techniques that can non-invasively identify mitochondrial clustering trends in mixed cell populations<sup>12</sup>. However, as shown in this work, wide-angle light scattering analysis could provide one method to perform non-invasive, label-free distribution analysis when supplemented with image analysis and pattern recognition tools.

## 5.3 Methods

To evaluate the correlation between mitochondrial placement and scattering pattern composition, we utilized a rapid light-scattering simulation method—the `mtPatterns` algorithm<sup>10</sup>—to generate wide-angle scattering patterns for the different diseases-related mitochondrial distributions described in the previous section. The resulting database of scattering images was evaluated using a set of standard biomedical image analysis routines<sup>1,7</sup>. From this evaluation, image characteristics with the greatest predictive value were identified using a bank of attribute selection methods<sup>2</sup>. The selected attributes were then analyzed for their relationship to organelle distribution, and used in six different classifiers to demonstrate their predictive ability. Image generation and simulation setup, image analysis methods, and attribute selection methods are described in detail below. Results and a discussion of the data will be presented in the following section.

### 5.3.1 Simulation Setup

For the tests described in this study, we based simulation model parameters on known diseased and healthy cell morphologies. As per the cellular size values in described in Sec. 5.2.1, all simulations used an average outer cell radius of  $8\mu m$ , a nuclear radius of  $4\mu m$ ; mitochondria were then distributed within the available cytoplasm (*i.e.* between the outer wall and the nucleus, a volume of approximately  $151\mu m^3$ .) Using these parameters as a base, we implemented the four mitochondrial distributions shown in Fig. 5.1—*perinuclear*, *diffuse*, *peripheral*, and *aggregate* clustering. As noted above, distribution dimensions were based on an analysis of published MitoTracker cell stains, organelle distribution histograms, and cell imaging results.

To cover the wide range of examples presented in the biomedical literature<sup>12–14,26,27</sup>, each test included two different perinuclear distributions ( $r = 4.0–4.8\mu m$  and  $r = 4.0–5.6\mu m$ ), two peripheral distributions ( $r = 6.4–8.0\mu m$  and  $r = 7.8–8.0\mu m$ ), two diffuse distributions ( $r = 4.0–8.0\mu m$  and  $r = 5.0–7.0\mu m$ ), and three aggregate distributions ( $r_d = \{1.0, 1.5, 2.0\}\mu m$ ). While these tests encompass the range of radial distributions found in the literature, future implementations could be easily adapted and refined to accommodate a specific distribution range, non-radial model, or medical threshold.

We created examples of each distribution type at six concentration levels: when they contained 83, 125, 250, 375, 500, and 677 mitochondria—the range expected for mammalian cells<sup>22</sup>. The effective size of mitochondria—in terms of their achievable spacing within a cell—was also varied for each test model, and results are presented for spherical organelles

placed using effective diameters of  $d_{mt} = 0.25\mu m$ ,  $d_{mt} = 0.75\mu m$ , and  $d_{mt} = 1.0\mu m$  (as per Sec. 5.2.1). For each combination of distribution type, effective organelle size, and concentration level, we performed fifteen simulations, each with a different randomly generated organelle arrangement within the specified distribution (*i.e.* fifteen different random seeds) to simulate cell rotation and other variability inherent in real experimental systems. This resulted in a total of 1455 different valid cell models for simulation and analysis (models where all organelles could not be placed within a containing volume were not included).

Using these cell models as input, we applied the `mtPatterns` scattering simulation algorithm<sup>10,11</sup> to generate the final database of 1455 scattering images. The `mtPatterns` algorithm is described in Pilarski *et al.*<sup>10,11</sup>, and initialized with an incident light wavelength of  $\lambda = 632nm$ , a receptive field area of  $3mm \times 3mm$ , and with the center of the cell model placed  $d = 5mm$  below the center of the receptive plane. As indicated by Pilarski *et al.*, this setup gives a viewable side-scatter region between  $77.3^\circ$  and  $106.7^\circ$  in both the  $\theta$  and  $\phi$  axes (*e.g.* a cone of  $\simeq 30^\circ$ )<sup>10</sup>.

### 5.3.2 Image Analysis

For this work, we utilized three common approaches to image texture analysis, as described by Rangayyan: *statistical texture measures*, *Haralick texture measures*, and *Law texture energy*<sup>1</sup>. A full list of the resulting texture attributes is shown in Tab. 5.1. As described above, these measures each focus on different aspects of an image’s composition in terms of grey-level intensity, and together form a compact but descriptive image signature. Source images were  $100 \times 100$  pixels ( $px$ ) in size. For GLCM computation, images were reduced in size to  $50 \times 50px$ , with 64 luminosity levels. As per Rangayyan, we used a pixel distance of one, and averaged over relationships at  $0^\circ, 45^\circ, 90^\circ$ , and  $135^\circ$ <sup>1</sup>. For this work we also implemented two additional Law measures,  $S5x2$  and  $S5x4$ , that measured spot content at two and four times the scale of the standard  $S5$  filter. This was done by reducing the size of the filter’s input image by a factor of two and four, respectively. All texture analysis algorithms were implemented as described by Rangayyan<sup>1</sup>. After analysis, each vector of features (statistical, Haralick, and Law—twenty-one attributes in total) was labeled with the super-class of the organelle distribution used to generate the image—*aggregate*, *perinuclear*, *diffuse*, or *peripheral*. Once labeled in this way, data points could be used for attribute selection and classification.

**Table 5.1:** List of texture analysis attributes generated during image processing; terms defined as in Rangayyan<sup>1</sup>

<b>Attribute</b>		<b>Attribute</b>	
H1	homogeneity	VAR	variance
H2	contrast	SKEW	skewness
H3	correlation	KURT	kurtosis
H4	sum of squares	L5	luminosity
H5	local homogeneity	E5	edge content
H6	sum avg.	S5	spot content
H7	sum var.	S5x2	spot content
H8	sum entropy	S5x4	spot content
H9	entropy	R5	ripple content
H10	diff. var.	W5	wave content
H11	diff. entropy		

(a) Haralick Attributes

(b) Statistical and Law Attributes

### 5.3.3 Attribute Selection

For the task of identifying texture attributes that relate to organelle distribution, we applied five of the attribute selection mechanisms implemented by Hall *et al.* in their 2003 benchmarking study<sup>8</sup>. This includes three subset identification methods—CFS Subset Evaluation, Consistency, Wrapper Bayes Network—and two ranking methods—ReliefF and Information Gain<sup>8</sup>. Following suggestions by Witten and Frank, we also used the Chi-Squared and OneR rank-based selection routines as a comparisons<sup>2</sup>. As done by Hall *et al.*, all selection tasks were performed using the Weka environment (v3.6.0), a widely used Java-based data mining toolkit<sup>2</sup>. All twenty-one texture attributes (described in Tab. 5.1) were presented to each attribute selection routine using 10-fold cross-validation (cv10)—the data was divided into ten folds, trained on nine of the folds and then tested on the remaining, previously unseen, fold<sup>2,28</sup>. Attributes were then sorted by the number of times they were selected in the top five or top three by selection routines (or, for subset methods, included in > 50% or 100% of the validation folds, respectively).

### 5.3.4 Classification

To study the use of scattering image texture in the classification and prediction of organelle distributions within single cells, combinations of the twenty-one texture attributes were given as input to six different machine learning systems: a *Random Decision Tree*, a *Support*

*Vector Machine*, a *Multilayer Perceptron*, a *Bayes Network*, a *OneR* rule system, and a *Naive Bayes* learner. This was done to cover a wide range of algorithmic assumptions and methods of handling the data. As in attribute selection, tests were performed using 10-fold cross-validation (*i.e.* systems were tested for classification accuracy only on data they had never seen in training). For these tests, we used the standard Weka<sup>2</sup> implementations of each classifier: `trees.RandomTree`, `functions.SMO`, `functions.MultilayerPerceptron`, `bayes.BayesNet`, `rules.OneR`, and `bayes.NaiveBayes`. Classifiers were provided with the entire set of 1455 valid data points; the class of each point was determined by its superclass label: *aggregate*, *perinuclear*, *diffuse*, or *peripheral*.

### 5.3.5 Extensions to the mtPatterns Algorithm

For use in this work, the `mtPatterns` algorithm of Pilarski *et al.* was extended by adding a collision checking routine to ensure the realistic placement of each scattering point given a set of physical constraints—when placing organelles within the available distribution volume, a spherical boundary was used to check for collisions with the cell wall, nucleus, and other organelles. In this way, the scattering points placed by the `mtPatterns` algorithm could maintain realistic spatial relationships, and avoid positions that would be not physically achievable for organelles with volumes similar to those listed in Sec. 5.2.1. All simulations in this study used the updated form of the `mtPatterns` algorithm<sup>11</sup>.

## 5.4 Results and Discussion

The result of the attribute selection process is shown in Tab. 5.2. The first three columns (subset methods) list the number of times out of ten each attribute was selected for a subset by each method, while the final columns (ranking methods) indicate an attribute’s relative rank as decided by each method.

As determined by rank ordering the results of the attribute selection tests—shown in Tab. 5.3—the five most predictive attributes were found to be Law features measuring the spot content of the images at two different scales, S5 and S5x2, the image’s wave content W5, edge content E5, and ripple content R5. These attributes were consistently selected and positively ranked by almost all of the attribute selection routines. Given that wide-angle mitochondrial scattering images are semi-ordered or disordered arrangements of bright spots, known to vary in size and spacing with the distribution and number of organelles<sup>10,29</sup>, it follows that the selected Law measures would contain predictive information about distribution class.

**Table 5.2:** Attribute selection results using 10-fold cross-validation; terms as defined in Sec. 5.3.3

Attribute	CFS	Consistency	WBN	ReliefF	OneR	ChiSquared	InfoGain
Variance	–	2	1	19	20	20	20
Skewness	–	2	–	20	14	18	17
Kurtosis	–	1	1	21	18	19	19
H1	9	–	2	13	12	12	13
H2	3	7	–	8	7	8	8
H3	–	5	–	16	15	15	16
H4	–	10	–	18	19	17	18
H5	–	3	1	15	13	11	12
H6	–	–	1	17	17	16	15
H7	10	–	–	11	6	10	9
H8	–	2	–	12	11	14	11
H9	–	9	–	10	16	13	14
H10	1	–	1	7	10	9	10
H11	6	–	–	9	8	7	7
L5	1	–	7	14	21	21	21
E5	–	5	–	2	5	5	5
S5	10	10	10	3	1	1	1
S5x2	10	3	4	1	2	2	2
S5x4	10	9	1	4	9	6	6
W5	4	3	1	6	3	3	3
R5	9	1	1	5	4	4	4

*CFS, Consistency, WBN (Wrapper BayesNet):* number of tests in which an attribute was selected.  
*ReliefF, InfoGain, ChiSquared, OneR:* rank order of attributes (1=best)

**Table 5.3:** Attribute ranking results; terms as defined in Sec. 5.3.3

Attribute	Votes (Top 5*)	Votes (Top 3**)	Rank
S5	7	7	1
S5x2	5	5	2
W5	3	3	3
E5	5	1	4
R5	5	0	5
S5x4	3	1	6
H4	1	1	7
H7	1	1	7
H1	1	0	8
H2	1	0	8
H3	1	0	8
H9	1	0	8
H11	1	0	8
L5	1	0	8

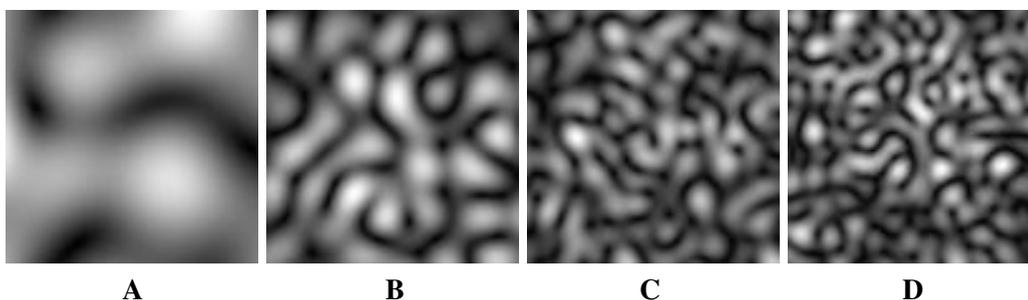
\* ranked in the top five, or was selected in 50% or more of the tests (*i.e.*  $> 4$ )

\*\* ranked in the top three, or was selected in 100% of the tests (*i.e.* 10)

In terms of the images themselves, we observed distinct qualitative and quantitative differences between the scattering patterns of each distribution type. In the aggregate case, the scattering pattern consisted of a small number of broad and smoothly changing bright intensity patches (Fig. 5.2, A). For the peripheral distribution, the number of intensity regions was much larger, while the size of each region was only a fraction of the image’s width (Fig. 5.2, D). The other distributions were found to fall between these two extremes, in relation to distribution size and type (Fig. 5.2, B,C).

Figure 5.3 clearly demonstrates this correlation between image complexity and distribution size—as the outer radius of the distribution increased, there was a corresponding increase in the complexity of the image. This increase was much more pronounced as the *average* radius of the distribution increased—*i.e.* an increase in both the inner and outer radius values (Fig. 5.3B), as opposed to just an increase in outer radius (Fig. 5.3A).

The increase in image complexity can be numerically quantified in terms of image texture attributes. Figure 5.4 shows a plot of Law texture energy for distributions with increasing outer radius, and outer/inner radius. Each data point represents the average of fifteen different samples. There is a clear upward progression of texture values to match the visual increase in image complexity observed in Fig. 5.3. This suggests that the methods described in this work can be applied even in application domains without fixed distribution types, and



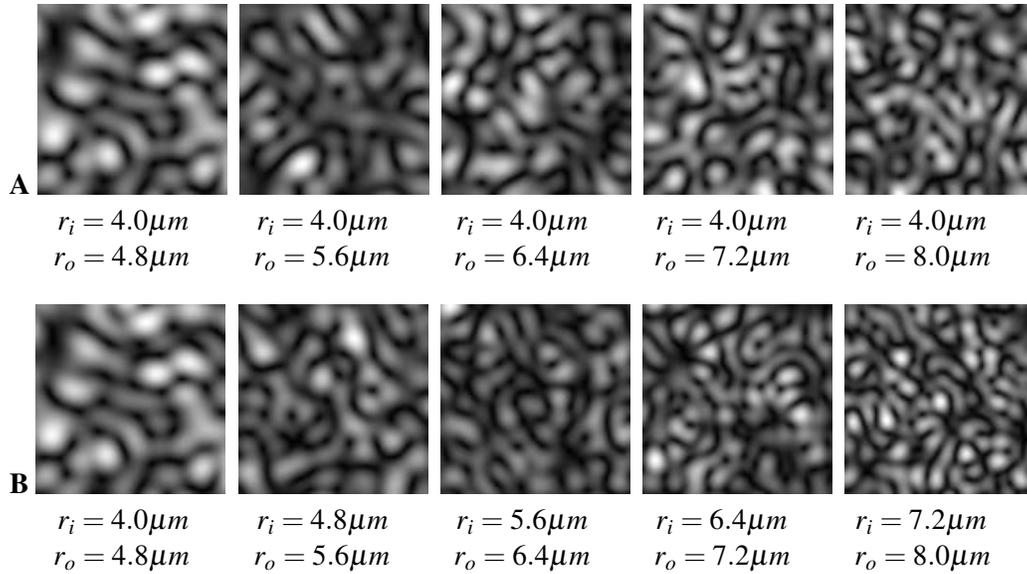
**Figure 5.2:** Qualitative comparison of scattering patterns from the four different mitochondrial distributions: (A) an *aggregate distribution* clustered within a  $2\mu\text{m}$ -radius spherical region next to the nucleus; (B) a *perinuclear distribution* arranged within  $0.8\mu\text{m}$  of the nucleus; (C) a *diffuse distribution*; and (D) a *peripheral distribution* arranged within  $0.8\mu\text{m}$  of the outer cell wall. All four examples used cell models with 250 organelles each of effective size  $d_{mt} = 0.25\mu\text{m}$ ; the distribution boundary is the only variable altered between images.

are applicable to environments where organelle distribution may instead be defined by one or more continuous variables.

As would be expected from Figs. 5.2 and 5.3, differences between organelle distributions—aggregate, perinuclear, diffuse, and peripheral—can be quantified using scattering image attributes. Fig. 5.5 shows histograms of Law E5, W5, S5, and S5x2 texture energy measures for the 1455 sample images. As noted above, these Law measures represent an image’s edge content (E5), wave content (W5), and spot content at two different scale sizes (S5 and S5x2). The purity—minimal class overlap—and coverage of each distribution’s curve along the texture axis shows that these Law features were able to effectively separate the different distribution types. Visual analysis verified that each attribute’s predictive ability—as prescribed by the attribute selection routines, Tab. 5.3—matched with the degree of class separation and coverage found in its histogram plot (*e.g.* compare the histograms shown in Figs. 5.5A, 5.5D, and 5.6B with the attribute ranking in Tabs. 5.2 and 5.3).

This quantitative assessment holds with qualitative observations. The smooth and infrequent intensity peaks of an aggregate distribution’s scattering pattern (Fig. 5.2, A) are reflected in its low E5, S5, S5x2, and W5 Law texture energy. Conversely, the complex pattern of a peripheral distribution (Fig. 5.2, D) was found to have much higher edge, wave, and spot texture energy. Haralick attributes also indicated a greater level of local homogeneity (H5) and lower contrast value (H2, Fig. 5.6) in aggregate samples, when compared to the other distributions, which is consistent with a visual analysis of image complexity.

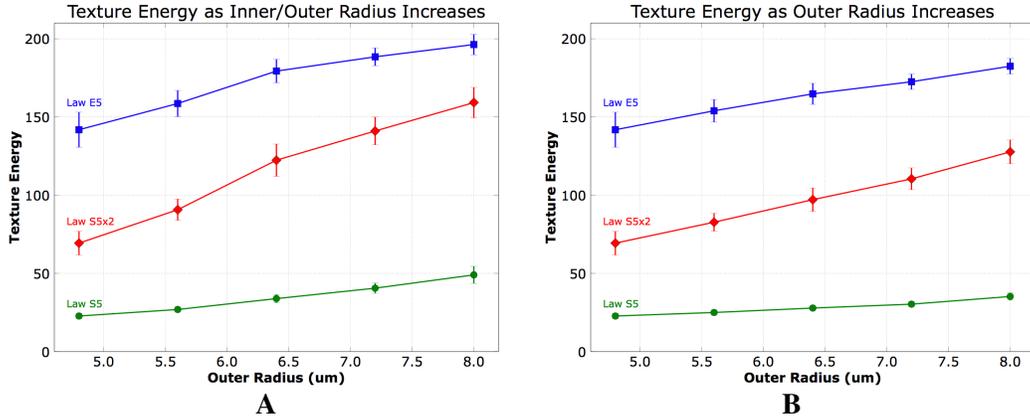
In terms of attribute selection, we observed an interesting difference between the selection choices of the *subset methods* and the *ranking methods*—some attributes consistently chosen for inclusion in predictive subsets were ranked poorly as individual predictors (*e.g.* H2



**Figure 5.3:** The relationship between the radius of an organelle distribution and the complexity of the resulting scatter pattern, for a fixed number of scatterers and two different distribution types: (A) a distribution with increasing outer radius and (B) a distribution of increasing outer and inner radius. Distributions contained 250 spherical organelles, each with an effective diameter of  $0.25\mu m$ .

and S5x4). As shown in Fig. 5.7, analysis of the respective histograms showed these particular attributes to be excellent candidates to separate aggregate distributions, and to a lesser degree perinuclear distributions, from all other distributions, but poor choices for finding differences between the diffuse and peripheral distributions. It seems likely that subset selections routines included them in subsets as complimentary features able to quickly detect aggregate distributions.

Using the same methods and dataset, it was also possible to identify the different size subclasses for each distribution type. As shown through the arrangement of sub-class center points in Fig. 5.7, there was a noticeable stratification within Law texture energy plots, with each major distribution super-class being divided up incrementally into overlapping sub-classes of increasing average radius (*e.g.* the average radius value for perinuclear sub-class n56,  $r_o = 5.6\mu m$ , is greater than that for n48,  $r_o = 4.8\mu m$ ). This is consistent with observations on the relationship between average radius and image complexity (Fig. 5.3), and gives further insight into the discrimination power of the different texture attributes. A comparison of sub-class histograms with the ranking values in Tabs. 5.2 and 5.3 shows that top-ranked (predictive) attributes appear to be able to not only separate the four major class distributions, but also accurately stratify sub-classes with regard to their radius. Due to the overlap between similar sub-classes (*i.e.* those with the same super-class, as shown



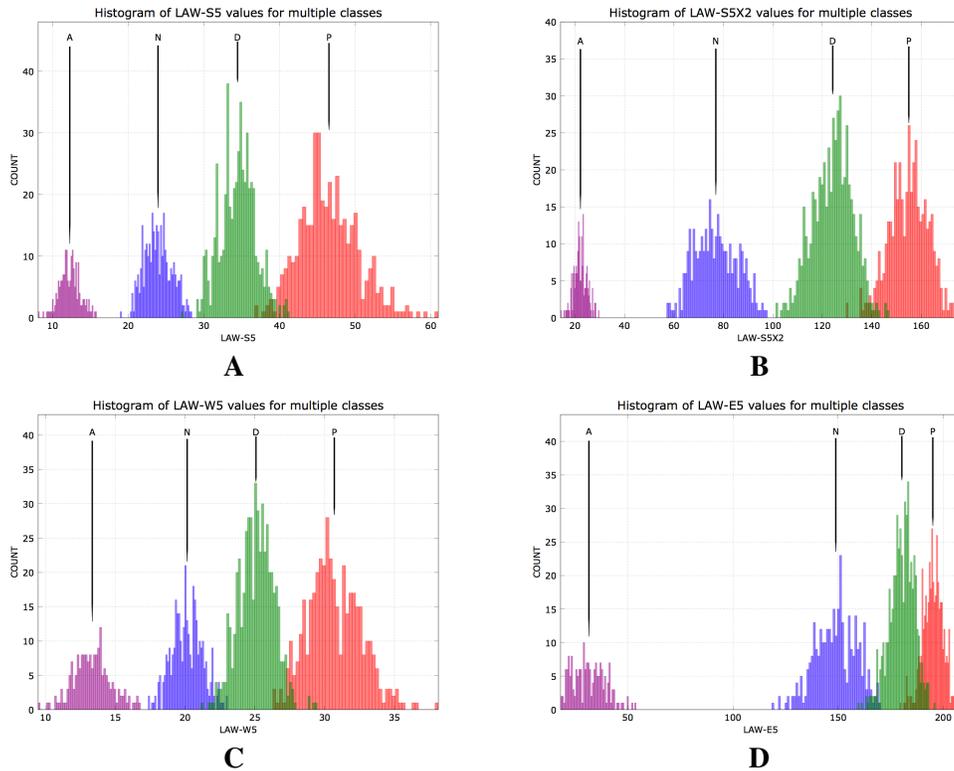
**Figure 5.4:** Relationship between distribution radius and texture energy for models with (A) an increasing outer radius and increasing inner radius  $r_i = r_o - 0.8\mu m$ , and (B) an increasing outer radius and a static inner radius  $r_i = 4.0\mu m$ . Shown for E5 (blue squares), S5 (green circles), and S5x2 (red diamonds) texture metrics. Each point is the average of fifteen different samples; error bars indicate standard deviation.

by related colours in Fig. 5.7), it follows that sub-class will be much more difficult to distinguish directly from a scattering plot. As described below, this is in fact the case. It is important to note, however, that even in sub-class histograms there is a clear delineation between different distribution super-classes.

Preliminary results indicate that the separation of sub-classes is more pronounced in samples with a greater number of organelles (*i.e.* sub-class overlap is inversely related to organelle concentration). Histograms from a data set with a concentration of 500 organelles per cell showed greater separation between sub-class means and less variance than the corresponding data set with a concentration of only 83 organelles per cell. This is likely due to the fact that random placement effects have a less pronounced impact on distribution shape when the number of organelles increases—organelles more consistently fill the available volume, reducing shape variability between random instances of each sub-class model.

In general, however, observed trends were largely invariant to organelle concentration and effective mitochondrial radius. It appears that fluctuations in distribution shape and size are the dominant causes of changes to scattering image composition; this is in agreement with previous work<sup>10,11</sup>. Changes to the size and arrangement of a distribution were closely linked to consistent image attribute variation, while changes to the mitochondria concentration and effective organelle size did not lead to a significant shift in image attributes.

As scattering intensity can be shown to increase with the addition of organelles<sup>10</sup>, it is expected that the Law L5 (luminosity) texture measure would be a good indicator of con-

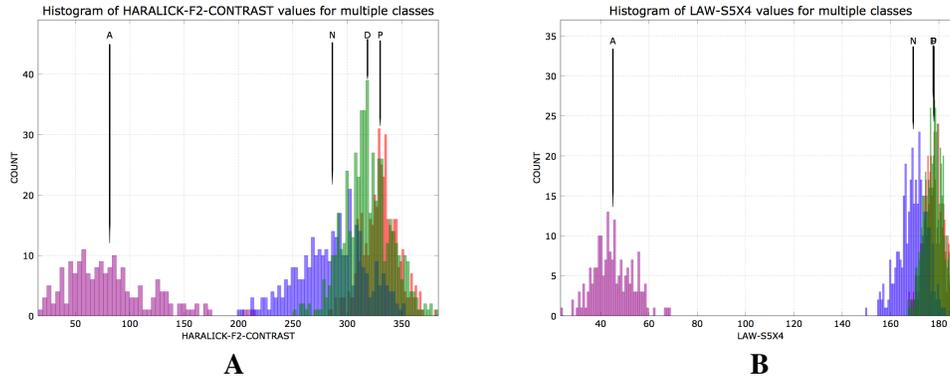


**Figure 5.5:** Histograms showing (A) S5, (B) S5x2, (C) W5, and (D) E5 Law texture energy for all 1455 sample images. Distribution class is indicated by colour—purple for aggregate, blue for perinuclear, green for diffuse, and red for peripheral. A labeled vertical line indicates mean value of each class. As shown by the purity of each class along the axis, these Law texture energy metrics are able to separate the four major distribution types with very little class overlap, making them good discriminators for mitochondrial arrangement.

centration when presented with un-normalized scattering data (only normalized image plots were available for use in the current study). Correlations between un-normalized image intensity and mitochondria concentration will be explored at length in future work.

#### 5.4.1 Classification of Distributions

Results from the classification tests are shown in Tab. 5.4. All classifiers showed a consistently high classification accuracy, and could reliably identify all four distribution types. Misclassification errors were almost exclusively due to discrimination errors between the diffuse class and the neighbouring regions of the perinuclear and peripheral classes. This can be seen in quantitative terms by examining the confusion matrix for each algorithm. As shown in Tab. 5.5, for learners using E5, S5, S5x2, and W5 as input, 13/15 of all mis-



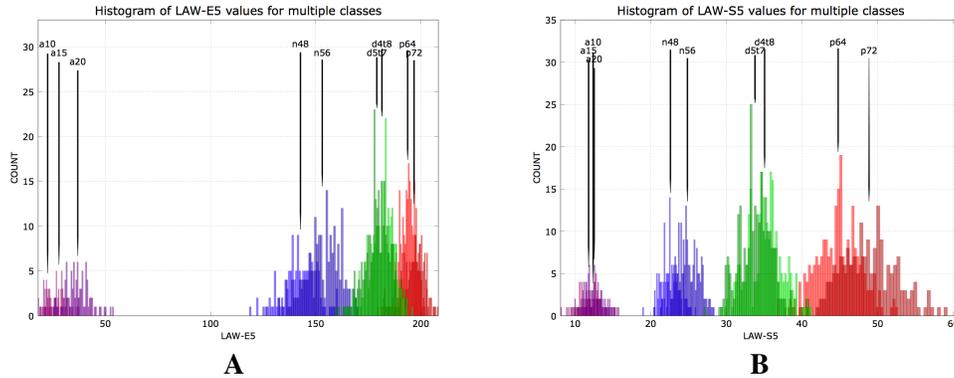
**Figure 5.6:** Histograms showing (A) Law S5x4 and (B) Haralick H2 values for all 1455 sample images. Distribution class is indicated by colour—purple for aggregate, blue for perinuclear, green for diffuse, and red for peripheral. A labeled vertical line indicates mean value of each class. Both these attributes were able to clearly discriminate between aggregates and all other classes, but were not effective separators for the other three distribution types.

classifications (for a multilayer perceptron, Tab. 5.5a) and 21/22 of all misclassifications (for Naive Bayes, Tab. 5.5b ) were between the diffuse and perinuclear classes.

This holds with our histogram analysis of all twenty-one texture attributes—none of the image attributes uniquely separated all four classes. As shown in Figs. 5.5 to 5.7, all attributes showed some class overlap between the diffuse distribution and the more central peripheral distribution (sub-class p64:  $r_i = 6.4\mu m$ ). For many attributes there was also an overlap between larger perinuclear distribution (sub-class n56:  $r_o = 5.6\mu m$ ) and the diffuse distribution. However, some attributes completely or almost completely separate the perinuclear and diffuse classes (*e.g.* S5 and S5x2); this accounts for the lower misclassification rate between these distributions in Tab. 5.5. Excluding these predictive attributes lead to a significant drop in classification accuracy, and a corresponding increase in confusion between the perinuclear and diffuse classes. This further confirms the predictive ability of the S5 and S5x2 attributes indicated in Tabs. 5.2 and 5.3.

As expected, each classifier’s performance changed depending on the number and type of attributes presented as an input set. Classification performance for six different input subsets is shown in Tab. 5.4—two small subsets containing the best parameters identified in Tab. 5.3, the subset identified by the CFS selection routine (H1, H7, H11, S5, S5x2, S5x4, R5), the subset identified by the Consistency selection routine (H2, H3, H4, H9, E5, S5, S5x4), the single best attribute, and a set containing all twenty-one texture parameters.

For all subsets, the OneR classifier chose to base its decisions on the S5 parameter, giving a classification accuracy of 98.7%. The performance of the two Bayesian classifiers and



**Figure 5.7:** Histograms showing (A) E5 and (B) S5 Law values for all 1455 sample images. Distribution sub-class is indicated by colour and a label indicating the distribution’s mean—purple for aggregate (a10,a15,a20), blue for perinuclear (n48,n56), green for diffuse (d4t8,d5t7), and red for peripheral (p64,p72). For both E5 and S5, the mean values for each sub-class are stratified with respect to radius (in ascending order) within their super-class.

the decision tree was found to improve as the number of parameters was reduced to only a small selection of predictive attributes (*e.g.* to S5,S5x2,E5, or simply S5). Bayesian algorithms and decision trees are adversely affected by redundant or noisy parameters<sup>2</sup>, which accounts for this preference toward a small, information-rich subset. Conversely, for the Multilayer Perceptron and Support Vector Machine, classification accuracy was found to increase with the number of predictive input parameters; these algorithms are known to effectively mitigate the effect of noisy parameters (*e.g.* L5) and redundant data<sup>2</sup>. However, their best performance was still obtained using the attribute subset identified by the Consistency algorithm, as opposed to the full set of twenty-one parameters. The best overall performance was 99.7%, given by the multilayer perceptron using the Consistency algorithm’s attribute subset—H2, H3, H4, H9, E5, S5, S5x4. This was followed by the Support Vector Machine at 99.5% using the same subset. The standard deviation for values in Tab. 5.4 was between approximately 0.6% and 0.8% (*i.e.*  $\pm 0.8$ –1.2 classified instances). This indicates that there is a statistically significant difference in the performance between classifiers for a given input set, and—to a lesser degree—performance between different input sets for each individual classifier (*i.e.* between best and worst cases for each classifier).

Classifiers tested on data points using sub-class labels showed a dramatically decreased classification accuracy. Despite good discrimination between sub-classes from different super-class distributions (*e.g.* between n56 and d4t8 or d5t7 and p64), a high misclassification rate between closely related sub-classes (*e.g.* between n48 and n56 or p64 and p72) lead to poor overall classifier performance for sub-class problems. As discussed above, and shown in Fig. 5.7, the large overlap between similar sub-classes makes consistent sub-class

**Table 5.4:** Classification accuracy using 10-fold cross-validation

Attributes	MLP	BayesNet	NaiveBayes	RandTree	SVM
All 21 Parameters	99.4	97.4	96.5	93.3	99.3
Consist. Parameters†	99.7*	96.7	96.3	95.3	99.5*
CFS Parameters††	99.3	98.4	98.2	98.1	99.0
4 Best: E5,S5,S5x2,W5	99.0	98.6	98.5	98.2	98.6
3 Best: W5,S5,S5x2	98.8	98.8*	98.8*	97.9	98.6
1 Best: S5	98.6	98.8*	98.8*	98.3*	98.4

\* *best classification rate*

† *H2, H3, H4, H9, E5, S5, S5x4*

†† *H1, H7, H11, S5, S5x2, S5x4, R5*

**OneR** accuracy was 98.7% for all subsets

**Table 5.5:** Example confusion matrices for two classifiers (best four attributes); row and column designations: (A)ggregate, peri(N)uclear, (D)iffuse, (P)eripheral. Numbers indicate how many times each true label (row header) was assigned to each classification category (column header)

A	N	D	P	←P	A	N	D	P	←P
180	0	0	0	<b>A</b>	180	0	0	0	<b>A</b>
0	329	1	0	<b>N</b>	0	330	0	0	<b>N</b>
0	1	519	5	<b>D</b>	0	1	513	11	<b>D</b>
0	0	8	412	<b>P</b>	0	0	10	410	<b>P</b>

(a) Multilayer Perceptron

(b) Naive Bayes

separation difficult. While accurate differentiation between super-class labels is sufficient for medical applications, the texture-based separation of closely related sub-distributions is a challenging problem that remains open for study.

## 5.5 Conclusions

In this chapter we showed how image analysis and pattern recognition methods can be used to identify relationships between wide-angle scattering pattern composition and the distribution of organelles (in this case mitochondria) within a cell. This was achieved using a library of scattering patterns generated by the `mtPatterns` algorithm and a series of standard image analysis, attribute selection, and machine learning tools. Results indicated that several components of scattering image texture are directly related to the arrangement of organelles within a cell. In particular, we observed that texture attributes could be used to

distinguish between medically important distribution types. Importantly, we also found that the predictive value of texture attributes is robust to changes in organelle concentration, effective size, and random placement. This provides a powerful set of techniques to interpret cellular light scattering signatures.

Scattering pattern composition—in terms of twenty-one image texture measures—was found to vary in a predictable and consistent way between different organelle distributions and aggregations. Notably, the complexity of scattering images is inversely proportional to the radius of the scattering distribution, where the complexity was reflected in the amount of edge, wave, and spot content detected in the images using Law texture measures. Very little overlap between classes was observed in Law texture energy histograms, and it was also possible to distinguish (to a first-order approximation) the average radius of a distribution from the data in these plots. Due to their ability to effectively separate scattering distributions, and their rapid processing time, Law features were found to be well suited for the analysis of mitochondrial scattering images.

Using these identified trends, it was possible to accurately classify different mitochondrial distributions directly from their wide-angle scattering patterns using Law and Haralick texture measures and standard machine learning techniques. In particular, Multilayer Perceptrons and Support Vector Machines were found to effectively distinguish each distribution type, especially when presented with texture attributes selected by the Consistency attribute selection algorithm. The classification of finely stratified sub-distributions with large inter-class overlap remains a challenging problem for future research.

We show that texture attributes provide an effective, robust method to analyze mitochondrial light scattering images, and that attribute selection can be used to improve classifier performance in this domain. This is important, as mitochondrial distribution has been shown to directly relate to cellular health and susceptibility to treatment<sup>12,13</sup>. Based on the agreement between `mtPatterns` simulations and published FDTD and experimental cytometry results<sup>9,10</sup>, these predictive trends are expected to extend well to scattering patterns from actual patient samples.

## References

- [1] R. M. Rangayyan, *Biomedical Image Analysis* (CRC Press, Boca Raton, 2004).
- [2] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. (Morgan Kaufmann, New York, 2005).
- [3] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Computational analysis of

- mitochondrial placement and aggregation effects on wide-angle cell scattering patterns,” in *Proceedings of SPIE*, A. Wax and V. Backman, eds., vol. 7187, p. 71870J (12 pages) (2009). URL <http://dx.doi.org/10.1117/12.809730>.
- [4] X.-T. Su, K. Singh, W. Rozmus, C. Backhouse, and C. Capjack, “Light scattering characterization of mitochondrial aggregation in single cells,” *Optics Express* **17**(16), 13,381–13,388 (2009). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-17-16-13381>.
- [5] C. Liu, C. E. Capjack, and W. Rozmus, “3-D simulation of light scattering from biological cells and cell differentiation,” *Journal of Biomedical Optics* **10**(1), 014007 (12 pages) (2005).
- [6] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1998).
- [7] L. G. Shapiro and G. C. Stockman, *Computer Vision* (Prentice Hall, Upper Saddle River, N.J., 2001).
- [8] M. A. Hall and G. Holmes, “Benchmarking attribute selection techniques for discrete class data mining,” *IEEE Transactions on Knowledge and Data Engineering* **15**(6), 1437–1447 (2003).
- [9] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, “2D light scattering patterns of mitochondria in single cells,” *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [10] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Rapid simulation of wide-angle scattering from mitochondria in single cells,” *Optics Express* **16**(17), 12,819–12,834 (2008). URL <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-17-12819>.
- [11] P. M. Pilarski and C. J. Backhouse, “Theoretical foundations for robust wide-angle scattering pattern analysis,” *Optics Express* **submitted**, 5 pages (2009).
- [12] S. Sikder, J. M. G. Reyes, C. S. Moon, O. Suwan-apichon, J. H. Elisseeff, and R. S. Chuck, “Noninvasive mitochondrial imaging in live cell culture,” *Photochemistry and Photobiology* **81**(6), 1569–1571 (2005).
- [13] A. M. Villa and S. M. Doglia, “Mitochondria in tumor cells studied by laser scanning confocal microscopy,” *Journal of Biomedical Optics* **9**(2), 385–394 (2004). URL <http://link.aip.org/link/?JB0/9/385/1>.
- [14] P. L. Gourley, J. K. Hendricks, A. E. McDonald, R. G. Copeland, K. E. Barrett, C. R. Gourley, and R. K. Naviaux, “Ultrafast nanolaser flow device for detecting cancer in single cells,” *Biomedical Microdevices* **7**(4), 331–339 (2005).
- [15] G. M. Enns, “The contribution of mitochondria to common disorders,” *Molecular Genetics and Metabolism* **80**(1-2), 11–26 (2003).
- [16] M. Zeviani and S. Di Donato, “Mitochondrial disorders,” *Brain* **127**(10), 2153–2172 (2004).

- [17] D. C. Wallace, "Mitochondrial diseases in man and mouse," *Science* **283**(5407), 1482–1488 (1999).
- [18] J. S. Modica-Napolitano and K. Singh, "Mitochondria as targets for detection and treatment of cancer," *Expert Reviews in Molecular Medicine* **4**(9), 1–19 (2004).
- [19] D. Rube and A. van der Blik, "Mitochondrial morphology is dynamic and varied," *Molecular and Cellular Biochemistry* **256**(1-2), 331–339 (2004).
- [20] J. Bereiter-Hahn and M. Voth, "Dynamics of mitochondria in living cells: shape changes, dislocations, fusion, and fission of mitochondria," *Microscopy Research and Technique* **27**, 198–219 (1994).
- [21] T. Frey and C. Mannella, "The internal structure of mitochondria," *Trends in Biochemical Sciences* **25**(7), 319–324 (2000).
- [22] E. Robin and R. Wong, "Mitochondrial-DNA molecules and virtual number of mitochondria per cell in mammalian-cells," *Journal of Cellular Physiology* **136**(3), 507–513 (1988).
- [23] P. L. Gourley and R. K. Naviaux, "Optical phenotyping of human mitochondria in a biocavity laser," *IEEE Journal of Selected Topics in Quantum Electronics* **11**(4), 818–826 (2005).
- [24] M. Bartlett, G. Huang, L. Larcom, and H. B. Jiang, "Measurement of particle size distribution in mammalian cells in vitro by use of polarized light spectroscopy," *Applied Optics* **43**(6), 1296–1307 (2004).
- [25] R. Chignola, A. Del Fabbro, C. D. Pellegrina, and E. Milotti, "Ab initio phenomenological simulation of the growth of large tumor cell populations," *Physical Biology* **4**(2), 114–133 (2007).
- [26] A. Hallmann, R. Milczarek, M. Lipinski, E. Kossowska, J. H. Spodnik, M. Wozniak, T. Wakabayashi, and J. Klimek, "Fast perinuclear clustering of mitochondria in oxidatively stressed human choriocarcinoma cells," *Folia Morphologica* **63**(4), 407–412 (2004).
- [27] M. Stojkovic, S. A. Machado, P. Stojkovic, V. Zakhartchenko, P. Hutzler, P. B. Goncalves, and E. Wolf, "Mitochondrial distribution and adenosine triphosphate content of bovine oocytes before and after in vitro maturation: correlation with morphological criteria and developmental capacity after in vitro fertilization and culture," *Biology of Reproduction* **64**(3), 904–909 (2001).
- [28] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification: Second Edition* (Wiley Interscience, New York, 2001).
- [29] P. M. Pilarski and C. J. Backhouse, "A method for cytometric image parameterization," *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.

## Chapter 6

# A Method for Cytometric Image Parameterization

### 6.1 Introduction

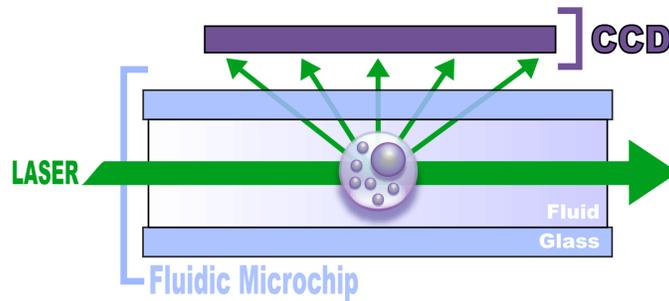
In this chapter we demonstrate a multi-agent computational pipeline that is able to extract features from a two-dimensional laser scattering image, cluster these features into spatially distinct regions—or bands—and extract a set of parameters relating to the structure and variability of intensity regions within the image. This parameterization can then be used to infer medically relevant properties of the scattering object. It also facilitates the decomposition of a complex scattering pattern into contributions from different cellular components, making it possible to prepare experimental images for the organelle-related wide-angle analysis presented in the previous chapter.

As described in earlier chapters, there is a great need for methods to extract and recognize patterns in cellular scattering images<sup>1-5</sup>. Scattering patterns contain vital information about the scattering source, and their interpretation facilitates diagnostic techniques ranging from the analysis of protein and DNA structure from X-ray diffraction<sup>6-8</sup>, to the assessment of cell health based on patterns of laser light scattered by cellular components<sup>1-3,5</sup>. In perhaps the best known example, Watson and Crick used information from patterns seen in two-dimensional X-ray scatter plots to infer the double-helix nature of DNA<sup>8</sup>. In assessing cellular structure, Sem'yanov *et al.* and Ghosh *et al.* recognized regular patterns in one-dimensional cell scattering plots, and were able to use a parameterization of these patterns to extract microstructural cell information<sup>9-11</sup>.

Scattering pattern analysis techniques are especially crucial in light of new medically relevant optical analysis methods—specifically the development of the wide-angle cyto-

---

<sup>1</sup>A version of this chapter has been published. Pilarski and Backhouse. 2006. *Optics Express*. 14(26): 12720–12743.



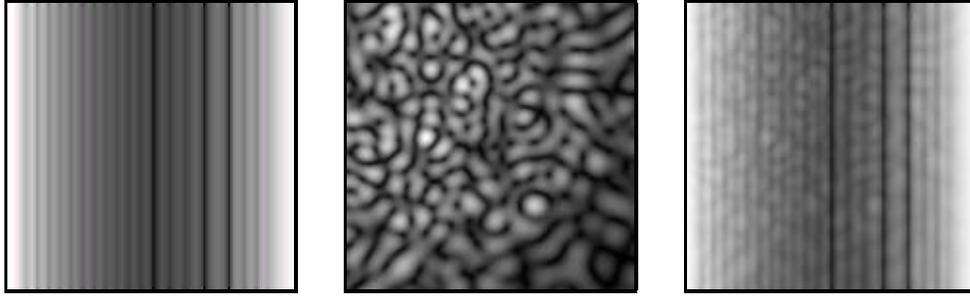
**Figure 6.1:** Schematic diagram of a wide-angle cytometer. It includes a fluidic channel, a laser source, and a two-dimensional charge-coupled device (CCD).

ter. Wide-angle cytometry devices are rapid, cost effective systems able to capture two-dimensional scattering patterns from a single cell or particle suspended within a fluidic channel. In these devices, laser light is propagated through a cellular body, where it scatters and is collected by a digital imaging device (as described by Singh *et al.*<sup>1,2</sup>). A schematic diagram of a wide-angle cytometer is shown in Fig. 6.1.

Building on traditional cytometry schemes—which typically only capture scattered light at a few fixed angles or an angular slice—these label-free (*i.e.* non-fluorescent) detection devices provide extensive information about the internal structure of cells and are highly relevant to medical diagnostic practices<sup>1,2</sup>. It is important to be able to rapidly ascertain small deviations in cell structure, as the structure of a cell can be an indicator for the progression of diseases (such as cancer) in patients<sup>4,5</sup>. However, to infer cell structure from two-dimensional scattering plots, feature extraction methods must be developed to extract and parameterize intensity patterns in cytometric scattering images. In this chapter we present a parameterization approach based on local shape features in scattering images. This method is used to extract both a geometric idea of image structure and also a set of numerical features that complement known texture parameters to more fully interpret a wide-angle pattern.

## 6.2 Light Scattering Through Complex Cellular Bodies

Previous work has shown that when light scatters through the cellular body it generates a complex and information-rich pattern of overlapping intensity regions. These regions are created by interfering waves propagating through a variety of cellular structures with differing size and optical properties<sup>4</sup>. Based on our current understanding of the scattering mechanisms present in biological cells (as indicated experimentally<sup>1,2,4,9</sup> and through numerical simulation<sup>12–15</sup>), these two-dimensional scattering images are typically comprised



**Figure 6.2:** Simplified example images containing features known to be present in experimental and numerically simulated scattering patterns: a series of vertical intensity bands, like those found in micro-structural scattering (left), and a number of randomly placed high-frequency intensity regions, characteristic of nano-structural Rayleigh scattering (middle). Varying levels of high- and low-frequency intensity variation may be present in a single image, leading to complex, information-rich image structures (right). These simulated images were generated by the methods explained in the Sec. 6.5

of a set of scattering bands of varying intensity and width, with a number of additional high-frequency intensity regions (*e.g.* resembling those in Fig. 6.2). For examples of experimentally acquired scattering signatures, please see the recent work of Singh *et al.*<sup>1,2</sup>.

Scattering intensity contributions in cells typically come from three sources: large cell structures with diameter  $d$  greater than the incident wavelength  $\lambda$  (geometric scattering,  $d > \lambda$ , on the order of micrometers), cell structures slightly less than the wavelength of incident light (Mie scattering,  $\lambda/15 < d < \lambda$ ), and very small organelles (Rayleigh scattering, sizes on the order of nanometers,  $d < \lambda/15$ )<sup>4</sup>. These lead to three general image cases.

In the first case (geometric scattering, and Mie scattering as  $d$  approaches  $\lambda$ ), the scattered light will form large regular intensity bands, which—in the case of our wide-angle cytometers—appear as vertical stripes in captured wide-angle scattering images<sup>2</sup>. While bands may arc at low scattering angles (as shown by the images of Singh *et al.*<sup>2</sup>), they appear approximately linear over smaller solid angles—particularly in the side-scattering region (*e.g.* Fig. 6.2, left). These larger intensity bands are most prominent (*e.g.* highest intensity) in the forward and back scatter regions of a  $180 \times 180$  degree scattering image, and are primarily due to the geometry of the cell wall and the larger organelles within the cell<sup>2-4,12</sup>.

In the second case, combining the influence of both large and medium-sized microstructural elements (*e.g.* both geometric scatterers and larger Mie scatters), a scattering image may contain bands that vary greatly in intensity along their length. Interference can lead to lighter or darker regions positioned within the intensity band structure.

For cellular scattering, the presence of smaller micro- and nano-scale cellular structures (like the mitochondria, which are primarily responsible for scattering at large angles<sup>5</sup>) will lead to a set of small randomly distributed intensity regions. The number, frequency, and size of the regions relates to the internal complexity of the cell. This is a result of the third case: Rayleigh scattering (and also Mie scattering where  $d$  approaches  $\lambda/15$ ). Intensity contributions from spatially distributed organelles will constructively and destructively interfere to create a number of high-frequency intensity regions (*e.g.* Fig. 6.2, middle).

The end result is a complex scattering pattern that is comprised of interfering contributions from high-frequency intensity components and a series of vertical intensity bands (such as in Fig. 6.2, right), and which indicates the detailed internal morphology of the cellular body. The combination of image cases one+two, one+three, or one+two+three will all lead to images similar to the one presented in Fig. 6.2, right. We have observed this complex structure in wide-angle cytometry images<sup>2</sup> and numerical Finite Difference Time Domain (FDTD) simulations<sup>15</sup>.

### 6.3 Computational Approaches to the Inverse Problem

Computational methods have done little to take advantage of this rich image structure. One of the major factors inhibiting the development of wide-angle diagnostic devices is the computational effort needed to analyze the scattered light signatures. To deduce cellular information from scattered laser light we must somehow solve the inverse scattering problem for light through biological cells. This inverse scattering problem involves recreating the geometric parameters of a cell based on the observed path of light propagating through its cytoplasm. This is a largely unsolved problem, and any direct mathematical methods are either computationally intractable and/or lead to non-unique solutions<sup>3</sup>. While numerous attempts have been made to simulate the effects of scattering in cellular bodies, a method for quickly inferring the geometric structure of a cell based solely on its light scattering data still eludes researchers<sup>3</sup>.

Given the challenge of solving the inverse problem for scattering from a living cell, the literature to date has focused on the empirical classification of cells based on their scatter at a few specific angles or an angular slice through the center of the full two-dimensional scattering pattern (commonly called the “indicatrix”). It is evident from the rich structure of the scatter patterns (along both the  $\phi$  and  $\theta$  axis) that there is far more information present than is contained in simple angular slices.

Techniques have been developed to address this problem by mathematically calculating the potential scattering patterns of cells<sup>12-14</sup>. In these ‘forward’ iterative methods, hypothetical cell geometries are used to generate simulated scattering signatures, which are then compared to experimental results. Further work has been done to use these calculated scattering patterns with evolving computer programs (such as genetic algorithms and neural networks) to interpret scattering data from crystals, proteins<sup>6,7</sup>, and single cells<sup>11</sup>. These methods largely involve the creation and verification of multiple potential structures (*e.g.* “generate and test” through repeated FDTD simulations<sup>12</sup>). These scattering simulations may take days to complete, and require the use of large parallel-processing computer arrays.

As shown by the work of Sem’yanov *et al.*, Ghosh *et al.*, and Ulanowski *et al.*, a more computationally tractable method is to effect a ‘parametric solution’ to the inverse scattering problem<sup>3,9-11,16</sup>. In this two-step method (feature extraction and pattern recognition), they parameterize some aspect of a scattering pattern and use a set of mathematical relations<sup>3,9</sup>, fast Fourier transforms<sup>10</sup>, or standard data mining techniques<sup>11</sup> to relate the extracted parameters to the initial structure of the scattering source. This process is rapid by comparison to iterative methods, and may allow a degree of structural generalization that alleviates some of the problems caused by non-unique forward solutions.

However, extracting viable parametric information from information-rich wide-angle scattering signatures presents a number of computational challenges. Because of complex cellular geometries, intensity bands may partially overlap in some places, the maximum intensity of each band may differ greatly from that of its neighbours, and the ambient background intensity is not consistent over the entire image. In addition, band boundaries are smooth gradients, not sharp intensity level transitions. These characteristics make it quite difficult to extract relevant features from an image and group them into meaningful regions.

While researchers have addressed the individual components that make up this high-level segmentation problem (*e.g.* feature detection/extraction, connected components labeling, noise rejection, region clustering), to the best of our knowledge no groups have developed a way to extract and analyze the full range of information present in two-dimensional cytometric scattering images. This problem involves partitioning two-dimensional scattering images into spatially distinct regions and extracting high-level semantic information (*i.e.* image parameters) from the detected regions. In this work we integrate and extend upon several tested image segmentation and computer vision techniques to enhance the diagnostic capacity of wide-angle cytometry systems through the automated shape-based parameterization of scattering plots.

### 6.3.1 Recent Segmentation Work

Computer vision and image segmentation lie at the heart of most medical image analysis schemes<sup>17–23</sup>. These are widely studied areas of research that are covered extensively in the literature. For the interested reader, Shapiro and Stockman (computer vision), Rangayyan (biomedical image analysis), and Pal and Pal (image segmentation) provide excellent reviews of the relevant background literature and techniques<sup>24–26</sup>.

While there are many possible methods to segment wide-angle scattering images, after surveying the body of current segmentation literature we chose to design our system within the framework of a shape-based multi-agent image processing environment (described below) due to its demonstrated power, flexibility, and novelty. Multi-agent segmentation systems (such as that of Liu *et al.*<sup>18,27,28</sup>) have been thoroughly tested in a number of image processing situations, and demonstrate comparable or superior performance when compared to traditional methods. In addition, the distributed nature of multi-agent systems is a benefit for future hardware implementation. As such, they provide a solid basis for the development of a cytometric image processing pipeline.

Cytometric image parameterization is primarily a high-level segmentation problem. A number of effective algorithms developed to subdivide an image into its component parts, using everything from texture information<sup>20,24,26,29–31</sup> and Markov Random Fields<sup>32</sup> (shown to be computationally demanding<sup>25,32</sup>), to standard image processing techniques<sup>24–26</sup>, models based on the human visual processing system<sup>33–36</sup>, and complex image processing networks<sup>17,21,22,37</sup>.

In addition, a large body of recent image segmentation work relies on the use of multi-agent swarms, including particle swarm optimizations<sup>23,38</sup>, evolutionary autonomous computational agents<sup>18,27,28,39–41</sup>, and ant-colony optimizations<sup>42</sup>. These multi-agent ('swarm') systems are composed of a population of autonomous or semi-autonomous 'agents' that collaborate (directly, indirectly, and/or competitively) to achieve a common goal. In this context, an agent is defined as a independent computational unit with a set of internal states and action rules; an agent's future behaviour depends on its current condition, programmed rules, and the state of its environment<sup>43</sup>. (Multi-agent systems are finding widespread use in engineering and computer science applications, ranging from process optimization to computer vision, population modeling to industrial control; Engelbrecht provides a good introduction to this topic<sup>43</sup>.)

All of these segmentation algorithms have one thing in common: they attempt to break a complex image into a set of smaller regions, where each region is homogeneous with respect to one or more parameters (*e.g.* contrast, intensity, texture)<sup>25</sup>. The effectiveness

of each method varies depending on the size, texture, orientation, and shape of features contained in an image; no single approach will work well for every image<sup>25</sup>. In most cases, image sub-division is a two stage process—an image is segmented into smaller sections which are then clustered into groups based on some similarity metric<sup>29,32</sup> (*i.e.* the split-and-merge or region-growing approach, recently used for tracking cells in diagnostic images<sup>44</sup>).

Liu *et al.* have recently proposed several interesting agent-based approaches to region detection and segmentation. They demonstrate a segmentation system capable of rapidly labeling homogeneous tissue regions in detailed brain scan images<sup>27</sup>, and present several methods to quickly detect edges and track shape information via a swarm of evolving computational instances (agents)<sup>18,28</sup>. In their swarm intelligence approach to image segmentation, the behavior of an agent is influenced by the agent's relation to other agents and local texture information (contrast threshold, regional pixel intensity deviation, and mean pixel value) contained in the analyzed image<sup>18,27,28</sup>. Their methods typically outperform traditional image processing techniques, and are successful over a diverse range of input data. Liu *et al.*'s method has distinct advantages in that it is highly parallel (a benefit for future hardware implementations), has proved successful in complex medical imaging environments, and facilitates a distributed feature clustering procedure.

Localized action and communication are the key components of most agent-based systems. Bourjot and colleagues have recently shown that a multi-agent system, based on the web-spinning behavior of social spiders, can effectively classify regions of homogeneous color in photographic images<sup>39</sup>, and ant colony optimizations have been used in autonomous vehicle navigation to detect roadways in low-contrast environments<sup>42</sup>. The work of Ghrist and Lipsky with self-assembling tile structures demonstrates an effective method for high-level organization with no centralized control<sup>45</sup>, and Omran *et al.* further show how particle swarm optimizations can dynamically cluster information during image segmentation<sup>38</sup>. The distributed shape classification of Mirzayans *et al.*<sup>41</sup>, and Wang and Yuan's agent-based face identification<sup>40</sup> also use local neighbourhoods to detect prominent features.

We use components of these successful swarm / image processing techniques (<sup>25,30,40,41,44</sup>) to complement the approach of Liu *et al.* and refine our system for use in a scattering analysis situation. We have also developed a set of unique algorithms to fully parameterize the detected image features in a way amenable to detailed scattering analysis.

Unlike most previous swarm segmentation work, our system does not involve agent reproduction or movement; the added complexity of agent dynamics, agent evolution, and agent fitness evaluation (with the additional possibility of incomplete feature detection) offsets any noticeable improvement for our particular application.

### 6.3.2 Computational Challenges

To parameterize scattering images we need to be able to detect continuous intensity regions and characterize them with respect to their spatial orientation within the image, their intensity profile, and their relationship to other parts of the image. This allows us to numerically represent the shape, size, and placement of low and high frequency intensity structures present in scattering images (as described above).

The complex image texture in cytometric scattering images makes simple feature detection problematic<sup>29</sup>. It is not possible to simply extract contiguous regions—corresponding to intensity bands—based solely on the raw intensity of the pixels (*e.g.* basic threshold-based region/edge detection<sup>25</sup>); the high intensity point of one band may be equal in value to the background intensity at another point in the image. Feature detection methods based on local information have proved useful in solving this problem<sup>25</sup>: compensation techniques such as adaptive thresholding<sup>24,25</sup>, and the contrast-based thresholding in Liu *et al.*'s “Local Stimulus”<sup>18</sup> have been effective at reducing the effect of differing background levels. In these systems an image is divided up into sections and the detection threshold is set independently for each region. Due to the success of this approach (as described in recent work<sup>18,24,25,27</sup>), our feature detection method uses adaptive thresholding (within the framework of Liu *et al.*'s “Local Stimulus”<sup>18</sup>) to compensate for varying background intensity.

Another challenge is “region bridging”, defined as the partial overlap of two intensity regions along a small portion of their boundary. In some circumstances (*e.g.* low-resolution input data and/or input images that contain greatly varying band width due to complex scattering source structure) small groups of high intensity pixels will form in regions of overlap between two distinct regions. This can cause two separate intensity bands to be classified as one continuous region, greatly (and erroneously) altering the final parameter extraction.

Wang and Yuan demonstrate an effective method for separating partially blended (*i.e.* weakly connected) regions based on the number of shared boundary pixels<sup>40</sup>; only if the number of pixels linking two regions is greater than a set threshold will two regions be merged into a single region. Wang and Yuan's technique effects a specialized form of the “erosion” and “opening” operators, commonly used to separate weakly connected image regions in binary morphology problems<sup>24</sup>. We use a similar bridge-eroding method in the feature detection and clustering stages of our pipeline to mitigate the effect of region bridging.

In addition to the problems of feature detection and clustering, there is the additional challenge of extracting a relevant numerical parameter set from the segmented images (*i.e.* extracting “region properties”<sup>24</sup>). Contiguous and homogeneous regions must be extracted as numerical entities for later parametric analysis. We use a form of localized communica-

tion (based on the widely used classical connected component labeling<sup>24</sup>) to organize the detected image features into a set of regions. These regions are parsed to extract a set of useful image parameters. As shown by the work of Sem'yanov *et al.*, Ulanowski *et al.*, and Maltsev, once an input image has been reduced to parametric form it is possible to infer some information regarding the internal structure of the scattering source from the extracted parameter values; mathematical relations and supervised learning algorithms were previously used to determine cell size and hemoglobin concentration from the parametric profile of scattering intensity slices<sup>3,9,11</sup>.

In this chapter we present a computational intelligence parameterization method as the first step in a parametric solution to the inverse scattering problem for laser light through a biological cell. Our method combines and builds upon a series of successful image processing methods (image segmentation<sup>25,30,44</sup>, multi-agent systems<sup>18,27,40,41,45</sup>, and computer vision<sup>24</sup>) to identify and group samples of local texture information into high-level patterns (*e.g.* semantic information such as intensity band location and structure). While our system is designed for cytometry problems involving vertical intensity bands with randomly distributed high-frequency components, its modular analysis pipeline, numerical representation of regions, and independent parameterization routine (all described in the following sections) give it the flexibility to be easily adapted for a variety of other cytometric image analysis situations (*e.g.* those with different band structures/orientations and/or arbitrarily shaped intensity regions).

To the best of our knowledge, our technique is the first computational system designed to comprehensively parameterize full wide-angle scattering signatures. We show that our system is able to identify the overall structure and relationships present in a scattering image. The resulting parameterization scheme, built from the numerical characterization of intensity bands and independent intensity blobs, can be used in the identification of cellular morphology. The end goal of this work is to facilitate the rapid division of experimental samples into healthy and diseased categories for expedient medical diagnosis. A pattern recognition system to infer detailed cellular structure from image parameterization values will be presented in future work (in preparation).

## 6.4 The Computational Pipeline

We present a computational intelligence pipeline (called Cythe) to effectively segment, cluster, and parameterize cytometric scattering images. The problem can be described as follows:

1. Given an input scattering image  $\mathbf{I}$  of size  $u \times v$ , where each of the  $u \cdot v$  pixels represents an 8-bit grey-scale intensity value, how can we effectively segment the image into its component intensity bands and sub-band regions?
2. Furthermore, once the salient features of the image have been identified, how can we extract relevant parametric information from these features and use this information to categorize the initial input image  $\mathbf{I}$ ?

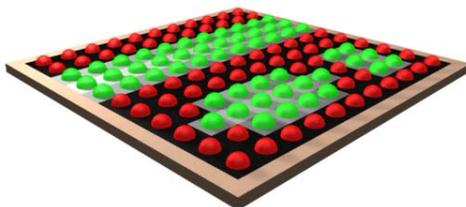
Previous work has shown viable two-stage image segmentation systems: in stage one all salient pixels are labeled with one or more identifiers; in stage two all labeled pixels are clustered and grouped according to some similarity or congruency metric<sup>24,25,29,31,32</sup>. In this work we use an additional stage to organize the clustered regions and extract a set of relevant parameters.

After performing an initial user-specified image size reduction, the first stage of our pipeline (*feature detection*) is responsible for creating and fixing a population of computational agents ( $\mathbf{A}$ ) to the salient features of the target image (as in the approach of Liu *et al.*<sup>18,27</sup>). This stage effectively labels all the pixels corresponding to relevant intensity regions; an explanation of saliency determination will be presented in the following sub-section. Stage two (*feature clustering*) is responsible for clustering the fixed agent population ( $\mathbf{A}$ ) into a list of spatially distinct regional groups ( $\mathbf{G}$ ). The final stage of the pipeline (*post-processing*) removes large-scale image noise, creates a band-like grouping structure from identified regions, and extracts a parametric representation ( $\mathbf{P}$ ) of the input data.

Detailed explanations of each stage are presented in the following subsections, which also describe the parameterization equations and the four major algorithms used to implement the individual stages of the pipeline. These are: the `agent_fixation()` routine, which is responsible for fixing the agent population to the salient image features; the `propagate_id()` routine, responsible for clustering agents into connected groups; the `scrub()` routine, which removes image noise and erroneous groupings; and the `join()` routine, which joins groups into a band-like structure of super-groups. A frame from an animated example of the complete pipeline is presented in Fig. 6.3.

### 6.4.1 Feature Detection

The first stage of the Cythe parameterization pipeline takes the input image  $\mathbf{I}$ , scales it to user specified dimensions  $u \times v$ , renders the image as a two-dimensional array<sup>37</sup>, and creates an agent population  $\mathbf{A}$  equal in size to the number of pixels in the image grid. A single



**Figure 6.3:** Frame from an animated movie of the complete Cythe pipeline processing an example 10 pixel by 10 pixel image. Agents are represented by colored hemispheres—green indicates ‘fixed’ agents, while red indicates ‘unfixed’ (dead) agents.

agent is assigned to every pixel in the image grid. These agents then use the information available in their local neighbourhood to detect features and sort themselves into regions; this is the standard approach used in most agent-based image processing systems<sup>18,27,30,41</sup>. To proceed we must elaborate on several definitions:

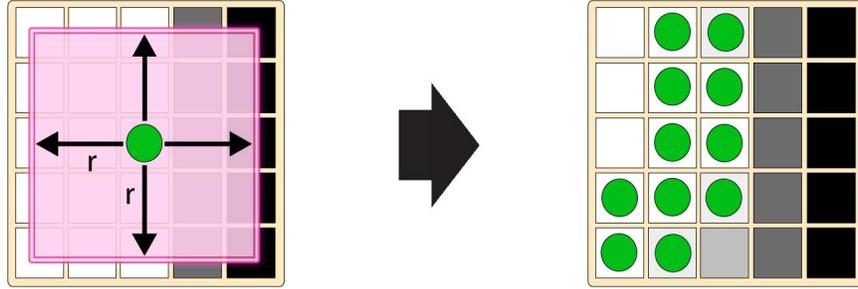
*Definition (Agent):* An agent is a single computational unit that is assigned to a pixel or region of the image grid  $\mathbf{I}$ . Each agent has a number of internal states and potential actions, and can alter these internal states and/or perform actions based on the information present in a localized area of the image grid  $\mathbf{I}$ .

*Definition (Agent Neighborhood):* The agent neighborhood  $\mathbf{N}$  is a  $n \times n$  region of the image grid  $\mathbf{I}$  centered on the agent location  $x_a, y_a$ . This region determines where an agent will look for and communicate with other agents (as in Liu *et al.*’s “Neighbouring Region of an Agent”<sup>18</sup>).

*Definition (Agent View Radius):* The agent view radius  $\mathbf{R}$  is a  $(2r + 1) \times (2r + 1)$  region of the image grid  $\mathbf{I}$  centered on the agent location  $x_a, y_a$  (Fig. 6.4, left). This area helps determine agent feature detection preferences, and the pixels within this area are used in the calculation of Average Pixel Intensity  $\mu_a$ . This is akin to the image region used in the “local binary pattern and contrast measure” of Ojala and Pietikainen<sup>30</sup> and the area used to acquire local stimulus by Liu *et al.*<sup>18,27</sup> and Mirzayans *et al.*<sup>41</sup>.

*Definition (Average Pixel Intensity):* This value, denoted  $\mu_a$ , is the average pixel intensity value that agent  $a$  observes within its view radius  $\mathbf{R}$ . Average Pixel Intensity is equivalent to the “mean pixel value” component of Liu *et al.*’s texture definition, as used in their multi-agent feature detection routine<sup>18</sup>.

During the feature detection stage of the pipeline, each agent calls on a fixation routine—`agent_fixation()`—to determine its immediate behavior<sup>18,41</sup>. When the fixation routine is called, the agent will perform one of two actions: an agent will affix to (and therefore identify as a salient region) a pixel at image grid location  $\mathbf{I}(x_a, y_a)$  if the pixel has an intensity



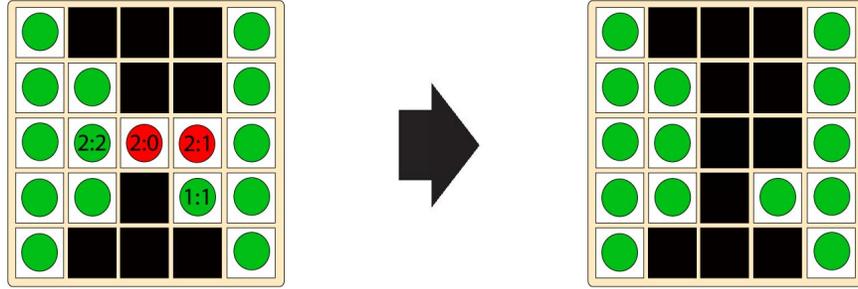
**Figure 6.4:** Agent fixation is determined by comparing the image intensity at an agent’s position to the average intensity ( $\mu_a$ ) within its view radius (left). After the `agent_fixation()` routine, members of the agent population will be fixed on areas of high intensity relative to the local image texture (right - shown here for an agent view radius of  $r = 1$ ). This adaptive process detects edges independent of differing background levels. Pixel color indicates 8-bit intensity, from 0 (black) to 255 (white).

value greater than the agent-computed average pixel intensity  $\mu_a$ , or, if this condition is not satisfied, the agent  $a$  will be removed from the agent population  $\mathbf{A}$ . In this way, agents are able to detect salient intensity edges in the image  $\mathbf{I}$  independent of differing background intensity values (*i.e.* a fixation routine based on an agent’s relation to its “average pixel intensity” is an adaptive thresholding function, as described by Pal and Pal<sup>25</sup>).

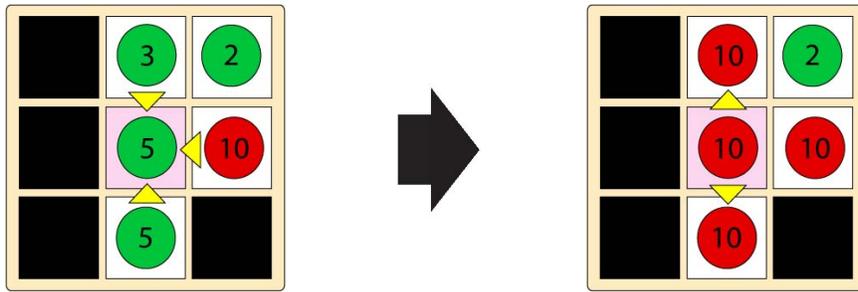
After the entire agent population has been polled for fixation, only agents that reside on salient pixels will remain in the population<sup>41</sup> (Fig. 6.4, right). To aid in effective region segmentation, we then scan the entire fixed agent population and remove all agents with more horizontal neighbours than vertical neighbours. Recalling the vertical nature of the intensity bands present in our scattering images (as discussed in the introductory section), we see that this helps eliminate any horizontal intensity ‘bridges’; much like a horizontally selective version of the “opening” operator used in binary image analysis<sup>24</sup>, the removal of these weakly connected ‘bridges’ facilitates region discrimination (as shown by Wang and Yuan<sup>40</sup>). An example of the removal process is shown in Fig. 6.5.

### 6.4.2 Feature Clustering

Once the agent population has completely labeled all relevant pixels in the image grid, a clustering process—`propagate_id()`—takes over to form the population  $\mathbf{A}$  into a set of spatially distinct regions  $\mathbf{G}$  (*i.e.* it links all adjacent agents to create spatially connected sub-regions). `propagate_id()` is a form of the classical connected components labeling algorithm<sup>24</sup>, traditionally used to identify spatially connected image features. Each time the `propagate_id()` routine is called, a sweep is done over the entire agent population; each agent in the population polls all other agents in its local neighbourhood ( $\mathbf{N}$ ) for their cur-



**Figure 6.5:** An example of horizontal bridge removal (before and after removal, left and right respectively), following the agent fixation shown in Fig. 6.4. Green circles indicate fixed agents. Red circles represent agents that will be removed, severing the connection between the two minimally connected bands. Numbers inside the pixels represent the ratio of horizontal to vertical neighbours within the 4-neighbourhood of a given agent (H:V).



**Figure 6.6:** Two parts of a single `propagate_id()` cycle for an active agent (center pixel). Initially, the agent surveys its local neighbourhood and records the ID values of its neighbours (left). Seeing there is a higher ID in the area (shown in red), it takes on this ID value and re-broadcasts the new ID to its neighbourhood (right). This leads to an agent neighborhood that is homogeneous with respect to the highest ID value.

rent ID value. Based on its initial scan, an agent records the highest ID value,  $id_{max}$ , in its local neighborhood. The agent then re-propagates the value  $id_{max}$  to all neighbors with ID values less than  $id_{max}$ , and the receiving agents take on the new maximum ID value. The entire agent population is iterated through until no further ID changes are observed<sup>30</sup>. At this point all agents in a separate physical region will share a unique ID number. A single iteration of the propagation process is shown graphically in Fig. 6.6.

It is important to note that ID propagation occurs in an agent’s 4-neighbourhood (*i.e.* to agents left, right, above, and below the agent, but not on diagonal corners<sup>24</sup>). This aids in band discrimination and removes additional band bridges. Due to the close horizontal proximity of bands in the scaled image **I**, it was found that communication within an agent’s full 8-neighbourhood could lead to a number of intensity regions being erroneously grouped into a single region. Allowing diagonal communication between agents did not facilitate any useful connections beyond those gained through purely horizontal and vertical transmission.

Since every agent starts with a unique ID value, the clustering process guarantees that every connected image region will have a common unique identifier<sup>24</sup>; we can now form a set of agent groups ( $\mathbf{G}$ ), where each group ( $\mathbf{G}$ ) contains a list of spatially connected agents ( $\mathbf{A}_g$ ) that share a common ID value ( $\mathbf{A}_g$  is a non-overlapping subset of the initial population  $\mathbf{A}$ ).

### 6.4.3 Post-Processing

As in previous work, a feature detection stage followed by a clustering stage is able to effect image segmentation. However, to utilize (and parameterize) the detected regions in the context of scattering image analysis, we require a third stage to organize and parse the segmentation results.

After the creation of homogeneous ID regions, several post-processing routines take over to remove high-level noise, join vertically correlated regions into a band hierarchy (*i.e.* create super-groups out of related regions), and extract the final parametric representation of the input file. The first process—`scrub()`—searches through the list of agent groups  $\mathbf{G}$  and removes all groups (and their member agents) smaller than a given percentage of the image size from  $\mathbf{A}$  and  $\mathbf{G}$  respectively; the removal size threshold  $\beta$  can be empirically set by human users to match the input image conditions. This method of removing small connected objects was used by Prasad *et al.* to eliminate background noise in their cellular tracking system<sup>44</sup>.

Each group that survives the `scrub()` routine is then analyzed for its dimensions and center point  $(g_x, g_y)$ . This effects a simple geometric characterization of all surviving groups in  $\mathbf{G}$ .

Next, horizontally related regions are connected into band-like structures using the `join()` routine, a simplified variant of the standard one-dimensional k-means clustering algorithm<sup>46</sup>. As in the k-means algorithm, `join()` creates list of super-groups and assigns one or more image regions  $g$  to each super-group  $g'$  based on the horizontal distance  $d = |x_{g'} - x_g|$  between the group center and the super-group center. Assignment occurs if  $d$  is less than a user defined threshold  $\delta$  (specified as a percentage of the image size), and each group may be assigned to only one super-group. A super-group's center is iteratively re-calculated based on the location of its member sub-groups. The `join()` process continues until every group has been assigned, clustering image regions with respect to their horizontal proximity. This allows the recognition of vertical bands in a scattering image while still retaining the detailed statistics of each individual sub-group. As such, `join()` creates a region hierarchy out of the agent population which can be stored at minimal cost for later retrieval and parameter estimation.

**Table 6.1:** The set of useful image parameters (**P**)

#	Parameter	Description
1	$B$	The number of bands in the image
2	$BS_{min}$	The minimum band spacing
3	$BS_{max}$	The maximum band spacing
4	$BS_{avg}$	The average band spacing
5	$aBW_{min}$	The average over all bands of minimum band width
6	$aBW_{max}$	The average over all bands of maximum band width
7	$aBW_{avg}$	The average over all bands of average band width
8	$aBW_{dev}$	The average over all bands of band width deviation
9	$aBI_{min}$	The average over all bands of minimum band intensity
10	$aBI_{max}$	The average over all bands of maximum band intensity
11	$aBI_{avg}$	The average over all bands of average band intensity
12	$aBI_{dev}$	The average over all bands of band intensity deviation
13	$aBI_{nn}$	The avg. over all bands of nearest-neighbour band intensity dev.

#### 6.4.4 Parameterization

In the last step of the Cythe pipeline, the super-group hierarchy is traversed and cross-referenced with the initial image **I** to extract a number of useful global parameters **P** (shown in Tab. 6.1). These parameters describe the overall structure and inherent complexity (in terms of spatial frequency components) of the image **I**, and are used to numerically represent the image features generated by light scattering through a biological cell (*i.e.* the number of regions, their size/shape, their relation to each other, and the variance of region width and intensity).

While finding a direct correlation between scattering signatures and the initial model parameters of a FDTD simulation or the structure of a cell has been shown to be an unsolved problem<sup>3</sup>, the parameters in **P** allow us to infer structural information from the presence of intensity regions with varying spatial frequency. The knowledge that certain cellular structures will generate intensity regions of a given spatial frequency allows relationships to be made between the extracted image parameters **P** and the initial layout of the scattering source. From our initial experiments, it is expected that there will be direct correlations between these parameters and the underlying cell model parameters; we have found that this is true for relations between small organelle content and several  $aBI / aBW$  parameters (work in preparation).

In this case, each super-group  $g'$  extracted by the Cythe pipeline corresponds to a detected intensity band ' $b$ ' in the scattering image. Based on previous cytometry work and FDTD simulation experiments<sup>1,2,15</sup> (which demonstrate the presence of vertical intensity bands in our scattering images) we found it most effective to use a band-based parameterization

scheme. In this approach, the small high-frequency intensity areas resulting from smaller scattering centers are effectively described by variations to the width and intensity of existing intensity bands (*i.e.* the presence and magnitude of high-frequency intensity fluctuation is indicated by changes to parameters 5–13, Tab. 6.1). A similar parameter set could be created for images without an observable band-like structure.

These parameters are extracted from the final super-group hierarchy through a series of mathematical operations, shown in Eqs. (6.1) and (6.2) below. Every detected super-group is analyzed with Eqs. (6.1) and (6.2), and the resulting values are combined into the set of parameters  $\mathbf{P}$ . Width statistics are derived by iterating through the agent population, intensity statistics are derived by taking a single-pixel wide intensity sample down the vertical center line of each super-group, and band spacing statistics are generated by comparing the horizontal centers of all super-groups.

$\min_x^b(y)$  and  $\max_x^b(y)$  are defined as the minimum and maximum angular values that still contain pixels belonging to band  $b$  at the vertical image position  $y$ . The function  $\text{intensity}(x^b, y)$  is the 8-bit intensity value at the horizontal center point  $x$  of band  $b$ , at the vertical position  $y$ . Set  $\mathbf{Y}^b$  is the set of vertical values for band  $b$ . The functions  $\min()$ ,  $\max()$ , and  $\text{avg}()$  are the standard minimum, maximum, and average operations performed on the list of values for a band. Band spacing ( $BS$ ) is defined as the distance between the horizontal centers of two neighbouring bands:  $|x^b - x^{b+1}|$ . Values for the maximum, minimum, and average band spacing are calculated using the standard operations.

$$BW^b(y) = \max_x^b(y) - \min_x^b(y) \quad (6.1a)$$

$$BW_{min}^b = \min(BW^b(y), y \in \mathbf{Y}^b) \quad (6.1b)$$

$$BW_{max}^b = \max(BW^b(y), y \in \mathbf{Y}^b) \quad (6.1c)$$

$$BW_{avg}^b = \text{avg}(BW^b(y), y \in \mathbf{Y}^b) \quad (6.1d)$$

$$BW_{dev}^b = \frac{1}{|\mathbf{Y}^b|} \sum_{y \in \mathbf{Y}^b} |BW^b(y) - BW_{avg}^b| \quad (6.1e)$$

$$BI^b(y) = \text{intensity}(x^b, y) \quad (6.2a)$$

$$BI_{min}^b = \min(BI^b(y), y \in \mathbf{Y}^b) \quad (6.2b)$$

$$BI_{max}^b = \max(BI^b(y), y \in \mathbf{Y}^b) \quad (6.2c)$$

$$BI_{avg}^b = \text{avg}(BI^b(y), y \in \mathbf{Y}^b) \quad (6.2d)$$

$$BI_{dev}^b = \frac{1}{|\mathbf{Y}^b|} \sum_{y \in \mathbf{Y}^b} |BI^b(y) - BI_{avg}^b| \quad (6.2e)$$

$$BI_{nm}^b = \frac{1}{|\mathbf{Y}^b|} \sum_{y \in \mathbf{Y}^b} |BI^b(y) - BI^b(y-1)| \quad (6.2f)$$

There is a dramatic increase in the amount of information available when we compare the number of values in this extended parameter set to the number of indicatrix parameters derived from one-dimensional scattering intensity slices. We expect this increase in parametric image information will lead to a corresponding increase in the predictive power of future classification systems. Intensity band relationships (such as band spacing  $BS$ , Tab. 6.1) can be used to predict the nature of larger cell structures<sup>9</sup>, while variations in region width and region intensity due to high-frequency image components (Params. 5–13, Tab. 6.1) may be used to detect the presence and number of micro- and nano- scale cellular organelles (work in preparation).

The final step in any automated diagnostic system is a method to deduce cellular structure from the extracted scattering pattern parameters  $\mathbf{P}$ . As described in earlier chapters, there are a number of potential machine learning approaches that can be used to associate extracted parameters with a labeled dataset to create a classifier with predictive power<sup>46,47</sup>.

## 6.5 Analysis Methods

We employed two testing methods to verify the validity of the CytHe system: qualitative image analysis, and a quantitative statistical breakdown. For our qualitative analysis we presented the system with images representative of all three cellular scattering cases described in Sec. 6.1 (*e.g.* intensity bands with a number of randomly placed intensity regions, as in Fig. 6.2, right). Due to the difficulty surrounding quantitative segmentation analysis, our statistical breakdown was performed on images containing the first two scattering cases (intensity bands and bands with interference). This is explained below. In both cases our test images closely matched experimental scattering patterns<sup>1,2</sup> and numerical FDTD simulations<sup>12</sup>, both visually and in the magnitude of the output parameters.

In an ideal testing environment we would be able to use FDTD simulations and experimental data to verify the success of our segmentation system. However, to numerically analyze system accuracy it is necessary to identify the ‘true’ segmentation and parameterization of experimental data. As ‘true’ image boundaries are subjective in all but the simplest segmentation problems, most segmentation evaluation methods rely on qualitative boundary assessments for comparison values<sup>48</sup>; the few attempts at true quantitative evaluation typically rely on correlation data, and still involve comparisons with a manual (*i.e.* human) segmentation<sup>25,32,48</sup>.

Thus, to quantitatively verify the validity of the Cythe extraction pipeline we used a mathematical model to create a set of viable test images. These images contained a fixed number of vertical intensity bands of varying intensity and width, irregularly placed high-frequency intensity components, intensity band overlap, differing background levels, blurring, and poorly defined band boundaries (*i.e.* qualities we observed in experimental scattering images). Unlike manually measured experimental scattering patterns, these model images were numerous and provided a well defined set of ‘true’ parameter values (derived directly from our mathematical image model) with which to statistically validate Cythe’s parameter extraction.

Despite this, it was still difficult to objectively define the ‘true’ band width values. As bands are represented in our test images by smooth intensity gradients with no discrete edges, the ‘true’ band width parameter ( $BW^b(y)_{real}$ ) was measured as the horizontal distance between band points where the pixel intensity was 80% of the band’s maximum intensity, relative to a black background. This width most accurately reflected observations about real scattering band width. Because of this approximate edge value definition, the validation data for band width parameters is slightly less precise than for other parameters, as seen in the following section.

These quantitative test images contained a more regular distribution of high-frequency intensity components than was found in experimental images or our qualitative analysis images; high-frequency intensity regions were randomly placed only on intensity bands, as in image cases one and two, Sec. 6.1. This additional regularity was needed generate reliable true values for band parameterization—images containing a completely random distribution of high-frequency regions (as expected from Rayleigh scattering, image case three) would suffer from the same subjective evaluation problems as real experimental data.

Thus, each quantitative test image consisted of a varying number of Gaussian intensity regions superimposed on a series of vertical intensity bands. Like real scattering patterns, our test images contained bands of varying width and maximum intensity that were placed at intervals across a black background. The intensity profile of individual bands, the size

and orientation of Gaussian intensity regions, and the variation of maximum band intensity across the image were picked to match the intensity profiles expected in actual scattering images. Finally, a  $5 \times 5$  Gaussian blur was applied to the images to smooth out any unrealistic intensity transitions.

These test images were then presented to the Cythe system for analysis. Each test image was processed by the full computational pipeline (*i.e.* feature extraction, feature clustering, and post-processing) to produce a set of output parameters ( $P_{\text{cythe}}$ ). Another set of parameters were derived directly from the mathematical model used to generate the test images; these parameters ( $P_{\text{real}}$ ) represented the ‘true parameter values’ used in the creation of the test images. We then inspected how well the true parameters  $P_{\text{real}}$  matched the parameter values extracted by our pipeline  $P_{\text{cythe}}$  (*i.e.* how well they demonstrated a correlated linear relationship that allowed accurate prediction of the true parameter values). Both the true parameter set  $P_{\text{real}}$  and Cythe parameter set  $P_{\text{cythe}}$  included all thirteen parameters outlined in Section 6.4. As explained in the previous section, this band-based parameterization scheme (calculated with Eqs. (6.1) and (6.2)) can be used to represent the influence of both large scattering structures and nanostructure-derived high-frequency intensity variation.

To assess the pipeline’s ability to detect changes in band width and band intensity, these tests were performed on 162 sample images. Two different test sets were generated. The first set ( $T_1$ : 143 images) was used to determine the system’s ability to detect variation in band width and band intensity—parameters primarily influenced by the presence of smaller scattering sources. In this set the number of intensity bands was held constant while the number of Gaussian intensity regions present in the image was varied between zero and fifty. The second set ( $T_2$ : 19 images) was used to test the system’s ability to detect changes in band structure and spacing, which relate to scattering from larger microstructural cellular objects. In test  $T_2$ , the number of intensity bands was varied between two and twenty, while the number of Gaussian regions inserted into the image was held constant.

After each test set the Cythe parameter extractions were compared to the true parameters. System success was determined by measuring how closely the Cythe parameters matched the true parameters, as evaluated with a range of statistical tests for correlation and similarity (described in the following section).

This procedure is similar to the comparison metric of Bovenkamp *et al.*, where the plot of human v.s. machine solutions was compared to a unity slope to determine accuracy<sup>21</sup>. In the absence of any methods to objectively compare and evaluate segmentation schemes<sup>25,32,48,49</sup>, this approach allowed a quantitative characterization of system success.

In addition to these quantitative test images, we performed a series of qualitative tests on a range images of containing features from all three scattering image cases presented in Sec. 6.1 (*e.g.* Fig. 6.2, right). This was done to determine the system’s ability to remove region bridges, detect regions in noisy images, and join detected regions into a structural hierarchy. For these tests, we generated a test series consisting of images with vertical intensity bands, randomly distributed high-frequency intensity regions, and vertical intensity bands overlaid with a pattern of randomly distributed high-frequency regions of irregular shape and size. Sample images and the corresponding results are presented in the following section.

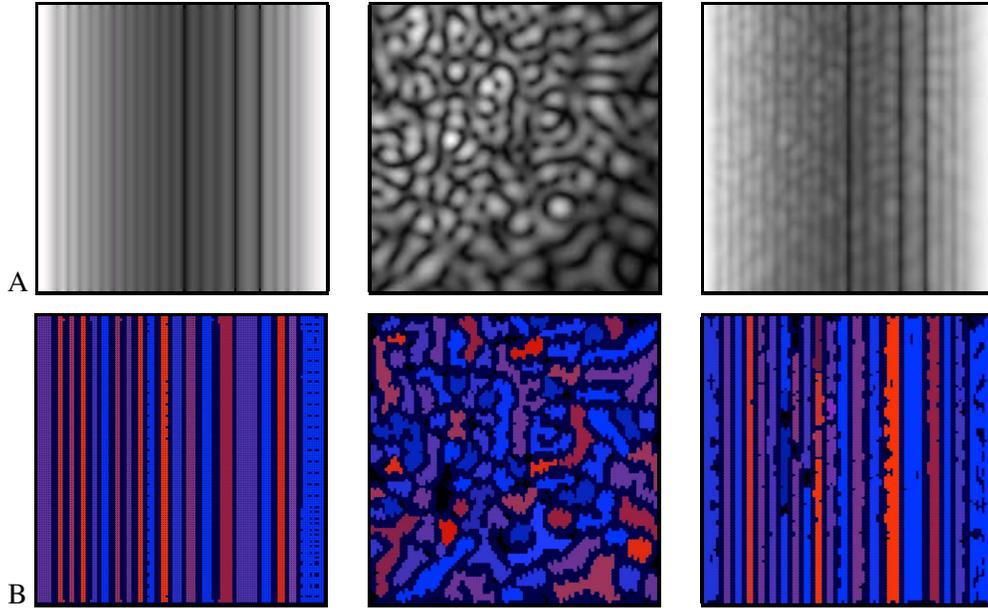
For these tests the scrub() and join() thresholds ( $\beta, \delta$ ) were set to 0.0018% and 0.028% respectively. The agent view radius was  $\mathbf{R} = 5 \times 5$ , and the agent neighbourhood was  $\mathbf{N} = 3 \times 3$ . Images were reduced to  $\mathbf{I} = 125 \times 125$ . These values were empirically derived on a small subset of the images, and subsequently used on the full set of test images without modification; one parameter setting performed well for an entire family of images.

## 6.6 Results

As a summary of the following performance assessment: the Cythe pipeline was able to detect relevant image features (*e.g.* intensity band pixels) in test images, remove horizontal region bridges, and cluster the detected features into a set of spatially distinct regions ( $\mathbf{G}$ ). These regions were then used to harvest a parametric representation ( $P_{cythe}$ ) of the initial image that directly matched the known parameters of the input image ( $P_{real}$ ). As described earlier, the parameter set  $\mathbf{P}$  serves to numerically capture the structure of both the large bands and small high-frequency intensity regions present in scattering images. This section will begin with qualitative verification results, and conclude with a quantitative numerical assessment of Cythe.

### 6.6.1 Qualitative Assessment

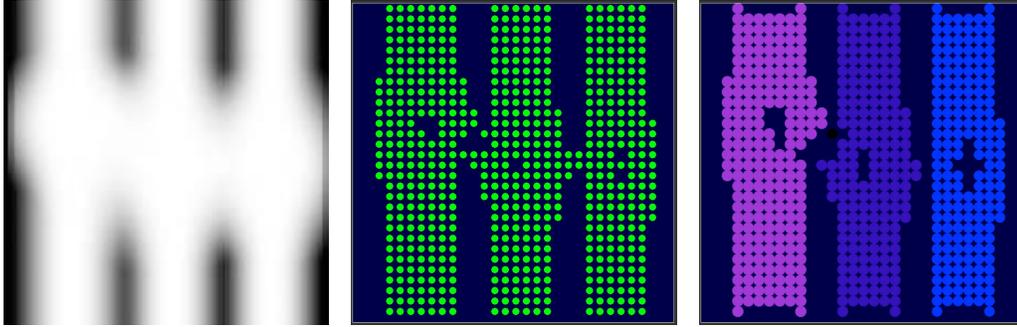
As shown in Fig. 6.7, the Cythe pipeline was able to affix the agent population ( $\mathbf{A}$ ) to the vertical intensity bands in the test image. A visual comparison of the Cythe labeling (Fig. 6.7, right) with the initial image (Fig. 6.7, left) showed that the Cythe extraction matched quite closely with our expectations from test image. We also observed that the system was able to detect the presence and magnitude of width variations in the band structure (Fig. 6.7, bottom row).



**Figure 6.7:** A visual comparison of Cythe region detection (row B) with the initial test image (row A) for images with vertical intensity bands (left), high-frequency intensity regions (middle), and high-frequency regions overlaid onto a band structure (right, similar to those observed in FDTD simulations). Region color was assigned based on each group’s unique ID value; all regions were verified to contain distinct ID values.

In addition to being able to detect linear bands, Cythe was able to detect small, arbitrarily shaped intensity regions of varying brightness (Fig. 6.7, middle). As shown by the difference between Fig. 6.7 left and right, Cythe was also able to detect the level of high-frequency variation present in images containing high-frequency components that overlap a pattern of vertical intensity bands. This observation further supports the efficacy of the band-based parameterization scheme **P**. Random intensity regions (like those expected from Rayleigh scattering) were indicated by width and intensity deviations within the detected band structure—their intensity contributed to, and noticeably altered, the shape of existing bands.

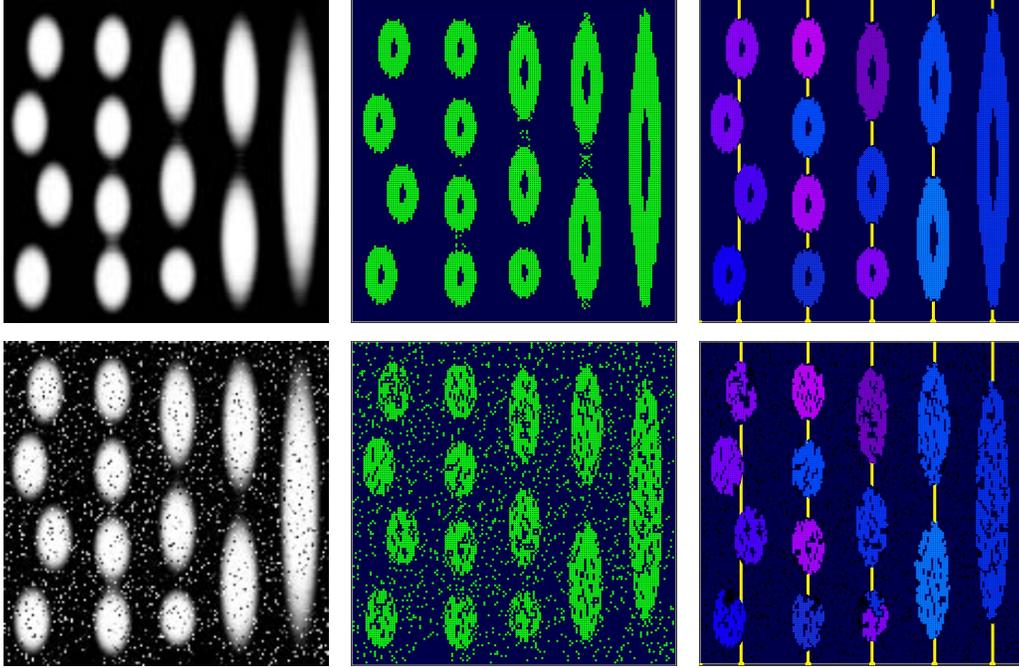
We found that Cythe was able to remove parameter-degrading horizontal intensity bridges and use the clustering stage to group the agent population (**A**) into a set of distinct regions (**G**). The removal of horizontal bridging can be seen in Fig. 6.8, and the ability to form a population into spatially connected regions can be seen by the homogeneous region colors in Figs. 6.7 and 6.8. As shown in the difference between the two agent populations in Fig. 6.8 (middle and right), we found that horizontal bridges less than three pixels in width were removed during the feature detection stage. In addition, the use of a 4-neighbourhood for communication in the feature clustering stage prevented distinct bands from being classified as a single region due to any remaining weak connections (Figs. 6.7 and 6.8, right).



**Figure 6.8:** A visual example of horizontal bridge removal. LEFT: the initial test image. MIDDLE: the agent population directly after the `agent_fixation()` routine; there are three bridges at this point. RIGHT: final region identification after post-processing; weak connections between bands did not adversely affect region identification—the two horizontal bridges were removed in the feature detection stage, and the diagonal propagation restriction prevented ID leaking over the remaining bridge (which was subsequently removed by the `scrub()` routine). Green dots represent fixed agents (middle), and different colors in the clustering image indicate spatially distinct regions (right).

The `join()` routine constructs a set of vertical bands  $g'$  out of the detected image regions  $\mathbf{G}$ . For noisy images this process would not be possible without prior use of the `scrub()` routine to filter out small unconnected intensity regions. Figure 6.9 illustrates the use of the `join()` and `scrub()` routines in the creation of a vertical band structure for simple images with and without 10% of the images pixels assigned a random 8-bit intensity noise value (*i.e.* random or independent noise, as expected from dust on a lens or bad CCD pixels). While portions of the agent population affixed to noise-related pixel clusters (6.9, bottom middle), the `scrub()` routine removed these small groups and the pipeline identified the same regions found in the noise-free image (6.9, right). In addition, the detected regions were joined into the same band structure for both the noisy and noise-free image (as shown by the number and horizontal position of the yellow vertical lines, Fig. 6.9, right). This lead to the same parameters being extracted for both the noisy and noise-free images. Similar performance was observed for Poisson/counting noise, though high levels of Gaussian noise required the use of a larger scrub threshold due to larger detected noise regions. A join threshold of  $\delta = 0.08$  was used for the tests in Fig. 6.9.

In addition to accurately parameterizing our model test images, Cythe was able to extract realistic parameters for a large set of FDTD scattering simulation images containing many arbitrarily shaped randomly distributed high-frequency intensity regions, as derived from complex cell structures with varying physical characteristics and organelle distributions (work in preparation).



**Figure 6.9:** Extraction of a band hierarchy for a simple noise-free image (top row) and for the same image with 10% of the images pixels assigned a random 8-bit intensity value (*i.e.* random noise; bottom row): the initial image (left), the agent population after the `agent_fixation()` routine (middle), and the final regions after post-processing (right). Yellow lines indicate band position ( $x_{g'}$ ), and coloured regions in the post-processing image indicate spatially distinct regions  $g$ .

## 6.6.2 Quantitative Assessment

We found that the parameters extracted by *Cythe* from the test images ( $P_{\text{cythe}}$ ) allowed us to accurately predict the true parameters  $P_{\text{real}}$  extracted from the initial test images. This was statistically determined by calculating the correlation coefficient ( $r$  – the amount of covariance in the two populations, a good indicator of segmentation accuracy<sup>25</sup>), the statistical significance of the correlation ( $P(r)$  – the probability that correlation value  $r$  could have arisen by pure chance for a given sample size), the chi-squared significance ( $P(\chi^2)$  – the probability of both input and output variables coming from the same distribution), and the standard error ( $SE$ ) for each population of input/output variables (all calculated as per Taylor<sup>50</sup>, using Eq. 12.11 and Sec. 12.4 on Pgs. 268, 271–275 for  $P(\chi^2)$ , Eq. 9.15 and Sec. 9.4 on Pgs. 217–220 for  $r / P(r)$ , and Eq. 4.14 on Pg. 102 for  $SE$ ).

This comparison is shown in tabular form for tests  $T_1$  (band intensity parameters, Tab. 6.2, and band width parameters, Tab. 6.3) and  $T_2$  (band number/spacing parameters; Tab. 6.4). These tables present the statistics for an image reduction size of  $\mathbf{I} = 125 \times 125$ . From statistical theory<sup>50</sup>,  $r$  values greater than 0.216 (test  $T_1$ , 143 samples) and 0.561 (test  $T_2$ , 20

**Table 6.2:** Statistical analysis for band intensity parameters

Parameter	Description	$r$	$P(r)$	$P(\chi^2)$
$aBI_{avg}$	Avg. Band Intensity Average	0.992	<0.0001	1.000
$aBI_{min}$	Avg. Band Intensity Minimum	1.000	<0.0001	1.000
$aBI_{max}$	Avg. Band Intensity Maximum	0.998	<0.0001	1.000
$aBI_{dev}$	Avg. Band Intensity Deviation	1.000	<0.0001	1.000
$aBI_{nn}$	Avg. Band Intensity Deviation (NN <sup>1</sup> )	1.000	<0.0001	1.000

<sup>1</sup> nearest neighbour.

**Table 6.3:** Statistical analysis for band width parameters

Parameter	Description	$r$	$P(r)$	$P(\chi^2)$
$aBW_{avg}$	Avg. Band Width Average	0.386	<0.0001	1.000
$aBW_{min}$	Avg. Band Width Minimum	0.872	<0.0001	0.528
$aBW_{max}$	Avg. Band Width Maximum	0.724	<0.0001	1.000
$aBW_{dev}$	Avg. Band Width Deviation	0.907	<0.0001	0.286

samples) indicate a statistical correlation (*i.e.* a probability  $P(r) < 0.01$  that the correlation score could have originated by pure chance). These threshold  $r$  values are based on the sample sizes used in our experiments. As shown in Tabs. 6.2–6.4, our derived values are consistently greater than these minimum values for statistical correlation. Similarly, chi-squared significance values approaching  $P(\chi^2) = 1.00$  indicate no difference in the distribution of input and output values.

The uncertainty in each parameter was estimated by adding Poisson/counting noise (*i.e.* each pixel was varied according to a normal distribution equal to the square root of the pixel value) and processing the resulting image by the same method as the test data sets. This was done for 56 images, allowing the extraction of a standard deviation that then allowed the calculation of chi-squared significance values.

As described in Sec. 6.4, the parameters shown in Tabs. 6.2 and 6.3 are used to characterize the intensity contributions from smaller Mie and Rayleigh scattering sources, while the parameters in Tab. 6.4 characterize the intensity contributions from larger Mie and geometric scattering objects.

With regard to the system’s ability to correctly identify deviations in band intensity (test  $T_1$ ), we found that Cythe was able to identify the magnitude and variance of intensity to a high degree of certainty. At an image size of  $\mathbf{I} = 125 \times 125$  Cythe was able to correctly identify the number of bands in every test image. The input and output intensity parameters ( $aBI_{min/max/avg/dev/nn}$ ) showed strong correlation, as indicated by the  $r$ ,  $P(r)$ , and  $P(\chi^2)$  val-

**Table 6.4:** Statistical analysis for band number/spacing parameters

Parameter	Description	$r$	$P(r)$	$P(\chi^2)$
$B$	Number of bands	1.000	<0.0001	1.000
$BS_{min}$	Minimum band spacing	0.995	<0.0001	1.000
$BS_{max}$	Maximum band spacing	0.993	<0.0001	1.000
$BS_{avg}$	Average band spacing	1.000	<0.0001	1.000

ues (Tab. 6.2). Standard error for these intensity parameters was less than half an intensity step on a 8-bit intensity scale.

The close relationship between input and output parameters was also evident for band width and band width deviation parameters ( $aBW_{min/max/avg/dev}$ ), as shown by the values in Tab. 6.3. As explained in the previous section, difficulty defining the ‘true’ width values in the test images lead to greater variability in the evaluation statistics  $r$  and  $\chi^2$ . While width statistics (Tab. 6.3) did show lower correlation between input and output values than the other parameters (Tab. 6.2, Tab. 6.4), all other values represented an excellent fit. Width values were still well above the thresholds for chance correlation, as indicated by  $r > 0.261$ ,  $P(r) \ll 0.01$ . Despite having a high degree of correlation, the parameters  $aBW_{dev}$  and  $aBW_{min}$  exhibited a low  $P(\chi^2)$ , and further investigation showed that this deviation in input/output distribution similarity was due to a shallow (*i.e.*  $< 0.5$ ) regression slope between the input and output parameter sets. Considering the lack of ‘true-value’ precision when quantitatively analyzing spatial parameters in this situation, the set of width statistics in Tab. 6.3 sufficiently demonstrated a distinct relation between the actual layout of the test images and the Cythe parameter extraction.

In addition to band width and intensity parameters, we observed that the system was able to accurately determine the number and spacing of bands (test  $T_2$ ). As shown in Tab. 6.4, the correlation coefficient ( $r$ ) for each band-structure parameter approached 1.0 (*i.e.* perfect correlation). This indicates a one-to-one correspondence between the input parameters  $P_{real}$  and the output parameters  $P_{cythe}$ . For the parameters  $BS_{min}, BS_{max}, BS_{avg}$  there was a standard error of less than 1.1% of the image width for both reduction levels. There were no band number ( $B$ ) identification errors in test  $T_2$ , and the chi-squared significance test for all parameters in Tab. 6.4 indicated no statistical difference between the input and output parameters.

With respect to the magnitude of observed values from Eq. 6.1, we found a typical range of 0.72–9.6 pixels for  $aBW_{min/max/avg}$ , and 0.0–0.99 pixels for  $aBW_{dev}$ . For Eq. 6.2, intensity parameter values were typically between 127.2–157.2 for  $aBI_{min/max/avg}$ , 0.59–4.29 for  $aBI_m$ , and 0.26–72.7 for  $aBI_{dev}$ . Band spacing parameters varied greatly depending on the number of detected intensity bands in a sample image; for our tests we found spacing

parameters between 6.46–14.8 (Test T1) and 6.48–72.2 pixels (Test T2). The standard deviation of parameter values observed under conditions with counting noise and random noise was much less than the total parameter range observed during these tests.

As the size of images presented to the system increased (with the Agent View Radius being held constant), we found that Cythe began to identify small erroneous bands within the larger regions of the image. At an image size of  $\mathbf{I} = 150 \times 150$  the pipeline incorrectly identified one extra band in 67 of the 143  $T_1$  tests images. This led to a noticeable decline in correlation values, and incurred a corresponding increase in standard error. It is apparent that the relationship between image size and Agent View Radius plays a role in feature detection; this will be discussed in the following section.

## 6.7 Discussion

### 6.7.1 Remarks on Feature Detection

The success of the Cythe feature detection system is in a large part due to the use of regional texture information to affect agent fixation. We chose to use an adaptive local thresholding method based on its success within other texture-based segmentation problems and its compatibility with agent-based image processing (as shown by a large body of previous work<sup>18,20,27,30,41</sup>).

Edge detection is by its very nature a local undertaking<sup>25</sup> and thus lends itself well to an agent-based framework. By determining fixation based on an adaptive local threshold ( $\mu_a$ , the average intensity value within an agent's view radius  $\mathbf{R}$ ), Cythe was able to effectively label all edges irrespective of the differing background intensity levels found in scattering images. By setting the adaptive threshold level greater than the local average (as in the `agent_fixation()` routine), the system consistently labeled the high-intensity side of all edges, isolating the band regions from the lower-intensity background.

Using an Agent View Radius of  $\mathbf{R} = 5 \times 5$ , we found  $\mathbf{I} = 125 \times 125$  to be the most appropriate size for image reduction. At this image size and view radius, the fixation routine was able to accurately divide the image into spatially distinct regions regardless of differing background levels and gradient slopes. The distance between identified band edges was small enough that the two edge-labeling agent populations for a given band connected along the center of their band. This allowed bands to be detected as continuous units in both our model test images and complex FDTD scattering simulations.

We found that it was important to select a view radius close to the size of target image features in the reduced image  $\mathbf{I}$ . The two edge regions of a single band may not connect if the Agent View Radius is significantly smaller than the band size. This led to the identification of extra bands by the clustering stage. By varying the size of the view radius (*i.e.* the adaptive thresholding region<sup>25</sup>) to match the image reduction level, feature extraction remains accurate at any image size (though increased image size comes with an increased computational cost, as described below). This follows from recent work in image saliency detection and model matching<sup>33,51</sup>.

### 6.7.2 Remarks on Clustering

The `propagate_id()` routine was a reliable and effective way to cluster the labeled pixels into contiguous regions. This routine, which was based on the connected components labeling algorithm commonly used in region identification problems<sup>24</sup>, provided a simple way to identify groups of connected agents. Much like the self-organizing tile behavior shown by Ghrist and Lipsky<sup>45</sup>, our system was able to effectively perform long-range organization through simple local interactions. In addition, the distributed approach lends itself well to parallelization—one of the major advantages of multi-agent systems<sup>43</sup>.

The removal of band bridging (as described in Section 6.4) was essential in the accurate clustering of spatially distinct regions. The agent fixation stage eliminated direct horizontal communication over bridges by eroding bridging agents (shown above in Fig. 6.8), while the use of an agents 4-neighbourhood for communication prevented ID propagation over any remaining (weak) junctions between band protrusions. Without the removal of band bridging it was impossible to successfully parameterize complex images. A similar restriction on the union of weakly connected regions has proved effective in other segmentation and image identification situations<sup>24,40</sup>. The assumption that there will be no strong horizontal links between bands follows from the structure of experimental scattering images and our understanding of cellular scattering mechanisms.

Using an agent neighbourhood of size  $\mathbf{N} = 3 \times 3$  further prevented erroneous ID propagation between distinct bands. By only allowing communication between adjacent agents, ID information was not able to travel over gaps between neighbouring intensity bands.

Due to the nature of the local interactions and the multiple sweeps through the agent population, the ID propagation routine was found to be the largest computational component of the parameterization pipeline. As the proper choice of agent view radius and image reduction size decreased the final size of experimental images to approximately  $\mathbf{I} = 125 \times 125$ , scalability was not an issue for our application. In the case of larger images, the use of a

union-find structure (described in<sup>24</sup>) in the connected components algorithm would greatly reduce the number of iterations through the agent population (though it would require a higher level of centralized control).

### 6.7.3 Remarks on Post-Processing and Parameterization

The `join()` routine was found to be an effective way to model the structure inherent in scattering images. It has been shown that changes in the relationships between full bands are indicative of large changes in cellular structure<sup>9</sup>. By linking several vertically aligned regions to a single band structure, we were able to analyze the relationship between whole band units while still retaining specific information regarding the variation present in each band and its associated sub-regions.

As shown in the results section, the `join()` routine managed to consistently group smaller regions into cohesive bands, even in the presence of noise. Noisy images were divided into the same number of bands (*i.e.* super-groups) as noise-free images (Fig. 6.9). This is important for the parameterization stage of the pipeline; band discrimination plays a large part in the calculation of band-based statistics, which in turn contain vital information about the nature of the scattering source.

The `scrub()` routine helped the parameter extraction process by removing any large noise regions that remained after the initial image reduction. By keeping the scrub threshold low, large features were still preserved (*e.g.* Fig. 6.9) while small agent clusters were rejected as noise; this parameter can be tuned to the specific nature (ambient noise and feature size) of the images under analysis. However, it should be noted that setting the scrub level too low can cause erroneous band identification, whereby bands of small pixel mass may appear near the edges of each real band. Extra bands will distort the extracted parameter space and should be avoided.

With regard to the selection of the variables for `join()` and `scrub()` (*i.e.*  $\beta$  and  $\delta$ ): these values are derived empirically based on observations regarding the size of noise artifacts inherent in a scattering image and the approximate size and frequency of bands in the image. Once selected, the parameters performed effectively on the entire test set. If significant changes are made to the ambient background noise or the angular range of a scattering image, the system threshold levels may need to be re-calculated.

While there are a number of other potential frequency analysis methods (such as Fast Fourier Transforms) which could be used to ascertain spatial frequency information from a scattering image, the Cythe parameterization routine allowed the extraction of a related

structure of image regions within the context of a scattering situation (effectively embedding frequency information within an interpretable band-like structure). This relationship information is of use to human observers, both for validating the extracted parameter data and for comparing results with those previously published in the cell scattering literature (which generally reference the number and size of bands, or the angular location and span of intensity band maxima).

#### 6.7.4 Remarks on Image Size Reduction

Image size reduction was found to play an important role in both the generalization of region boundaries and the rejection of low-level noise, as it influences the degree of abstraction applied to the input image<sup>33,34</sup>. A similar reduction approach is used in saliency-based vision systems to detect high-level features in natural scenes, where detail (and the associated noise) is sacrificed to rapidly form an accurate structural impression of the image<sup>33,34</sup>.

Appropriate choice of image reduction size depends on the size of the Agent View Radius, and the number, spacing, and width of intensity bands within an image. A large reduction to an image with very narrow bands or important high-resolution features could merge independent intensity regions, or render some relevant features undetectable. Failing to reduce an image with wide bands could lead to erroneous band detection or extra computational cost / increased run times. We found that disparity between image reduction size and Agent View Radius either lead to the identification of too many small intensity regions (*e.g.* when the true  $abW_{avg} \gg \mathbf{R}$ ) or the grouping of many distinct initial regions into a small number of larger features (*e.g.* when the true  $abW_{avg} \ll \mathbf{R}$ ).

Image reduction was also essential for manageable run times, as un-reduced experimental cytometer images are typically greater than 700 pixels on a side. At an image reduction size of  $\mathbf{I} = 100 \times 100$ , an entire pipeline run took approximately six seconds. At a reduction size of  $\mathbf{I} = 300 \times 300$  or larger, runs lasted two minutes or more. All performance tests were conducted on a Pentium IV desktop computer. The entire Cythe pipeline and all related routines were implemented in the Python programming language.

#### 6.7.5 Remarks on Versatility

Cythe is expandable and may be readily adapted to new scattering situations; once intensity regions are segmented and numerically represented (as in the group list  $\mathbf{G}$ ) it is possible to test for any number of spatial relationships. In addition, the pipeline should be robust to variations in expected image structure. By adjusting the post-processing settings, agent

view radius, and image reduction size, Cythe can be made to detect intensity regions with greatly varying geometric properties. Furthermore, due to the local nature of the parameter calculation equations, slight band curvatures should not adversely affect parameter extraction.

In the event that images with different (*e.g.* non-vertical) spatial relationships need to be analyzed, the `join()` routine may be modified or replaced to create a different region hierarchy, and orientation-related changes may be made to the band bridge removal process. This would also allow identification of randomly placed intensity regions, such as would be generated by a field of Rayleigh scatters, without imposing a band structure on the intensity data. The additional analysis of intensity region perimeter and area would allow further distinctions to be made between differing region types (*e.g.* independent regions and full bands). This would facilitate the parameterization of heterogeneous images consisting of horizontal bands, tightly grouped region clusters, blobs arranged without a band structure, or any other arbitrary cluster shape.

To this end, we have used Cythe within other applications, including the identification of geometric objects in natural scenes, and the detection of bright fluorescent regions during the genetic analysis of cell populations<sup>52</sup>.

The final goal of the Cythe system is the classification of biological samples based on light scattering. Our preliminary results have shown that cell classes (*e.g.* those with features indicating cell health or malignancy) typically reside at the extremes of the possible parameter space (work in preparation). The difference between cell classes appears to be much greater than the variation due to noise, such as from imperfections in a fluid wave-guide or CCD. Thus, based on the parameter deviation indicated from random and counting noise (as discussed above), measurement noise should only moderately detract from Cythes classification ability and the correlations we observed between input and output parameters.

## 6.8 Conclusions

In this work we present a multi-agent system (Cythe) to parameterize laser scattering images of the kind produced by a wide-angle 2D cytometer. Extending upon a solid base of tested image processing methods, Cythe uses a three-stage pipeline of feature detection, feature clustering, and post-processing to create a parametric representation of an input scattering image. The resulting parameter set numerically represents the complex image features created by light scattering through a cellular body. This facilitates a parametric solution to the inverse scattering problem of laser light through a single biological cell.

Comparison of the Cythe-extracted parameter sets with those derived from a mathematical image model show that our pipeline is able to accurately extract information about the structure and variation present in an image. In addition to our model test images, Cythe is able to parameterize complex FDTD scattering images containing a number of randomly distributed high-frequency intensity regions. Cythe was also able to effectively extract information from images without a noticeable band structure, and has been successfully modified to help detect and parameterize fluorescent genetic material in populations of stained cells<sup>52</sup>.

This is possible through the combination of an adaptive feature detection system, an agent-based clustering scheme, and a set of post-processing routines that reject noise and extract high-level information about the relationships between image features. Once a parameter set has been extracted from a scattering signature, it is possible to infer cellular structure from regularities in the extracted parameters<sup>1,9-11</sup>. For example, we have observed correlations between organelle content in simulated cells and several of the intensity and width parameters present in Eqs. 6.1 and 6.2 (work in preparation). To date, no other group has developed an computational system to extract detailed parametric information from wide-angle cytometric scattering signatures.

A rapid method to infer cell characteristics from the information contained in 2D light scattering plots is essential to the further development of wide-angle cytometry systems. We have developed a method of predicting micro- and nano-structural cellular information from the parameters generated by Cythe. A system integrating Cythe with a machine learning classifier to characterize cellular organelle content will be presented in future work.

## References

- [1] K. Singh, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, "Analysis of cellular structure by light scattering measurements in a new cytometer design based on a liquid-core waveguide," *IEE Proc.-Nanobiotechnology* **151**(1), 10–16 (2004).
- [2] K. Singh, X. Su, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, "A miniaturised wide-angle 2-D cytometer," *Cytometry A* **69A**, 307–315 (2006).
- [3] V. P. Maltsev, "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [4] P. L. Gourley, "Biocavity laser for high-speed cell and tumour biology," *Journal of Physics D: Applied Physics* **36**(14), R228–R239 (2003).
- [5] P. L. Gourley and R. K. Naviaux, "Optical phenotyping of human mitochondria in a biocavity laser," *IEEE Journal of Selected Topics in Quantum Electronics* **11**(4), 818–826 (2005).

- [6] P. Chacon, F. Moran, J. F. Diaz, E. Pantos, and J. M. Andreu, "Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm," *Biophysical Journal* **74**(6), 2760–2775 (1998).
- [7] P. Chacon, J. F. Diaz, F. Moran, and J. M. Andreu, "Reconstruction of protein form with X-ray solution scattering and a genetic algorithm," *Journal of Molecular Biology* **299**(5), 1289–1302 (2000).
- [8] J. D. Watson and F. H. C. Crick, "Molecular structure of nucleic acids - a structure for deoxyribose nucleic acids," *Nature* **171**(4356), 737–738 (1953).
- [9] K. A. Sem'yanov, P. A. Tarasov, J. T. Soini, A. K. Petrov, and V. P. Maltsev, "Calibration-free method to determine the size and hemoglobin concentration of individual red blood cells from light scattering," *Applied Optics* **39**(31), 5884–5889 (2000).
- [10] N. Ghosh, P. Buddhiwant, A. Uppal, K. Majumder, H. S. Patel, and P. K. Gupta, "Simultaneous determination of size and refractive index of red blood cells by light scattering measurements," *Applied Physics Letters* **88**, 084,101 (3 pages) (2006).
- [11] Z. Ulanowski, Z. Wang, P. H. Kaye, and I. K. Ludlow, "Application of neural networks to the inverse scattering problem for spheres," *Applied Optics* **37**(18), 4027–4033 (1998).
- [12] C. Liu, C. E. Capjack, and W. Rozmus, "3-D simulation of light scattering from biological cells and cell differentiation," *Journal of Biomedical Optics* **10**(1), 014,007 (12 pages) (2005).
- [13] R. Drezek, A. Dunn, and R. Richards-Kortum, "Light scattering from cells: finite-difference time-domain simulations and goniometric measurements," *Applied Optics* **38**(16), 3651–3661 (1999).
- [14] R. Drezek, A. Dunn, and R. Richards-Kortum, "A pulsed finite-difference time-domain (FDTD) method for calculating light scattering from biological cells over broad wavelength ranges," *Optics Express* **6**(7), 147–157 (2000).
- [15] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, "2D light scattering patterns of mitochondria in single cells," *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [16] K. Sem'yanov and V. P. Maltsev, "Analysis of sub-micron spherical particles using scanning flow cytometry," *Particle & Particle Systems Characterization* **17**(5-6), 225–229 (2000).
- [17] N. Richard, M. Dojat, and C. Garbay, "Automated segmentation of human brain MR images using a multi-agent approach," *Artificial Intelligence in Medicine* **30**(2), 153–175 (2004).
- [18] J. Liu, Y. Y. Tang, and Y. C. Cao, "An evolutionary autonomous agents approach to image feature extraction," *IEEE Transactions on Evolutionary Computation* **1**(2), 141–158 (1997).
- [19] M. Schmidt, "Automated brain tumor segmentation," Master's thesis, University of Alberta (2005).

- [20] C. E. Priebe, D. J. Marchette, and G. W. Rogers, "Segmentation of random fields via borrowed strength density estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(5), 494–499 (1997).
- [21] E. G. P. Bovenkamp, J. Dijkstra, J. G. Bosch, and J. H. C. Reiber, "Multi-agent segmentation of IVUS images," *Pattern Recognition* **37**(4), 647–663 (2004).
- [22] E. Duchesnay, J. J. Montois, and Y. Jacquelet, "Cooperative agents society organized as an irregular pyramid: a mammography segmentation applications," *Pattern Recognition Letters* **24**(14), 2435–2445 (2003).
- [23] M. P. Wachowiak, R. Smolikova, Y. F. Zheng, J. M. Zurada, and A. S. Elmaghraby, "An approach to multimodal biomedical image registration utilizing particle swarm optimization," *IEEE Transactions on Evolutionary Computation* **8**(3), 289–301 (2004).
- [24] L. G. Shapiro and G. C. Stockman, *Computer Vision* (Prentice Hall, Upper Saddle River, N.J., 2001).
- [25] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognition* **26**, 1277–1294 (1993).
- [26] R. M. Rangayyan, *Biomedical Image Analysis* (CRC Press, Boca Raton, 2004).
- [27] J. M. Liu and Y. Y. Tang, "Adaptive image segmentation with distributed behavior-based agents," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(6), 544–551 (1999).
- [28] J. M. Liu, H. Jing, and Y. Y. Tang, "Multi-agent oriented constraint satisfaction," *Artificial Intelligence* **136**(1), 101–144 (2002).
- [29] D. K. Panjwani and G. Healey, "Markov random-field models for unsupervised segmentation of textured color images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17**(10), 939–954 (1995).
- [30] T. Ojala and M. Pietikainen, "Unsupervised texture segmentation using feature distributions," *Pattern Recognition* **32**(3), 477–486 (1999).
- [31] A. K. Jain and K. Karu, "Learning texture discrimination masks," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(2), 195–205 (1996).
- [32] S. Lee and M. M. Crawford, "Unsupervised multistage image classification using hierarchical clustering with a Bayesian similarity measure," *IEEE Transactions on Image Processing* **14**(3), 312–320 (2005).
- [33] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20**(11), 1254–1259 (1998).

- [34] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition - a gentle way," in *Biologically Motivated Computer Vision, Second International Workshop (BMCV 2002)*, pp. 472–479 (Tubingen, Germany, 2002).
- [35] A. Sha'ashua and S. Ullman, "Structural saliency: the detection of globally salient structures using a locally connected networks," in *The International Conference on Computer Vision (ICCV)*, pp. 321–327 (Tarpon Springs, Florida, 1988).
- [36] M. Meister and M. Berry, "The neural code of the retina," *Neuron* **22**, 435–450 (1999).
- [37] X. L. Wu, "Image-coding by adaptive tree-structured segmentation," *IEEE Transactions on Information Theory* **38**(6), 1755–1767 (1992).
- [38] M. G. H. Omran, A. Salman, and A. P. Engelbrecht, "Dynamic clustering using particle swarm optimization with application in image segmentation," *Pattern Analysis and Applications* **8**(4), 332–344 (2006).
- [39] C. Bourjot, V. Chevrier, and V. Thomas, "How social spiders inspired an approach to region detection," in *International Conference on Autonomous Agents and MultiAgent Systems*, vol. 1, pp. 426–433 (Bologne, Italy, 2002).
- [40] Y. Wang and B. Yuan, "Fast method for face location and tracking by distributed behaviour-based agents," *IEE Proceedings - Vision, Image, and Signal Processing* **149**(3), 173–178 (2002).
- [41] T. Mirzayans, N. Parimi, P. Pilarski, C. Backhouse, L. Wyard-Scott, and P. Musilek, "A swarm-based system for object recognition," *Neural Network World* **15**(3), 243–255 (2005).
- [42] A. Broggi, M. Cellario, P. Lombardi, and M. Porta, "An evolutionary approach to visual sensing for vehicle navigation," *IEEE Transactions on Industrial Electronics* **50**(1), 18–29 (2003).
- [43] A. P. Engelbrecht, *Computational Intelligence: An Introduction* (John Wiley & Sons, New York, 2002).
- [44] B. Prasad, S. Du, W. Badawy, and K. Kaler, "A real-time multiple-cell tracking platform for dielectrophoresis (DEP)-based cellular analysis," *Measurement Science & Technology* **16**(4), 909–924 (2005).
- [45] R. Ghrist and D. Lipsky, "Gramatical self assembly for planar tiles," in *International Conference on MEMS, NANO and Smart Systems*, W. Badawy and W. Moussa, eds., pp. 205–211 (Banff, Alberta, 2004).
- [46] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification: Second Edition* (Wiley Interscience, New York, 2001).
- [47] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. (Morgan Kaufmann, New York, 2005).

- [48] J. K. Udupa, V. R. LeBlanc, Z. G. Ying, C. Imielinska, H. Schmidt, L. M. Currie, B. E. Hirsch, and J. Woodburn, "A framework for evaluating image segmentation algorithms," *Computerized Medical Imaging and Graphics* **30**(2), 75–87 (2006).
- [49] L. Bergman, A. Verikas, and M. Bacauskiene, "Unsupervised colour image segmentation applied to printing quality assessment," *Image and Vision Computing* **23**(4), 417–425 (2005).
- [50] J. Taylor, *An Introduction to Error Analysis*, 2nd ed. (University Science Books, Sausalito, California, 1997).
- [51] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision Research* **45**, 205–231 (2005).
- [52] P. M. Pilarski and C. J. Backhouse, "A method for cytometric image parameterization," *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.

## Chapter 7

# Reverse Monte Carlo Analysis for Wide-Angle Cellular Scattering

### 7.1 Introduction

Despite the potential for very rapid medical diagnostics based on the scattering of cellular organelles such as the mitochondria, current analysis methods are still unable to predict complex 3D cell structures directly from a wide-angle scattering image. This chapter describes a preliminary study in combining the `mtPatterns` algorithm with Reverse Monte Carlo methods—iterative algorithms that have been used to reconstruct 3D structure from a number of different data types, including X-ray diffraction patterns and nuclear magnetic resonance plots.

Here we apply RMC to the problem of identifying specific organelle distributions and placements from wide-angle cellular scattering patterns. However, RMC is deliberately a very general method<sup>7</sup>. This facilitates its first-order application to a number of very different problem domains, but means that each domain requires its own set of tailored search parameters for more effective, optimized search<sup>7</sup>. As the search domain in the cellular version of RMC (here abbreviated as cRMC) has key differences from that of traditional RMC problems—*e.g.* cellular geometry and the spacing/size/arrangement of scatterers in the scattering model—it is unclear what combination of search parameters and enhancements will allow the successful reconstruction of cellular scatterer structure.

As such, this chapter surveys a range of possible parameters and optimizations. Parameters sweeps are based on traditional RMC methods and known strategies for successful iterative search and local optima avoidance. The resulting comparison presents a first look at cRMC performance trends and highlights areas that merit further exploration. These initial results show that cRMC may have the potential to reproduce approximate 3D distributions of mitochondria through iterative image comparison and model refinement.

## 7.2 Background

In previous work, we showed several approaches for the parametric interpretation of wide-angle scattering signatures<sup>1-4</sup>. This provides a useful set of tools to look into the general structure of a cell. However, there are cases when it is important to know specific details about the internal placement of organelles—*e.g.* for exploratory medicine. In such cases, wide-angle cytometry holds the potential to examine structures equal to, or smaller than, the wave-length of light, without the need for expensive optical hardware<sup>5</sup>. Unfortunately, extracting exact geometric information from these 2D patterns is still an unsolved problem.

While a specific solution to the cellular inverse scattering problem remains elusive<sup>6</sup>, other disciplines have shown it is possible to infer 3D crystal structure directly from a two-dimensional scattering pattern. Of note, Reverse Monte Carlo methods have found wide use in predicting the structural arrangement of crystal, liquid, glass, polymer, and even magnetic particle systems<sup>7,8</sup>. These search methods function by creating a candidate model  $P_c$  of a scattering structure or particle system and iteratively comparing its simulated scattering pattern  $IM_c$  (or a resulting parameterization) with the experimental scattering signature  $IM_t$  from an unknown target structure  $P_t$ ; the goal is to infer the geometry of this target (*i.e.*  $P_t$ ). In most cases the difference between scattering images is calculated based on a  $\chi^2$  metric, where  $\sigma(x, y)$  is a measure of experimental error<sup>7</sup> (*n.b.* where unknown, this may be the magnitude at a point in the target image<sup>9</sup>). This “fitness function” is shown in Eq. 7.1.

$$\chi_n^2 = \sum_{x,y} \frac{[IM_c(x,y) - IM_t(x,y)]^2}{\sigma(x,y)} \quad (7.1)$$

Structure prediction is done via an iterative process; at every time step, the candidate model  $P_c$  is perturbed in some way (usually through the position shift of one of more atoms<sup>7</sup>), and its new simulated scattering signature is compared to the target signature. As per Eq. 7.2, if this results in a decrease in the  $\chi^2$  image difference, the new candidate distribution  $P'_c$  replaces  $P_c$ . Otherwise, the population remains constant. With a small  $\chi^2$ -dependent probability, poor distributions replace a more successful previous population—as in evolutionary computing (EC) and simulated annealing, this helps the algorithm avoid premature convergence to local optima<sup>7</sup>. During iteration, structural constraints are also imposed on the candidate model to steer the evolution of a solution and ensure a physically relevant candidate solutions.

$$P_c = \begin{cases} P_c, & \text{if } \chi_n^2 < \chi_{n-1}^2 \\ P'_c, & \text{if } p < \exp(-(\chi_n^2 - \chi_{n-1}^2)/2); \\ P_c, & \text{otherwise} \end{cases} \quad (7.2)$$

After a set number of iterations, or the onset of a termination condition (such as a low rate of change<sup>10</sup>), the candidate population  $P_c$  is returned as a possible solution to the true target structure  $P_t$ . This method has proved highly successful for fault detection, structural exploration, and crystal identification<sup>7</sup>. A description of the traditional RMC process can be found in McGreevy’s comprehensive review article on the subject<sup>7</sup>.

It is extremely important that sub-optimal solutions are at times accepted by the search process to avoid premature convergence, especially as solutions stabilize on known good candidates<sup>7,10</sup> (e.g. using Eq. 7.2). In traditional RMC, the probability of accepting a sub-optimal solution is tied to the value of a fitness-dependent exponential function<sup>7</sup>. McGreevy shows that this is an effective analog to the energetics-based Metropolis criterion used in Metropolis Monte Carlo (MMC)—another common, but less general, method for iterative structure prediction<sup>7</sup>. As in simulated annealing and some evolutionary algorithms, MMC uses a Boltzmann distribution to determine the “selective pressure” or generosity of the acceptance routine<sup>7,10</sup>. The MMC acceptance probability is therefore similar to Eq. 7.2, but contains a denominator term that changes during iteration:  $p < \exp(-(\chi_n^2 - \chi_{n-1}^2)/(K_b T))$ , where  $K_b$  is the Boltzmann constant and the cooling schedule  $T$  is a measure of system energy/order<sup>7,10</sup>. This gives the acceptance algorithm feedback as to the actual (or operator-defined) state of the search process<sup>7,10</sup>.

By contrast, RMC aims only to produce a viable model that is consistent within its own constraints and error definitions, and is not necessarily governed by defined energetics<sup>7</sup>—the rate of acceptance of sub-optimal solutions is therefore determined by a dynamic feedback signal that depends on the observed state of the search process (e.g.  $\chi_n^2 - \chi_{n-1}^2$ )<sup>7</sup>. This gives it the flexibility to deal with novel data types and avoid some of the over-fitting of errors experienced using MMC<sup>7</sup>. It also removes the need to determine an arbitrary cooling schedule for a domain with no thermal energetics, a lengthy task that if done improperly can severely compromise the quality and validity of results<sup>7</sup>. Despite these issues, the Metropolis criterion is a powerful approach that can—with a valid cooling schedule—effectively avoid premature convergence. This has been shown in evolutionary computing (memetics) tasks, where the cooling schedule—and thus selective pressure—was tied to the diversity of individuals in a candidate population<sup>10</sup>. Due to the wide range of possibilities, however, the evaluation of such cooling schemes is beyond the scope of the current study.

**Table 7.1:** The RMC algorithm for scattering prediction

---

0: **set** target image  $IM_t$   
1: **initialize** candidate population  $P_c$   
2: *for*  $n \in N$  || termination:  
3:      $P'_c = P_c$   
4:     *if*  $P_{jump}$ : randomly **replace** element in  $P'_c$   
5:     *else*: **move** random element in  $P'_c$  by  $\delta$   
6:      $IM_c = mtPatterns(P'_c)$   
7:     **evaluate**  $\chi_n^2$  difference between  $IM_c$  and  $IM_t$   
8:     **update**  $P_c$  based on  $\chi_n^2$  and  $\chi_{n-1}^2$   
9: **return**  $P_c$

---

### 7.3 RMC for Cellular Light Scattering

Recently, we demonstrated a rapid simulation method that was capable of producing realistic wide-angle scattering signatures from a distribution of mitochondrial scatterers<sup>1,3</sup>. This method—the `mtPatterns` algorithm—allows quick image generation and comparisons not previously possible. As such, it facilitates the use of RMC methods for the cellular scattering problem. By treating a mitochondrial scattering distribution  $\mathbf{r}_n$  as the RMC candidate distribution  $P_c$ , given a known scattering geometry—*i.e.* the location of scatters with respect to incident light and a receptive field—we can iteratively generate a set of candidate scattering signatures using the procedure described in Tab. 7.1.

As with transferring any method to a new domain, the number of possible parameter combinations is vast. Similarly, some procedures and parameters used in the traditional domain may fail in the new domain, requiring alternate approaches. With this in mind, we use the following work to explore the transferability of RMC to the cellular domain. The results present a first-order sweep through the possible parameter space to identify areas for future improvement. Simplicity of solutions is emphasized, and where possible we draw on standard methods from both RMC and EC. This study is not meant to describe a finished or fully optimized method; rather, it presents a constructive starting point for future cRMC implementations.

#### 7.3.1 Mutation Methods

Variation or modification of the sample distribution (“mutation”) is the fundamental procedure for improving a candidate solution in RMC and many EC implementations<sup>10</sup>. The

magnitude of the mutation, and the type needed, is highly application specific; many very successful mutation schemes have been described in the literature, but none are effective in every situation<sup>10</sup>. In this work we use the standard variation procedure from RMC, but supplement it with deterministic (dynamic) parameter control methods as outlined in Eiben and Smith<sup>10</sup>.

As done in traditional RMC<sup>7</sup>, at each iteration one scatterer (organelle) is randomly selected and moved via a Gaussian distribution triple  $G_{\theta,\phi,r}(0, \delta)$ . This move is centered on the organelle's location with a step size ( $\delta$ ) given in microns. Maximum movement ranges and boundary checking (*e.g.* for collisions between organelles) are used to prevent the violation of physical constraints<sup>7</sup>. This is the cRMC analog to hard shell potential constraints<sup>7</sup>. Eiben and Smith describe how adaptive parameters can facilitate the early acquisition of good solutions in a rugged search space without deterring the exploitation of identified optima later in the search process<sup>10</sup>. However, depending on the nature of the search domain, it can be advantageous to focus on certain types of exploration mechanisms, both dynamic and static. As such, tests were performed with the mutation step size  $\delta$  being held static (SM), varied linearly during the run (DM), and set to zero (deactivated).

Computational intelligence studies have shown that in some cases random shifts to other portions of the search space can help iterative search climb out of local minima<sup>10,11</sup>. To examine this possibility for cRMC, we implemented a routine that can, in lieu of the standard Gaussian movement, randomly relocate one member of the population with a probability  $P_{jump}$ . If  $p < P_{jump}$ , one scatterer in  $P'_c$  is replaced with a new randomly located organelle whose coordinates are picked from a uniform distribution  $U([r_m, r_M]; \theta; \phi)$  (where  $r_i$  and  $r_o$  are the minimum and maximum radii of a cellular distribution volume, and  $\theta, \phi$  are angular coordinates; Eq. 7.3). As shown by a number of groups, "hyper-mutating" a population beyond the bounds of normal movement can possibly access a more viable region of the search landscape<sup>10-12</sup> (*e.g.* relocating an organelle to the far side of the nucleus). This is comparable to classical RMC schemes where atoms are added, modified, or deleted during the course of a run<sup>7</sup>. Thus, for each mutation type, we also tested the effect of keeping the jump probability  $P_{jump}$  static (SJ), varying it linearly over the course of a run (DJ), and setting it to zero (deactivated). This allows classical RMC (no  $P_{jump}$ , static mutation) to be compared to more adaptive methods to determine the best combination for cRMC.

$$P'_c[i] = \begin{cases} U([r_m, r_M]; \theta; \phi), & \text{if } p < P_{jump} \\ P'_c[i] + G_{\theta,\phi,r}(0, \delta), & \text{otherwise} \end{cases} \quad (7.3)$$

### 7.3.2 Acceptance of Solutions

Accepting sub-optimal solutions is critical to avoiding local optima and finding narrow fitness basins in a rugged search landscape<sup>10</sup>. To help identify the best method for accepting sub-optimal solutions, we used both Eq. 7.2 from traditional RMC and a deterministic parameter control scheme that varies the likelihood of accepting sub-optimal solutions over the course of the run in response to a pre-defined schedule or system feedback. In this case, we experimented with a basic linear schedule. This allowed, with small probability, the system some free movement (FM) without penalty, encouraging the exploration of unseen sections of the search landscape. This was coupled with the option for a “flash-back” (FB) probability, whereby the population could return to a previously known best (or other solution point) if it did not improve after a set number of generations (*i.e.* a form of “elitism”, as described by Eiben and Smith<sup>10</sup>). Tests were performed for all possible combinations of these options, and compared to the acceptance criteria of classical RMC.

### 7.3.3 Fitness Metrics

Fitness metrics, or ways of measuring the success of a solution, drive the iterative search process. As such, identifying the best fitness representation for a given search problem is critical to the success of the method<sup>10</sup>. For the cRMC fitness metric, we explored the use of direct  $\chi^2$  pixel-based comparison (PBRMC) as in traditional RMC, and also a fitness metric based on texture attributes (texture-based comparison; TBRMC). As described by McGreevy, it is often advantageous to base fitness on a parameterization or feature-based representation of the experimental sample, instead of the sample itself<sup>7</sup>. For the TBRMC texture metric, the sum-squared-error (SSE<sup>15</sup>) between the target and candidate was computed for a vector of five Law texture parameters known to be sensitive to distribution shape—*i.e.* S5, S5x2, W5, R5, and E5<sup>2</sup>. As demonstrated in previous work, using texture information will decrease sensitivity to exact organelle position, but should increase robustness to population rotation and translation<sup>2,3</sup>.

The best option is likely a combination of the two approaches, as this could maintain a degree of specificity while steering the search process toward viable geometries through distribution-sensitive texture parameters. As such, tests could be performed using a fitness metric that is the normalized and balanced combination of both texture and direct-comparison  $\chi^2$  values. Texture and direct comparison functions provide preliminary results. Future work will present a more detailed discussion of a hybrid fitness function, and a complete analytical and comparative study is currently underway.

### 7.3.4 Test Cases

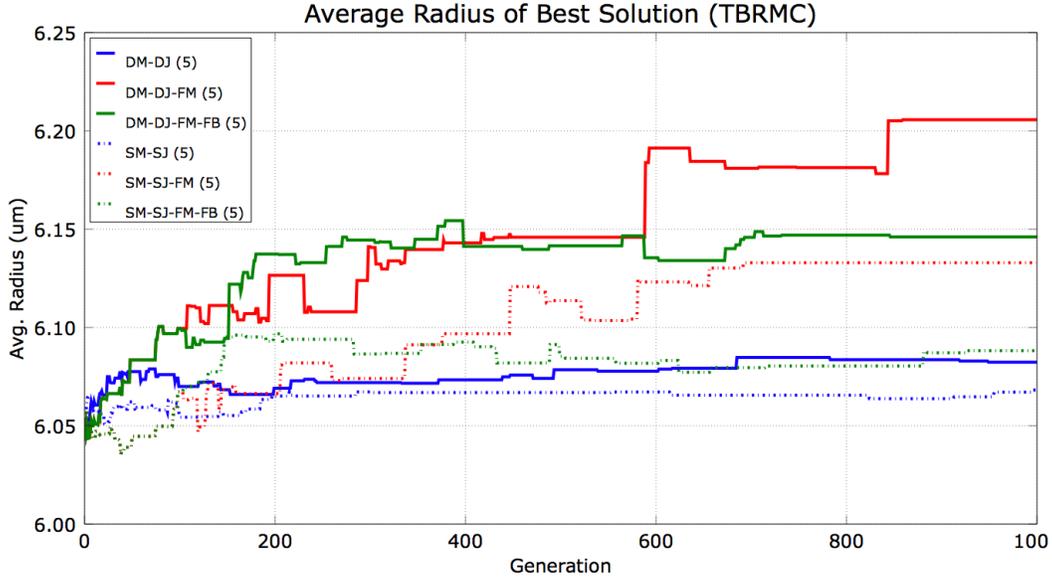
The TBRMC and PBRMC approaches were applied to two different target models: a *peripheral distribution* (organelles located within  $0.8\mu m$  of the cell wall) and a *perinuclear distribution* (organelles located within  $0.8\mu m$  of the nucleus); distribution geometry and simulation parameters were set according to Pilarski *et al.*<sup>2</sup>. As discussed previously, these distributions can be used as indicators for a cancer patient's susceptibility to treatment, and can be used to detect the progression of disease<sup>2,13</sup>.

Each model contained 100mt. For the target population  $P_t$ , organelles were randomly placed inside the distribution boundaries. For the starting candidate population  $P_c$ , organelles were randomly placed within the entire available volume (*e.g.* a diffuse distribution, as per Pilarski *et al.*<sup>2</sup>). As per our previous work<sup>2</sup>, the simulated cell parameters were based on a known cancerous cell type (Raji), typically found with a nuclear radius of  $4\mu m$  and a cell radius of  $8\mu m$ <sup>14</sup>. For this preliminary study, each RMC run lasted 1000 iterations, and was done for a number of different starting candidate populations for each target distribution.

## 7.4 Results

A quantitative comparison of the success of different mutation, acceptance, and flashback parameters is shown in Figs. 7.1 (TBRMC) and 7.2 (PBRMC). Traces in these plots indicate the average distribution radius of the best candidate solution; this is one measure of how close a candidate distribution is to the target distribution. For the tests shown in Figs. 7.1 and 7.2, a peripheral target distribution was used, as per the description in Pilarski *et al.*<sup>2</sup>. The average radius of this peripheral target distribution was  $7.6\mu m$ . The average starting radius for each candidate distribution was approximately  $6.05\text{--}6.15\mu m$ , and organelles were initially spread at random throughout the entire cytoplasm (*i.e.* a diffuse distribution<sup>2</sup>). Each trace indicates the average of five tests, each initialized with different random seeds for both the target and candidate distribution (*n.b.* the same starting and target models were used for both TBRMC and PBRMC to allow valid comparison between methods). Maximum variance for any given trace was approximately  $\pm 0.05\mu m$ , often less.

Traces are labeled with the combinations of dynamic and static movement and jump probabilities (DM-DJ / SM-SJ), free movement (FM), and flashback rate (FB) used in each set of tests. Dynamic mutation size (DM) was linearly varied from  $\delta = 4.0\mu m$  (start) to  $\delta = 0.25\mu m$  (end), dynamic jump probability (DJ) was linearly varied between  $P_{jump} = 0.1$  (start) and  $P_{jump} = 0.0$  (end). Static mutation size (SM) was set at  $\delta = 2.0\mu m$ , while static

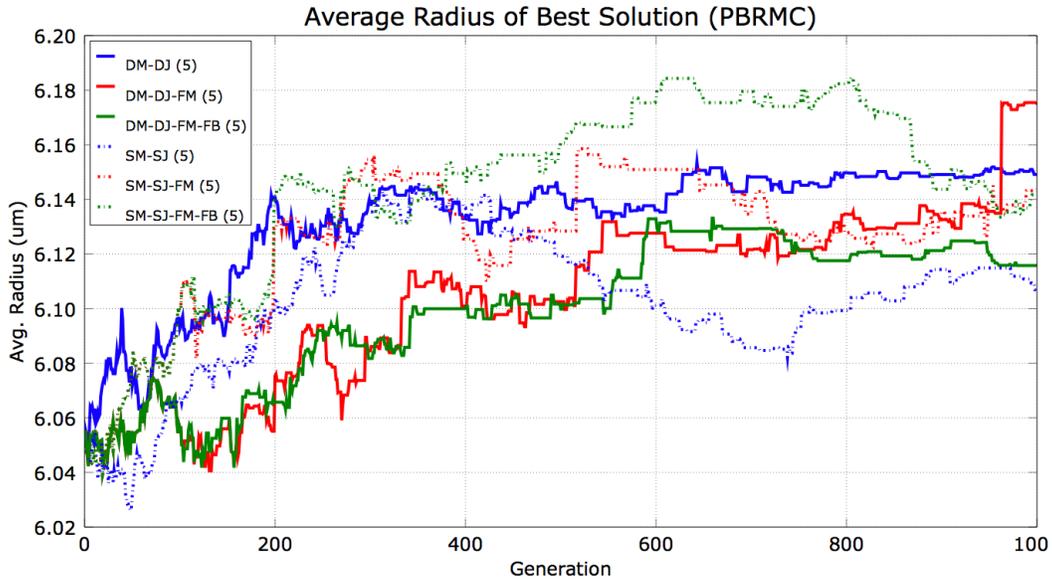


**Figure 7.1:** Comparison of mutation, acceptance, and flashback parameters for TBRMC using 100mt and a peripheral target distribution. Traces indicate the average distribution radius of the best candidate solution using combinations of dynamic and static movement and jump probabilities (DM-DJ / SM-SJ), dynamic free movement (FM), and flashback (FB). The higher the trace, the closer the system is to the avg. radius of the target distribution (in this case  $7.6\mu m$ ). Maximum variance for any given trace was approximately  $\pm 0.05\mu m$ , often less.

jump probability (SJ) was set at  $P_{jump} = 0.1$ . Where used, the dynamic free movement probability (FM) was linearly varied between 0.1 (start) and 0.0 (end), and flashback testing (FB) occurred every 100 generations. Comparison image size was  $25 \times 25px$ .

There are several important observations to take away from Figs. 7.1 and 7.2. For TBRMC, dynamic mutation schemes outperformed static schemes. In addition, dynamic schemes using the linear free movement probability showed the greatest potential for refining candidate solutions, as measured by their high final radius values for  $P_c$ . In particular, a small amount of blind exploration of sub-optimal solutions seems to allow the algorithms to more fully examine the search landscape and find better candidate models; this holds with expectations for both RMC and EC literature<sup>7,10</sup>.

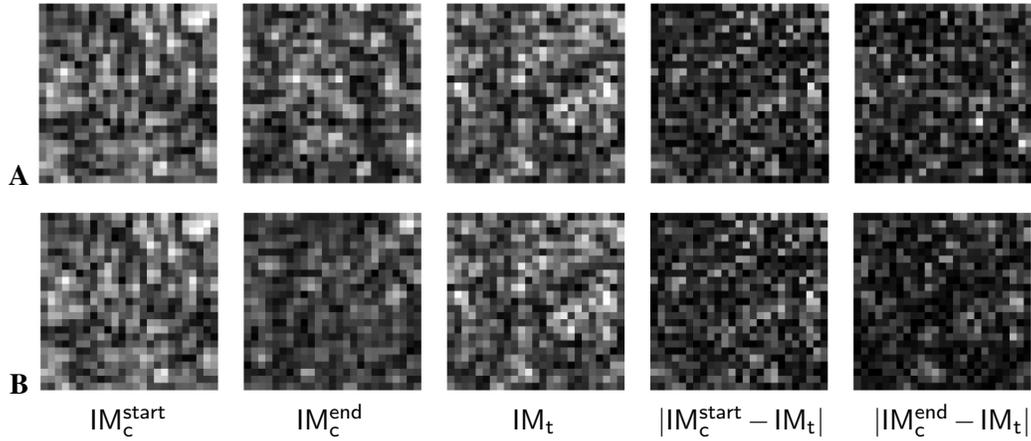
Comparing TBRMC to PBRMC, it was found that—unlike TBRMC—an increase in the fitness of the best candidate solution frequently did not correlate with a successful move toward the target distribution. This can be seen via the oscillations in Fig. 7.2. However, as with TBRMC, DM-DJ-FM produced the best final result after 1000 iterations. The difference between static and dynamic methods was less pronounced for PBRMC, and in some cases negligible. Static methods equaled or exceeded dynamic methods early in the iteration, but dynamic methods showed a steady increase to eventually match or exceed them.



**Figure 7.2:** Comparison of mutation, acceptance, and flashback parameters for PBRMC using 100mt and a peripheral target distribution. Traces indicate the average distribution radius of the best candidate solution using combinations of dynamic and static movement and jump probabilities (DM-DJ / SM-SJ), dynamic free movement (FM), and flashback (FB). The higher the trace, the closer the system is to the avg. radius of the target distribution (in this case  $7.6\mu m$ ). Maximum variance for any given trace was approximately  $\pm 0.05\mu m$ , often less.

It is important to note that methods using the traditional RMC acceptance criteria shown in Eq. 7.2 performed exactly the same as standard DM-DJ and SM-SJ—they were less successful than a simple linear probability for sub-optimal acceptance. Experiments with an increasing free movement probability that varied from 0.0 (start) to 0.1 (end) showed inferior performance to the decreasing free movement probability traces shown in Figs. 7.1 and 7.2—for both TBRMC and PBRMC, final performance for decreasing FM was better than DM-DJ, but significantly less than DM-DJ-FM. This indicates that for cRMC it may be advantageous to randomly explore the search space in the earlier stages of RMC iteration.

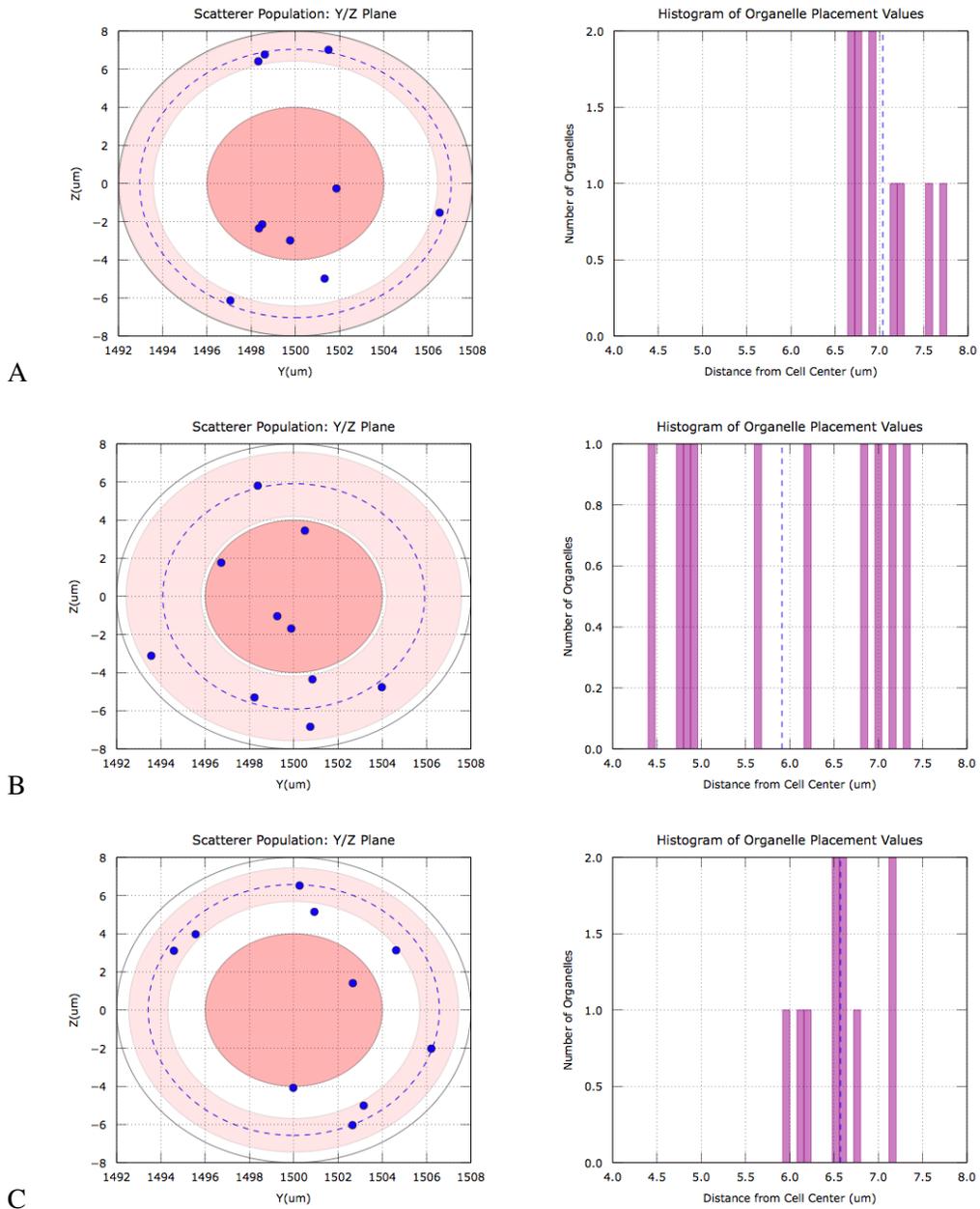
However, too much random exploration appears to be detrimental. For the DM-DJ case, a FM rate of 0.25 (start) was found to perform worse than a FM rate of 0.1 (start), as did a static FM rate that was held at 0.1 for the entire iteration. This indicates that while too much unconstrained movement will lead to high initial gains, in the middle and end of the search process it may negatively impact candidate solutions. Additional comparisons indicate that DM-DJ-FM dramatically outperforms tests with just a high static mutation or jump rate, and also tests where either the DM or DJ rate was increased as the iteration progressed (*i.e.* upwards linear change). These observations suggests that the performance of the DM-DJ-FM scheme is not just a factor of high initial mutation parameters—later refinement is in fact useful. This was the case for both peripheral and perinuclear target distributions.



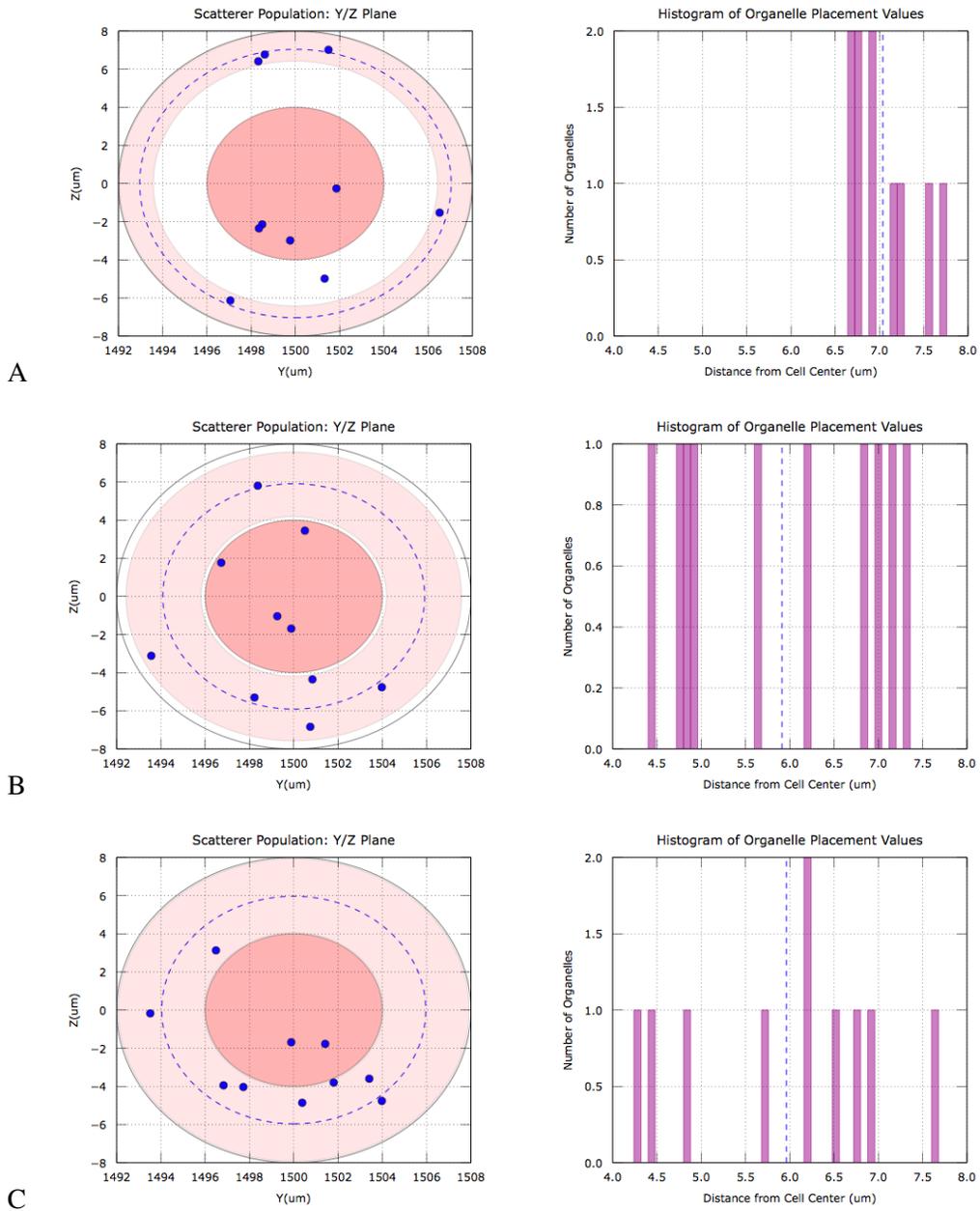
**Figure 7.3:** Comparison of scattering images from a peripheral target population  $IM_t$  containing 100 organelles with a DM-DJ-FM candidate population  $IM_c$  after 0 (start) and 1000 (end) iterations using (A) TBRMC and (B) PBRMC.

Fig. 7.3 shows a comparison of  $IM_t$  with the candidate scattering images captured at the beginning and end of DM-DJ-FM iteration—*i.e.* at 0 and 1000 iterations of the TBRMC and PBRMC tests shown in Figs. 7.1 and 7.2. While the actual composition of  $IM_c^{end}$  is different from the target image  $IM_t$  (as shown by the brightness in the difference images  $|IM_c^{end} - IM_t|$ ), it shares a number of visual similarities in terms of the shape and size of its component intensity regions. Candidate image texture characteristics such as peak shape and size begin to approach those of the target image toward the end of the TBRMC test. This increase in texture similarity was quantitatively observed as a significant decrease in the recorded SSE values over the course of TBRMC iteration. As would be expected, the increase in texture similarity between  $IM_c$  and  $IM_t$  was not as pronounced for candidate images generated using PBRMC (shown in Fig. 7.3, B); the  $\chi^2$  metric leads the algorithm to match the exact intensity profile, as opposed to texture quality. This increase in exact image similarity can be seen in the decreased intensity of  $|IM_c^{end} - IM_t|$  in Fig. 7.3, B.

As a graphical example of population improvement over the course of RMC, Figs. 7.4 and 7.5 show a comparison of a peripheral target population  $P_t$  containing ten organelles (A) with the candidate population  $P_c$  after (B) 0 and (C) 1000 RMC iterations (using TBRMC, Fig. 7.4, and the PBRMC  $\chi^2$  metric, Fig. 7.5). These tests both used a dynamic mutation size that shifted from  $\delta = 4.0\mu m$  (start) to  $\delta = 0.25\mu m$  (end), a dynamic jump probability between  $P_{jump} = 0.1$  (start) and  $P_{jump} = 0.0$  (end), a dynamic (linear) free movement probability between 0.1 (start) and 0.0 (end), and flashback testing every 50 generations; comparison image size was  $25 \times 25px$ .



**Figure 7.4:** Comparison of a peripheral target population  $P_t$  containing ten organelles (A) with the candidate population  $P_c$  after (B) 0 and (C) 1000 iterations using TBRMC. The cell distribution's Y/Z projection is shown on the left, with the shaded region indicating the region taken up by organelles, while a histogram showing the radius of component organelles is to the right (blue dotted line indicates the avg. radius of the distribution). After 1000 iterations, there is a noticeable shift in the candidate distribution, toward the periphery of the cell.



**Figure 7.5:** Comparison of a peripheral target population  $P_t$  containing ten organelles (A) with the candidate population  $P_c$  after (B) 0 and (C) 1000 iterations using PBRMC. The cell distribution's Y/Z projection is shown on the left, with the shaded region indicating the region taken up by organelles, while a histogram showing the radius of component organelles is to the right (blue dotted line indicates the avg. radius of the distribution). After 1000 iterations, there is only a slight shift in the candidate distribution, toward the periphery of the cell.

The cell distribution's Y/Z projection is shown on the left, with the shaded region indicating the radial region taken up by organelles, while a histogram showing the radius of component organelles is to the right (blue dotted line indicates the avg. radius of the distribution). After 1000 iterations using TBRMC, there is a noticeable shift in the candidate distribution toward the periphery of the cell. However, after 1000 iterations of PBRMC there is only a slight shift in the average radius of  $P_c$  toward the periphery of the cell.

Qualitative results for this ten organelle study held with those of the 100 organelle study shown in Figs. 7.3. Though TBRMC showed larger differences between  $IM_c^{end}$  and  $IM_t$  than PBRMC, reflected in the difference image  $|IM_c^{end} - IM_t|$ , image texture in  $IM_c^{start}$  was visually closer to that of the target image; *e.g.* as would be expected for a shift toward a peripheral distribution<sup>2</sup>, intensity regions are on average smaller than in  $IM_c^{start}$ .

Both qualitative and quantitative observations using both peripheral and perinuclear distributions indicate that TBRMC has the potential to identify cellular geometry, at least in general terms, from a wide-angle scattering signature. As expected, while a texture-based fitness metric was found to be less sensitive to the exact position of organelles, TBRMC was much more successful in determining the overall shape of a population than PBRMC. It is likely that by developing a hybrid of these two fitness metrics it would be possible to quickly approach the distribution type (using TBRMC) then focus in on the exact placement of organelles (using PBRMC). Work with expanded direct image comparison metrics, hybrid metrics, and additional parameter values is currently underway.

## 7.5 Conclusions

In this chapter, we demonstrated an extension of RMC analysis techniques to the wide-angle scattering domain, and showed how it may be used to help predict 3D organelle distributions for medically relevant cell types from wide-angle light scattering patterns. This is relevant to predictive systems, as the ability to recover a 3D distribution of organelles could help identify and characterize the progression of disease<sup>13,16</sup> and help predict a patient's response to therapy<sup>13,17</sup>. While these results are preliminary, and there is a range of parameter values and search methods left to explore, this work provides good a foundation for later studies.

From the current survey of the available parameter space, it was found that TBRMC methods using free movement generated the best final result, and improve quite rapidly at the beginning of the RMC iteration. However, they seem to plateau early in the iteration process (Fig. 7.1). This could be the result of premature convergence, and indicates the need

for tests to determine if a different (*e.g.* non-linear) criteria for accepting sub-optimal solutions, or a more complex method for varying mutation, would facilitate continued (and more rapid) improvement. PBRMC showed slightly worse performance when compared to TBRMC, but was more successful at actually matching the exact intensity pattern of the target scattering image (though not the target image texture).

While the initial acceptance of sub-optimal solutions (FM) may be one key to finding good candidate populations in cRMC, the exact free movement probability to use and the rate at which to vary its value during iteration remain open topics for study. The success of tests using free movement indicate the potential for more complex adaptive acceptance and mutation methods—*e.g.* gradient-based methods as used in evolutionary strategies, empirical potential structure refinement, or a true Metropolis criterion and associated cooling schedule based on fitness change or solution diversity<sup>7,10</sup>. Additional studies are needed to determine if flashback (elitism) does in fact help find good solutions, or if it actually hinders the identification of novel solutions (as seems to be indicated by Fig. 7.1).

To our knowledge, this preliminary work is the first exploration of RMC methods to infer specific aspects of 3D cellular geometry directly from its wide-angle scattering signature. This holds great potential for both patient diagnostics and exploratory research. Future work will expand these initial results, using comprehensive multi-objective fitness functions and more advanced mutation and free movement schemes to avoid the premature convergence of solutions.

## References

- [1] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Rapid simulation of wide-angle scattering from mitochondria in single cells,” *Optics Express* **16**(17), 12,819–12,834 (2008). URL <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-17-12819>.
- [2] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Computational analysis of mitochondrial placement and aggregation effects on wide-angle cell scattering patterns,” in *Proceedings of SPIE*, A. Wax and V. Backman, eds., vol. 7187, p. 71870J (12 pages) (2009). URL <http://dx.doi.org/10.1117/12.809730>.
- [3] P. M. Pilarski and C. J. Backhouse, “Theoretical foundations for robust wide-angle scattering pattern analysis,” *Optics Express* **submitted**, 5 pages (2009).
- [4] P. M. Pilarski and C. J. Backhouse, “A method for cytometric image parameterization,” *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.

- [5] X.-T. Su, "Light scattering in an integrated microfluidic waveguide cytometer," Ph.D. thesis, University of Alberta (2008).
- [6] V. P. Maltsev, "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [7] R. L. McGreevy, "Reverse Monte Carlo modelling," *Journal of Physics: Condensed Matter* **13**(46), R877–R913 (2001).
- [8] V. FitzGerald, D. M. Pickup, D. Greenspan, G. Sarkar, J. J. Fitzgerald, K. M. Wetherall, R. M. Moss, J. R. Jones, and R. J. Newport, "A neutron and X-ray diffraction study of bioglass with reverse Monte Carlo modelling," *Advanced Functional Materials* **17**(18), 3746–3753 (2007).
- [9] J. Taylor, *An Introduction to Error Analysis*, 2nd ed. (University Science Books, Sausalito, California, 1997).
- [10] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Natural Computing Series, 2nd ed. (Springer, New York, 2007).
- [11] L. N. DeCastro and F. J. VonZuben, "Artificial Immune Systems: Part I - Basic Theory and Applications," Technical Report TR - DCA 01/99, State University of Campinas (1999).
- [12] P. Musilek, A. Lau, M. Reformat, and L. Wyard-Scott, "Immune programming," *Information Sciences* **176**(8), 972–1002 (2006).
- [13] S. Sikder, J. M. G. Reyes, C. S. Moon, O. Suwan-apichon, J. H. Elisseeff, and R. S. Chuck, "Noninvasive mitochondrial imaging in live cell culture," *Photochemistry and Photobiology* **81**(6), 1569–1571 (2005).
- [14] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, "2D light scattering patterns of mitochondria in single cells," *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [15] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification: Second Edition* (Wiley Interscience, New York, 2001).
- [16] P. L. Gourley, J. K. Hendricks, A. E. McDonald, R. G. Copeland, K. E. Barrett, C. R. Gourley, and R. K. Naviaux, "Ultrafast nanolaser flow device for detecting cancer in single cells," *Biomedical Microdevices* **7**(4), 331–339 (2005).
- [17] A. M. Villa and S. M. Doglia, "Mitochondria in tumor cells studied by laser scanning confocal microscopy," *Journal of Biomedical Optics* **9**(2), 385–394 (2004). URL <http://link.aip.org/link/?JBO/9/385/1>.

## Chapter 8

# Conclusions

The work presented in this dissertation describes a new way of understanding the link between wide-angle scattering patterns and cellular structure. This is a critical research problem as light scattering patterns contain a wealth of medically relevant structural information<sup>1-5</sup>, such as cell-macroeometry<sup>1</sup>, optical properties<sup>2,5</sup>, and the distribution of mitochondria within a cell<sup>6</sup>.

It also impacts new advances in device design and development. Emerging LOC devices are poised to make an important contribution to clinical practice and exploratory medicine<sup>7</sup>. This is especially true for the wide-angle cytometer—through miniaturization and optimization, these diagnostic platforms hold the potential to dramatically reduce testing costs while increasing the speed, accessibility, and accuracy of tests<sup>4,8-10</sup>. Using the systems and methods developed in the previous chapters, researchers now have the ability to fully engineer and utilize these powerful optical tools.

As described in Ch. 2, previous attempts to solve the cellular inverse scattering problem fall into the context of direct, parametric, and iterative methods. These approaches can produce a range of information about a scattering body or distribution. Despite the lack of a direct mathematical inverse solution<sup>2</sup>, parametric and iterative methods have been highly successful for limited angles, 1D scattering scenarios, and related problem domains<sup>1-5,11</sup>. Prior to the work described in this dissertation, no comparable approaches had been demonstrated for the case of wide-angle cellular scattering.

This work approached the scattering problem from three angles—simulation and optical theory, image analysis, and biomedical pattern classification—addressing each of the major open problems identified from previous studies (presented in Ch. 2). Chs. 3 and 4 tackled the problem of scarce labeled data—experimental wide-angle scattering datasets are limited in their availability, and previous simulation methods were either too restrictive in their geometry or subject to large computational constraints<sup>4</sup>. The new simulation methods presented herein provide both a rapid method to generate massive quantities of accurate labeled

scattering data, and a new theoretical Fourier-diffraction approach to the interpretation of wide-angle signatures. As such, the patterns generated by this method were used as the primary data source throughout the remainder of the dissertation.

Following this, Ch. 5 demonstrated the use of pattern analysis, attribute selection, and machine learning classification to categorize large sets of scattering data. This showed that texture metrics are a viable and accurate way to analyze mitochondrial scattering information. Ch. 6 presented an extension to image processing and parametric feature extraction, providing a viable method to interpret shape information in 2D scattering signatures, such as the intensity artifacts created by scattering from combinations of cellular micro- and macrostructures. Finally, Ch. 7 shows how both simulation and feature extraction may be combined to determine 3D geometric information from a scattering signature through the use of iterative search methods.

Taken as a whole, this presents a new theoretical and methodological context to approach wide-angle scattering, with tools and techniques that should transfer well to other complex pattern analysis domains. The systems and methods described in the previous chapters have been presented at conferences, published in journals, and protected under patents pending. As such, they will provide a solid basis for others to continue exploring the simulation, interpretation, and classification of wide-angle cellular scattering patterns.

## **8.1 Impact and Contributions**

Cellular light scattering is a rapidly expanding field with links to a number of other disciplines, including but not limited to nanobiotechnology, cellular scattering theory, optical simulation, optical device design, computer vision, and machine learning. As such, this work embodies a diverse interdisciplinary contribution—a new cellular light scattering simulation method was presented, validated, and used to develop image processing techniques that enable rapid pattern-based medical analysis.

One major advance is the ability to perform rapid medically related classifications directly from wide-angle scattering signatures<sup>19,20</sup>. This processing power and theoretical background will prove invaluable in the development of next-generation LOC devices. Work on both general and specific solutions to the inverse scattering problem has provided the academic community with a rapid new way to simulate cellular light scattering<sup>21</sup>, the ability to effectively categorize scattering images based on highly predictive texture metrics<sup>19,20</sup>, a powerful segmentation system to extract shape-based features from wide-angle scattering images<sup>22</sup>, and a first approach to iterative 3D structural reconstruction. It has also demon-

strated a number of previously unknown correlations between scattering pattern behaviour and scatterer structure<sup>19,20</sup>, and sets firm limits on what can and cannot be observed using information from wide-angle scattering signatures<sup>23</sup>. This will have a significant impact on exploratory medicine and research into the origins of disease.

This dissertation presents new examples of both parametric (general) and iterative (specific) approaches to the inverse structure prediction problem. In the process, it demonstrates how detailed changes to a cell's internal structure contribute to its wide-angle scattering pattern. By making visible the qualitative and quantitative relationships between scattering patterns and scatterer structure, it is possible to identify key features of two-dimensional scattering images and relate them back to changes in cellular structure. This confirms trends presented in previous studies on limited angular slices, provides a platform to validate new simulation methods (*e.g.*<sup>4,12</sup>), and develops new relational rules and simulation methods which may be used in experimental sample classification and disease research. To date, the literature has only explored this style of analysis for one dimensional patterns from commonly available fixed-angle cytometry devices, neglecting the rich information content of newly accessible two-dimensional scattering patterns from wide-angle cytometers<sup>4</sup>.

Wide-angle cytometry is an emerging field, with only a select number of devices in operation on the global stage. Prior to this work, there were no feature analysis or image processing solutions specifically designed for use on a wide-angle scattering signature<sup>22</sup>. This work presents the first example of a computational framework to begin to analyze and identify relevant patterns in complex 2D cellular light scattering patterns. As evident from the sampling in previous chapters, cellular light scattering patterns are extremely complex and varied, and come with a unique set of image processing and pattern analysis challenges<sup>19,20,22</sup>. The methods developed over the course of this project are immediately transferable to other difficult image-based problem domains, and will expand both the pattern recognition community's knowledge of possible application domains and its available algorithmic toolset.

## 8.2 Future Directions

The work presented in this dissertation opens the door for additional research into the parametric analysis of scattering structures and their 3D reconstruction from patterns of scattered laser light. In particular, two new studies could apply and extend the work presented in the previous chapters. As preliminary work has been done on each, they are presented here as an example of future directions that use and build on the tools and theoretical basis developed herein.

### 8.2.1 Applied Parameterization for Pattern Decomposition

As described in Chs. 3–5, the intensity contribution of small scatterers is centered primarily on the side-scattering region, and takes the form of a number of small complex intensity regions. Through the pattern analysis methods in Ch. 5, it is now feasible to infer the arrangement of these organelle distributions directly from wide-angle scattering patterns.

While this is an effective way to process the side-scattering data, emerging cytometry devices will soon have the capability to consistently image a full  $180^\circ$  hemisphere around an experimental sample<sup>9,10</sup>. This presents the challenge of separating organelle-related intensity information from the contributions of cellular macro-structures. Light scattered by cellular components such as the nucleus and the cell wall is focused within the front and back scatter region<sup>5,10</sup>, but also appears as intensity bands throughout the entire scattering hemisphere<sup>9,12</sup>.

Using the shape-based image analysis methods described in Ch. 6, it is possible to detect and isolate a band-based intensity pattern. The band geometry resulting from the Cythe image segmentation process gives a compact description of contributions of larger macro-structures. Once identified, these intensity contributions could be separated from those of smaller organelles. Using the relationships between intensity maxima and particle size presented by Maltsev<sup>2</sup> and Sem'yanov *et al.*<sup>1</sup>, and the distribution analysis methods of Ch. 5, this would allow the parametric interpretation of both cellular micro- and macrostructures from a single image. In a similar fashion, the Cythe algorithm could be used to identify and subtract other kinds of characterized noise and device-related interference.

Preliminary results indicate that this approach warrants further investigation. We have applied the Cythe algorithm to an example dataset of hemispherical FDTD simulations from models containing cellular macrostructures surrounded by small organelles and shown the ability to extract a viable representation of the intensity band structure within the image. This enabled us to determine the mitochondrial content of small cell models. Further work using larger simulated datasets and experimental data could yield a powerful new method for robust parametric analysis.

### 8.2.2 Advanced Iterative Structure Prediction

Initial work with RMC methods in Ch. 7 showed it was possible to iteratively derive a first approximation of the 3D structure of a scatterer population. However, there are a number of places where classical RMC may be modified to better perform for the specific case of inverse light scattering from mitochondria. Some of these possible improvements are suggested below, as avenues for further exploration.

## **Cells with an unknown number of organelles**

In experimental situations, the true number of organelles in the target cell is likely unknown. It is important to be able to still derive candidate structures, even given this possibility. There are two approaches of note<sup>13</sup>: i) try a number of different candidate distribution sizes, run for a short time, and then perform intensive search on the best solution(s); this is effective, but very time consuming<sup>13</sup>, and ii) implement methods to add and subtract scatterers during the run; this means the system needs to be more accepting of sub-optimal solutions, and may increase run time<sup>13</sup>. Both of these should be explored, with reference to the current literature before proceeding—clustering systems commonly face this problem when determining how many clustering groups to apply to a population; there is no easy solution<sup>14</sup>.

## **Different fitness metrics, adaptive parameters, and acceptance of sub-optimal moves**

Direct  $\chi^2$  comparison or simple feature comparison may not be the best evaluation metrics for the cRMC, despite their use in traditional RMC methods. The fitness function is the primary aspect of operator control in iterative systems<sup>13</sup>. A detailed survey should be performed of other possible fitness assessment methods, including hybrid functions that combine direct comparison with additional image parameters to fully exploit both the comparison-based specificity and general identification capabilities of these metrics.

Another issue discussed in Ch. 7 is that the acceptance condition used in traditional RMC may be too restrictive in a search space with different physical constraints and large mutation parameters. Our experiments have shown that in most cases for cRMC, the traditional acceptance condition rarely allows sub-optimal placement. Possible reasons for this may be that the physical constraints of placing organelles within a cell take the candidate images too far from the target to be considered for acceptance, even when the solution approaches convergence. While more generous linear acceptance probabilities were explored in Ch. 7, it would be useful to examine other options for premature convergence avoidance, as discussed by Eiben and Smith<sup>13</sup>.

Finally, it would be useful to compare a number of different parameter combinations to find the subset most able to explore the topography of the scattering search domain. Adaptive methods like evolutionary strategies could be implemented as described by Eiben and Smith<sup>13</sup> to allow the RMC process to adjust its own parameters in response to fluctuations in the search landscape.

## Advanced search methods

Ultimately, RMC has serious limitations with regard to its ability to explore a rugged cellular search space. One option to overcome the premature convergence problems of classical RMC is the use of more advanced search methods drawn from the field of evolutionary computing (EC)<sup>13</sup>.

Evolutionary computing methods are a more advanced form of generate-and-test search, with a number of advantages over more primitive methods<sup>11,13,15-17</sup>. Over the past few decades, EC has progressed from its theoretical roots to become a widely accepted problem solving tool; it has been shown to successfully match or outperform other methods in some of the most challenging domains<sup>13,15,16</sup>. Despite its increasing use in everything from structure design to autonomous vehicle navigation and artificial intelligence<sup>13,16</sup>, EC is continually breaking new ground in specific and sometimes highly specialized problem domains. Notably, EC has shown excellent performance in very rugged search landscapes, the ability to avoid convergence to poor local optima, and an increased chance of finding good optima with very small basins of attraction<sup>13,15,16</sup>.

While RMC search could be thought of as an EC method with a single population member, there are a number of advantages to a true EC method. Through the use of diversity preservation metrics (*e.g.* island models, deterministic crowding, fitness sharing, or speciation<sup>13</sup>) the search population can more rapidly and evenly explore the search landscape, increasing the chance of finding good optima with very small basins of attraction. In addition, recombination allows effective transmission of good structure examples, something lost in a heuristic containing a single population member. As shown by a wealth of previous work, there are many convincing reasons for choosing an EC method when approaching the problem of generate-and-test inverse structure prediction<sup>11,18</sup>.

Several forms of EC could be implemented, and their performance compared to that of the classical RMC method described in Ch. 7. It is expected that EC search methods that utilize previously presented simulation and feature extraction techniques will lead to functional and robust specific solutions to the structure prediction problem. This may be coupled with content based image retrieval technology to expedite the search process.

## References

- [1] K. A. Sem'yanov, P. A. Tarasov, J. T. Soini, A. K. Petrov, and V. P. Maltsev, "Calibration-free method to determine the size and hemoglobin concentration of individual red blood cells from light scattering," *Applied Optics* **39**(31), 5884–5889 (2000).
- [2] V. P. Maltsev, "Scanning flow cytometry for individual particle analysis," *Review of Scientific Instruments* **71**(1), 243–255 (2000).
- [3] L. T. Perelman, V. Backman, M. Wallace, G. Zonios, R. Manoharan, A. Nusrat, S. Shields, M. Seiler, C. Lima, T. Hamano, I. Itzkan, J. Van Dam, J. M. Crawford, and M. S. Feld, "Observation of periodic fine structure in reflectance from biological tissue: a new technique for measuring nuclear size distribution," *Physical Review Letters* **80**(3), 627–630 (1998).
- [4] X.-T. Su, C. Capjack, W. Rozmus, and C. Backhouse, "2D light scattering patterns of mitochondria in single cells," *Optics Express* **15**(17), 10,562–10,575 (2007). URL <http://www.opticsexpress.org/abstract.cfm?URI=oe-15-17-10562>.
- [5] A. K. Dunn, "Light scattering properties of cells," Ph.D. thesis, University of Texas at Austin (1997).
- [6] S. Sikder, J. M. G. Reyes, C. S. Moon, O. Suwan-apichon, J. H. Elisseeff, and R. S. Chuck, "Noninvasive mitochondrial imaging in live cell culture," *Photochemistry and Photobiology* **81**(6), 1569–1571 (2005).
- [7] V. J. Sieben, "Self-assembly on microfluidic platforms," Ph.D. thesis, University of Alberta (2009).
- [8] K. Singh, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, "Analysis of cellular structure by light scattering measurements in a new cytometer design based on a liquid-core waveguide," *IEE Proc.-Nanobiotechnology* **151**(1), 10–16 (2004).
- [9] K. Singh, X. Su, C. Liu, C. Capjack, W. Rozmus, and C. J. Backhouse, "A miniaturised wide-angle 2-D cytometer," *Cytometry A* **69A**, 307–315 (2006).
- [10] X.-T. Su, "Light scattering in an integrated microfluidic waveguide cytometer," Ph.D. thesis, University of Alberta (2008).
- [11] P. Chacon, F. Moran, J. F. Diaz, E. Pantos, and J. M. Andreu, "Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm," *Biophysical Journal* **74**(6), 2760–2775 (1998).
- [12] C. Liu, C. E. Capjack, and W. Rozmus, "3-D simulation of light scattering from biological cells and cell differentiation," *Journal of Biomedical Optics* **10**(1), 014,007 (12 pages) (2005).
- [13] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*, Natural Computing Series, 2nd ed. (Springer, New York, 2007).

- [14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification: Second Edition* (Wiley Interscience, New York, 2001).
- [15] T. Back, U. Hammel, and H.-P. Schwefel, “Evolutionary computation: comments on the history and current state,” *IEEE Transactions on Evolutionary Computation* **1**(1) (1997).
- [16] A. P. Engelbrecht, *Computational Intelligence: An Introduction* (John Wiley & Sons, New York, 2002).
- [17] P. Chacon, J. F. Diaz, F. Moran, and J. M. Andreu, “Reconstruction of protein form with X-ray solution scattering and a genetic algorithm,” *Journal of Molecular Biology* **299**(5), 1289–1302 (2000).
- [18] G. Webster and R. Hilgenfeld, “An evolutionary computational approach to the phase problem in macromolecular X-ray crystallography,” *Acta Crystallographica, Section A: Foundations of Crystallography* **57**, 351–358 (2001).
- [19] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Computational analysis of mitochondrial placement and aggregation effects on wide-angle cell scattering patterns,” in *Proceedings of SPIE*, A. Wax and V. Backman, eds., vol. 7187, p. 71870J (12 pages) (2009). URL <http://dx.doi.org/10.1117/12.809730>.
- [20] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Robust pattern analysis techniques for the interpretation of wide-angle light scattering images from single cells,” *Journal of Biomedical Optics* **submitted**, 15 pages (2009).
- [21] P. M. Pilarski, X.-T. Su, D. M. Glerum, and C. J. Backhouse, “Rapid simulation of wide-angle scattering from mitochondria in single cells,” *Optics Express* **16**(17), 12,819–12,834 (2008). URL <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-17-12819>.
- [22] P. M. Pilarski and C. J. Backhouse, “A method for cytometric image parameterization,” *Optics Express* **14**(26), 12,720–12,743 (2006). URL <http://www.opticsinfobase.org/oe/abstract.cfm?URI=oe-14-26-12720>.
- [23] P. M. Pilarski and C. J. Backhouse, “Theoretical foundations for robust wide-angle scattering pattern analysis,” *Optics Express* **submitted**, 5 pages (2009).