Self Organization in Future Cellular Networks

Ali Imran

Submitted for the Degree of
Doctor of Philosophy
University of Surrey

UNIVERSITY OF SURREY

Centre for Communication Systems Research
Faculty of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, U.K.

April 2011

© Ali Imran 2011
Abstract

This PhD thesis focuses on developing novel self-organising functionalities in wireless cellular systems. Since the root cause of suboptimal performance in wireless cellular systems is the mismatch between its semi-static design and its dynamically changing environment, the main objective of self-organising functionalities is to countermeasure the effect of this mismatch. Towards this end, we classify wireless cellular system dynamics based on their time scale into three main classes - short, medium and long term dynamics - and develop a self-organising solution suitable for each class.

Through investigation of case studies of self-organising systems in nature, we first identify the desirable characteristics of self-organising systems. By building on these case studies we propose a general framework called Biomimetic Self Organising Framework (BSOF), for designing adaptive solutions in engineering systems that bear characteristics of self-organisation.

First major contribution of this thesis consists of a novel solution to cope with short term dynamics e.g. pop-up hotspots. This solution optimizes antenna tilts in a distributed manner for system-wide spectral efficiency optimization in face of heterogeneous user geographical distributions. The solution is developed analytically by applying BSOF and its performance is evaluated against centralised fixed-tilting benchmarks through system-level simulations. Results show a 30% improvement in average spectral efficiency along with advantages of a self-organising distributed solution i.e. very low signalling overhead, agility and scalability.

Second major contribution in this thesis provides a novel solution to cope with medium term dynamics e.g. uneven traffic load among cells. This solution optimises cell load through cell coverage adaptation in a distributed manner in order to minimise system-wide average call blocking. The analytical framework behind this solution is developed by following the steps of BSOF. The numerical results show 280% reduction in average blocking probability compared to no load balancing in place. The performance of this distributed solution is also compared against a bench mark of centralised control based load-balancing algorithm. Results show that a performance very close to the centralised solution can be obtained with proposed distributed solution, with added advantages of a distributed solution.

Our final major contribution aims for providing a self-organising functionality for long term dynamics e.g. demographical and socio-economic changes. First we develop a novel framework for long term performance characterisation of wireless systems in terms of three key performance indicators i.e. capacity, quality of service and energy efficiency. Then, by following the steps of BSOF, we develop a novel solution for self-organisation of frequency reuse and deployment architecture for joint optimization of spectral efficiency, fairness and energy efficiency.
Key words: Self Organization, Wireless Cellular Systems, Spectral Efficiency Optimization, Load balancing.
Dedication

Dedicated to My Grandfather

AKBAR ALI

Who is my first teacher, I'll always miss.
Declaration of Originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated by myself in the Centre for Communication Systems Research (CCSR), University of Surrey UK.
Acknowledgements

First of all, I am grateful to my Creator who gave us the cognition to learn, stimulated it through His unfathomable creations and inspired it through revelations like: “(Oh Mankind) Don’t you see, Don’t you think, Don’t you ponder upon (what you see)” - Al-Qurran.

I find myself deeply obliged to my supervisor Professor Rahim Tafazolli for many reasons; including trusting in my abilities and awarding me the opportunity to pursue PhD at CCSR, continuously motivating me to utilise my abilities to their fullest; instead of his extremely busy schedule, allowing me to always avail his visionary guidance and erudite advice that helped me to improve my research skills a lot.

My greatest gratitude goes to my PhD advisor Dr. Muhammad Ali Imran who, despite his pressing work load, has always been there whenever I needed help. His support knew no bounds: spanning from petty issues of latex to big issues of research directions. His knowledgeable guidance, constructive criticism and progressive feedback on my work have helped me a lot to reach this far.

I am also extremely grateful to my friends, with whom I spent my time during my PhD that made it one of the most cherishable periods of my life. In no particular order my deepest gratitude goes to Ammar Hussain Khan, Hassaan Touheed, Kamran Arshad and Muhammad Imran Rafique for being such valuable friends.

I will always run short of words to thank my parents for their prayers, unconditional love and unlimited support. I am also thankful to my siblings and closest friend of mine Atif Qayyoom, as the faith and utmost confidence they had in me have been a great moral booster through the thick and thin of this study period.

Finally, I would like to express my sincerest gratitude to my wife for her care, patience and support during this period. I am particularly thankful to her for sacrificing her share of my time to provide me space and extra time that I needed for my work. Without her continuous support and her understanding of the demands of my work this thesis would not have been possible.
## Contents

Abstract ............................................................................................................... iii
Dedication .......................................................................................................... v
Declaration of Originality ............................................................................... vi
Acknowledgements ......................................................................................... viii
Table of Contents ............................................................................................. viii
List of Figures ................................................................................................... xv
List of Tables .................................................................................................... xx
List of Acronyms ............................................................................................... xxi

1 Introduction

1.1 Context and Background ................................................................. 1
1.2 Motivation ......................................................................................... 3
1.3 Objective of Research .................................................................... 5
1.4 Scope of the Thesis ..................................................................... 5
1.5 Main Contributions and Achievements .................................. 7
1.6 List of Publications .................................................................. 9
1.7 Structure of Thesis .................................................................. 12
1.8 Conclusions ................................................................................. 14

2 Self Organization from Conception to Realization .............................. 15

2.1 Introduction .................................................................................... 15
2.2 What is Self Organization? ....................................................... 16
    2.2.1 Existing Notions ............................................................. 16
<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2 Some Refinements of Notion of Self Organization</td>
</tr>
<tr>
<td>2.2.3 Case Studies of Self Organization in Nature</td>
</tr>
<tr>
<td>2.2.3.1 Flock of Common Cranes</td>
</tr>
<tr>
<td>2.2.3.2 Shoal of Fish</td>
</tr>
<tr>
<td>2.2.4 Definition of Self Organization</td>
</tr>
<tr>
<td>2.3 Characteristics of Self Organization</td>
</tr>
<tr>
<td>2.3.1 Scalability</td>
</tr>
<tr>
<td>2.3.2 Stability</td>
</tr>
<tr>
<td>2.3.3 Agility</td>
</tr>
<tr>
<td>2.4 Distinction between Self Organization and Adaptability</td>
</tr>
<tr>
<td>2.5 Self Organization in Wireless Cellular Systems: A Brief Review</td>
</tr>
<tr>
<td>2.5.1 Individual Works</td>
</tr>
<tr>
<td>2.5.2 Major Research Projects</td>
</tr>
<tr>
<td>2.5.3 Standardization Activity on Self Organization</td>
</tr>
<tr>
<td>2.6 Designing Self Organization in Wireless Cellular Systems</td>
</tr>
<tr>
<td>2.6.1 Biominetics: A Bio Inspired Design Approach</td>
</tr>
<tr>
<td>2.6.1.1 Direct Approach of Biominetics</td>
</tr>
<tr>
<td>2.6.1.2 Indirect Approach of Biominetics</td>
</tr>
<tr>
<td>2.6.2 Considerations in Use of Biominetics</td>
</tr>
<tr>
<td>2.7 A Biominetic Self Organization Framework (BSOF)</td>
</tr>
<tr>
<td>2.7.1 Revisiting SO in Nature: Flock of Common Cranes</td>
</tr>
<tr>
<td>2.7.2 Constituents of Self Organization</td>
</tr>
<tr>
<td>2.7.2.1 Relationship Between SO-Objective and SO-Goal</td>
</tr>
<tr>
<td>2.7.2.2 Designing SO-Functions</td>
</tr>
<tr>
<td>2.7.2.3 Coopetition: An Interesting Principle Behind SO-Functions</td>
</tr>
<tr>
<td>2.7.3 A Generic Biominetic Self Organization Framework (BSOF)</td>
</tr>
<tr>
<td>2.8 Conclusions</td>
</tr>
</tbody>
</table>
# 3 Spectral Efficiency Optimization through SO of Antenna Tilting

3.1 Introduction ................................................. 45
3.2 Background and Related Work ......................... 46
3.3 System Model and Assumptions ......................... 51
3.4 Hot Spot Traffic Relief by Tilt Optimization through BSOF(TO-BSOF) ........................................ 55
   3.4.1 Problem Formulation: Identifying SO-Objective 56
   3.4.2 Transforming SO-Objective into SO-Goals ...... 57
   3.4.3 Decomposing SO-Goal to Design SO-Functions 64
   3.4.4 Solving the SO-Function: Tilt Optimization Problem within a Triplet 69
   3.4.5 Recap of Tilt Optimization through BSOF(TO-BSOF) 74
3.5 System Level Performance Evaluation ................. 76
   3.5.1 CDF of User Throughput ............................ 78
      3.5.1.1 Theoretical Throughput ......................... 80
      3.5.1.2 Throughput with LTE Modulation and Coding Schemes 81
   3.5.2 Expected Spectral Efficiency ....................... 85
   3.5.3 Fairness ................................................... 88
3.6 Practical Implementation of TO-BSOF ..................... 91
3.7 Conclusions ................................................. 92

# 4 Load Balancing through SO of Coverage

4.1 Introduction ............................................... 95
4.2 Background and Related Work ......................... 97
   4.2.1 Resource Adaptation (RA) based Load Balancing 97
   4.2.2 Traffic Shaping (TS) based Load Balancing ...... 98
   4.2.3 Coverage Adaptation (CA) based Load Balancing 100
      4.2.3.1 LB through Antenna Adaptation (AA) ........ 100
      4.2.3.2 LB through Power Adaptation (PA) .......... 102
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.4</td>
<td>Load Balancing with Hybrid Approach</td>
<td>105</td>
</tr>
<tr>
<td>4.2.5</td>
<td>Load Balancing through Relay Stations</td>
<td>105</td>
</tr>
<tr>
<td>4.2.6</td>
<td>Summary</td>
<td>107</td>
</tr>
<tr>
<td>4.3</td>
<td>System Model and Assumptions</td>
<td>108</td>
</tr>
<tr>
<td>4.4</td>
<td>Biomimetic SO Framework for Load Balancing (LB-BSOF)</td>
<td>109</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Problem Formulation: Identifying SO-Objective</td>
<td>109</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Transforming SO-Objective into SO-Goal</td>
<td>110</td>
</tr>
<tr>
<td>4.4.3</td>
<td>Decomposing SO-Goal into SO-Functions</td>
<td>116</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Design Considerations for SO-Functions and Practical Implementation</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td>4.4.4.1 Exemplary Deployment Architecture for Illustration of LB-BSOF</td>
<td>120</td>
</tr>
<tr>
<td>4.4.5</td>
<td>Using Two SO-Goals to Achieve Same SO-Objective</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>4.4.5.1 Coverage Adaptation at BS (BCA)</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>4.4.5.2 Coverage Adaptation at RS (RCA)</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>4.4.5.3 Access Link Adaptation (ALA)</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>4.4.5.4 Beam Switching of RS (BSR)</td>
<td>124</td>
</tr>
<tr>
<td>4.4.6</td>
<td>Recap of LB-BSOF</td>
<td>125</td>
</tr>
<tr>
<td>4.5</td>
<td>Operational Use Cases of LB-BSOF</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>4.5.1 SO at Micro level</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>4.5.2 SO at Macro level</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>4.5.3 SO at Global Level</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>4.5.3.1 Load Balancing Across Operators</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>4.5.3.2 Load Balancing Across Infrastructures</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>4.5.3.3 Self Healing</td>
<td>130</td>
</tr>
<tr>
<td>4.6</td>
<td>A Heuristic Algorithm for Execution of LB-BSOF</td>
<td>131</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Some key Assumptions</td>
<td>131</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Simulation Model</td>
<td>134</td>
</tr>
</tbody>
</table>
4.6.3 LB in Macro Cell based WCS ............................................ 135
4.6.3.1 LB-BSOF with Super cell of 3 cells ......................... 135
4.6.3.2 LB-BSOF with Super Cell of 6 Cells ..................... 139
4.6.4 LB in Heterogenous WCS .................................................. 149
4.6.4.1 LB with Femto cells ................................................... 151
4.6.4.2 LB with RS ............................................................... 152
4.7 Conclusions ........................................................................................ 155

5 Multi Objective Optimization through SO of Frequency Reuse & Deployment Architecture 157
5.1 Introduction ........................................................................................ 157
5.2 Context and Background ..................................................................... 159
5.2.1 Need for SO for Long Term Dynamics ................................... 159
5.2.2 Major System Optimization Objectives on Long Term .......... 161
5.2.3 Major Factors Effecting Long Term Performance of WCS . 162
5.2.4 Relevant Work ........................................................................... 164
5.3 A Performance Characterisation Framework (PCF) ...................... 166
5.3.1 System Model ........................................................................ 167
5.3.2 Relationship between the WCS Performance and Frequency
Reuse and Deployment (FD) .................................................. 168
5.3.3 Proposed Performance Metrics ............................................ 170
5.3.3.1 Effective Spectral Efficiency (ESE) .............................. 170
5.3.3.2 Service Area Fairness (SAF) ....................................... 176
5.3.3.3 Energy Consumption (EC) ....................................... 178
5.4 Long Term Performance of WCS's with different FD's ............. 180
5.4.1 System Model for Performance Evaluation ...................... 180
5.4.2 ESE of Various FD's for WCS ............................................ 182
5.4.2.1 ESE for Conventional FD's ....................................... 182
5.4.2.2 ESE for Relay Enhanced FD's .................................... 183
5.4.3 SAF of Various FD's for WCS ............................................ 186
5.4.3.1 SAF for Conventional FD ........................................ 186
5.4.3.2 SAF for Relay Enhanced FD's ................................. 187
5.4.4 Energy Consumption of Various FD's ............................ 188
5.4.5 Comparison of Performance of Conventional FD's with Relay Enhanced FD's ........................................ 188

5.5 A Novel Self Organising Frequency Reuse and Deployment Framework (SOFD) ........................................ 190
5.5.1 Problem Formulation: Identifying SO-Objective ............ 190
5.5.2 Transforming SO-Objective into SO-Goal ...................... 191
5.5.2.1 Case 1: General Optimization ................................. 192
5.5.2.2 Case 2: Targeted Optimization .............................. 192
5.5.3 Identifying SO-Functions ........................................... 194
5.5.4 Practical Implementation ........................................... 194
5.5.5 Numerical Results .................................................... 196

5.6 Multi-Objective Optimization Vs Constrained Optimization 201
5.7 Conclusions .............................................................. 202

6 Conclusions and Future Work 205
6.1 Conclusions .............................................................. 205
6.2 Summary of Key Achievements ....................................... 207
6.3 Specific Future Work Directions ..................................... 208
6.3.1 Future work on TO-BSOF ........................................... 208
6.3.2 Future work on LB-BSOF ......................................... 209
6.3.3 Future work on SOFD .............................................. 210
6.4 Some General Research Issues in SO .............................. 212
6.4.1 Parallel Operation of Multiple Time Scale SO ............... 212
6.4.2 Parallel Operation of Multiple Spatial Scope SO ............ 212
6.4.3 Challenges in Evaluation of Long Time Scale SO ............ 213
6.4.4 Enabling Self Organization ......................................... 213
6.5 Alternative Design Techniques towards SO ....................... 214
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.5.1 Use of Non Deterministic Techniques towards Design of SO</td>
<td>214</td>
</tr>
<tr>
<td>6.5.2 Supervised Learning</td>
<td>215</td>
</tr>
<tr>
<td>6.5.3 Unsupervised Learning</td>
<td>215</td>
</tr>
<tr>
<td>6.5.4 Deducive Learning</td>
<td>215</td>
</tr>
<tr>
<td>6.5.5 Reinforcement Learning</td>
<td>216</td>
</tr>
<tr>
<td>A Proof of Center of Gravity Proposition</td>
<td>217</td>
</tr>
<tr>
<td>B Proof of Optimality condition for Minimal Blocking in Homogeneous WCS</td>
<td>221</td>
</tr>
<tr>
<td>C Difference between Area Spectral Efficiency and Effective Spectral Efficiency and</td>
<td>225</td>
</tr>
<tr>
<td>D Brief Overview of Optimization Theory</td>
<td>229</td>
</tr>
<tr>
<td>D.1 Types of optimization problems</td>
<td>230</td>
</tr>
<tr>
<td>D.1.1 Convex Optimization</td>
<td>230</td>
</tr>
<tr>
<td>D.1.2 Nonlinear programming</td>
<td>231</td>
</tr>
<tr>
<td>D.1.3 Other Major Optimization Types</td>
<td>233</td>
</tr>
<tr>
<td>Bibliography</td>
<td>235</td>
</tr>
</tbody>
</table>
List of Figures

1.1 Time scale classifications of WCS dynamics ........................................... 6
2.1 An example of group flight of common cranes ........................................... 20
2.2 An example of shoaling behaviour of fish .............................................. 21
2.3 An illustration the possible approaches towards design of SO ..................... 34
2.4 Case Study of SO in nature: Flock of Common Cranes ............................... 38
2.5 Main elements of self organization working in flock of common cranes .......... 39
2.6 A biomimetic Self Organising Framework (BSOF) inspired by case studies of self organization in nature .............................................. 42
3.1 System model for problem formulation for tilt optimization ....................... 52
3.2 User geographical density characterization ............................................. 59
3.3 Transforming SO-Objective into SO-Goal ............................................ 60
3.4 Extracting Center of gravity from Locus of Gravity ................................ 63
3.5 Decomposition of SO-Goal into SO-Function using concept of Triplets ......... 68
3.6 Aggregate throughput in a Triplet .................................................... 70
3.7 A visual illustration of steps of TO-BSOF ........................................... 75
3.8 User geographical distribution for case 1 ............................................. 79
3.9 CDF of theoretical throughput for case 1 ............................................. 80
3.10 CDF of practical user throughput for case 1 ........................................ 82
3.11 CDF of practical user throughput for case 2 ........................................ 83
3.12 Gain of TO-BSOF in terms of expected spectral efficiency per user .......... 87
3.13 Impact of TO-BSOF on Fairness .................................................... 89
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.14</td>
<td>Impact of location of CG on TO-BSOF</td>
<td>90</td>
</tr>
<tr>
<td>4.1</td>
<td>Exemplary deployment architecture for illustration of LB-BSOF</td>
<td>121</td>
</tr>
<tr>
<td>4.2</td>
<td>A visual summary of LB-BSOF</td>
<td>126</td>
</tr>
<tr>
<td>4.3</td>
<td>A heuristic algorithm for execution of LB-BSOF for continuous optimization of user satisfaction</td>
<td>132</td>
</tr>
<tr>
<td>4.4</td>
<td>Key system level simulation parameters for evaluations of LB-BSOF</td>
<td>134</td>
</tr>
<tr>
<td>4.5</td>
<td>HB with super cell consisting of three macro cells</td>
<td>136</td>
</tr>
<tr>
<td>4.6</td>
<td>Geographically uniform user distribution in 3 sector WCS</td>
<td>137</td>
</tr>
<tr>
<td>4.7</td>
<td>User association per sector for uniform user distribution with three cells per super cell</td>
<td>138</td>
</tr>
<tr>
<td>4.8</td>
<td>Average HB with and without LB-BSOF as percentage of minimum blocking for macro, three sector, uniform user distribution scenario</td>
<td>139</td>
</tr>
<tr>
<td>4.9</td>
<td>SB with super cell consisting of three macro cells</td>
<td>140</td>
</tr>
<tr>
<td>4.10</td>
<td>HB and SB with super cell consisting of six macro cells</td>
<td>141</td>
</tr>
<tr>
<td>4.11</td>
<td>Non uniform user distribution used for evaluation of LB-BSOF performance</td>
<td>142</td>
</tr>
<tr>
<td>4.12</td>
<td>HB and SB with super cell consisting of six macro cells with non uniform user distribution</td>
<td>144</td>
</tr>
<tr>
<td>4.13</td>
<td>HB as percentage of minimum blocking achievable with ICC, with and without LB, both for uniform and non uniform user distribution with super cell consisting of six macro cells</td>
<td>145</td>
</tr>
<tr>
<td>4.14</td>
<td>User association with ICC, LB-BSOF and without LB, both for uniform and non uniform user distribution with super cell consisting of six macro cells</td>
<td>146</td>
</tr>
<tr>
<td>4.15</td>
<td>CDF of average load in cells, with ICC, LB-BSOF and without LB, both for uniform and non uniform user distribution with super cell consisting of six macro cells</td>
<td>147</td>
</tr>
<tr>
<td>4.16</td>
<td>Tx power of association control reference signal with ICC, LB-BSOF and without LB, for non uniform user distribution with super cell consisting of six macro cells</td>
<td>149</td>
</tr>
<tr>
<td>4.17</td>
<td>Energy Efficiency potential of LB-BSOF</td>
<td>150</td>
</tr>
</tbody>
</table>
List of Figures

4.18 Performance of LB-BSOF compared to no LB and ICC in Heterogeneous WCS containing Femto cells ............................................... 151
4.19 Loading in the heterogenous WCS with Femto Cells ......................... 152
4.20 HB in heterogenous WCS with RS. Blocking on coverage as well as access link is logged .......................................................... 153
4.21 Total HB in heterogenous WCS with RS ............................................... 154

5.1 Traffic variation over hours of the day [1] .............................................. 160
5.2 List of potential FD’s for future WCS .................................................. 181
5.3 List of potential FD’s for future WCS .................................................. 181
5.4 ESE for different number of sectors and relay stations per site and frequency reuse factors ............................................................... 182
5.5 ESE for different number of sectors and relay stations per site and frequency reuse factors ............................................................... 184
5.6 CDF of SINR distribution in the coverage area for different number of sectors per site and frequency reuse factors ................................... 185
5.7 CDF of SINR distribution in the coverage area for different number of sectors and relay stations per site and frequency reuse factors ................................... 185
5.8 SAF for different number of sectors frequency reuse factors ............... 186
5.9 SAF for different number of sectors and relay stations per site and frequency reuse factors ............................................................... 187
5.10 Fixed power consumption per site ..................................................... 189
5.11 Variable power consumption per site ................................................... 189
5.12 Total power consumption per site ..................................................... 190
5.13 SAF, ESE and EC normalized by their respective maximum value in conventional WCS ............................................................... 196
5.14 ESE, SAF and EC normalized by their respective maximum value in relay enhanced WCS ............................................................... 197
5.15 Value of general optimization utility function $v_g$ for different priorities in terms of ESE, SAF and EC for conventional WCS ........ 198
5.16 Value of general optimization utility function $v_g$ for conventional FD’s with different priorities in terms of ESE, SAF and EC.. 198
5.17 Value of targeted optimization utility function $v_t$ for convectional FDs with different target values of ESE, SAF and EC ............ 199

5.18 Value of targeted optimization utility function $v_t$ for relay enhanced convectional FDs with different target values of ESE, SAF and EC. ................................................................. 200

D.1 Optimization problem classification table ........................................ 230
List of Tables

3.1 System Level Simulation Parameters ............... 77
3.2 Threshold SINR's for LTE modulation and coding schemes [2] .. 85
## List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>Antenna Adaptation</td>
</tr>
<tr>
<td>ALA</td>
<td>Access Link Adaptation</td>
</tr>
<tr>
<td>AMC</td>
<td>Adaptive Modulation and Coding</td>
</tr>
<tr>
<td>ASE</td>
<td>Area Spectral Efficiency</td>
</tr>
<tr>
<td>BSR</td>
<td>Beam Switching of Relay Station</td>
</tr>
<tr>
<td>BCA</td>
<td>Base station Coverage Adaptation</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>BSOF</td>
<td>Biomimetic Self Organisation Framework</td>
</tr>
<tr>
<td>CA</td>
<td>Coverage Adaptation</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CDMA</td>
<td>Code Division Multiple Access</td>
</tr>
<tr>
<td>DA</td>
<td>Deployment Architecture</td>
</tr>
<tr>
<td>DL</td>
<td>Docitive Learning</td>
</tr>
<tr>
<td>ESE</td>
<td>Effective Spectral Efficiency</td>
</tr>
<tr>
<td>EC</td>
<td>Energy Consumption</td>
</tr>
<tr>
<td>FD</td>
<td>Frequency Reuse and Deployment</td>
</tr>
<tr>
<td>GSM</td>
<td>Global System for Mobile Communication</td>
</tr>
<tr>
<td>HB</td>
<td>Hard Blocking (i.e. blocking due to lack of resources)</td>
</tr>
<tr>
<td>HSR</td>
<td>Hotspot Radius (as percentage of cell radius)</td>
</tr>
<tr>
<td>ICC</td>
<td>Ideal Central Control based load balancing algorithm</td>
</tr>
<tr>
<td>LB</td>
<td>Load Balancing</td>
</tr>
<tr>
<td>LB-BSOF</td>
<td>Load Balancing through BSOF</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>LTE-A</td>
<td>Long Term Evolution-Advanced</td>
</tr>
<tr>
<td>MCE</td>
<td>Modulation and Coding Efficiency</td>
</tr>
<tr>
<td>MCS</td>
<td>Modulation and Coding Scheme</td>
</tr>
<tr>
<td>OFDM</td>
<td>Orthogonal Frequency Division Multiplexing</td>
</tr>
<tr>
<td>OFDMA</td>
<td>Orthogonal Frequency Division Multiple Access</td>
</tr>
<tr>
<td>PCF</td>
<td>Performance Characterisation Framework</td>
</tr>
<tr>
<td>PA</td>
<td>Power Adaptation</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>RA</td>
<td>Resource Adaptation</td>
</tr>
<tr>
<td>RCA</td>
<td>Relay station Coverage Adaptation</td>
</tr>
<tr>
<td>RL</td>
<td>Reinforced Learning</td>
</tr>
<tr>
<td>RS</td>
<td>Relay Station</td>
</tr>
<tr>
<td>SO</td>
<td>Self Organization</td>
</tr>
<tr>
<td>SO-Objective</td>
<td>Self Organization Objective</td>
</tr>
<tr>
<td>SO-Goal</td>
<td>Self Organization Goal</td>
</tr>
<tr>
<td>SO-Function</td>
<td>Self Organization Function</td>
</tr>
<tr>
<td>SAF</td>
<td>Service Area Fairness</td>
</tr>
<tr>
<td>SB</td>
<td>Soft Blocking (i.e. call rejection due to high interference)</td>
</tr>
<tr>
<td>SOFD</td>
<td>Self Organising Frequency Reuse and Deployment</td>
</tr>
<tr>
<td>SoM</td>
<td>Self Organising Maps</td>
</tr>
<tr>
<td>TDF</td>
<td>Traffic Distribution Factor (e.g. TDF=x% means x% users are in hotspots)</td>
</tr>
<tr>
<td>TO</td>
<td>Tilt Optimization</td>
</tr>
<tr>
<td>TO-BSOF</td>
<td>Tilt Optimization through BSOF</td>
</tr>
<tr>
<td>TS</td>
<td>Traffic Shaping</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Mobile Telecommunications System</td>
</tr>
<tr>
<td>WCS</td>
<td>Wireless Cellular Systems</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Context and Background

Quest for unprecedented higher data rates, better Quality of Service (QoS) and need for energy efficiency is pushing the cosmos of Wireless Cellular System (WCS) to its limits. On top of that, plethora of bandwidth-hungry applications and quality conscious services are mushrooming with every passing day. This fact complemented by the timely advent of myriad of mobile computing devices that can support these applications and services, has given rise to a rapidly mounting demand for broadband on the move. From service provider's point of view, this means that WCS should also have what it takes to support these newly born applications and services. Therefore, recently there are widely heard new big bangs ranging from physical layer technologies like OFDM/OFDMA and MIMO to system wide concepts like cognitive radio, in order to expand the cosmos of cellular networks. Self Organization (SO) is one such big bang that has the potential to open up new dimensions to expand the cosmos of wireless systems in terms of capacity, QoS, energy efficiency and cost effectiveness. In terms of formal definition, a system is said to have SO if it can organize itself without any
external or central control entity to achieve and maintain a certain objective or set of objectives [3, 4].

SO is intrinsically a nature inspired phenomenon. In nature there are many systems that can autonomously organize themselves to cope best with their continuously changing environments to achieve and maintain their desired optimal objectives. Most well known examples of such systems are; school of shoaling fish, swarming insects, herding sheep, synchronously flashing fire flies and flocking birds, to name a few. Each of these natural systems is capable to maintain its intended operation in optimal or near-optimal state despite of unpredictable variations in its operational environment. For example, a flock of birds maintains optimal flight efficiency while flying, indifferent to the number of birds in the flock or weather or wind conditions. Most importantly, these self organising natural systems are capable to operate without any central or external control. In a nutshell, SO system in nature are scalable, stable, and agile. These features make SO a very desirable feature in any engineering system in general.

An operational environment of a WCS is as dynamic and unpredictable as the environment faced by natural self organizing systems. Changing demography, user distribution, mobility patterns, traffic demands and geographical and socio-economic profile of the coverage area in which a WCS is deployed are few of the major causes behind the acute spatio-temporal dynamics of WCS. The root cause of underachievement of major performance objectives in WCS (namely the optimal capacity, QoS or energy efficiency) boils down to conventional WCS’s lack of ability to adapt itself to these spatio-temporal dynamics. Therefore, it is foreseeable that if SO can be designed into future WCS, a significant boost in performance in terms of all major system objectives can be achieved. This boost in performance is possible because a cellular system with SO capability can adapt its system design and operational parameters to suit the changing environment.

\footnote{These three characteristics will be investigated in detail in chapter 2 through case studies}
and thus make the most of available resources at all places at all times.

SO has been well investigated in the context of the ad-hoc and mesh networks as they are infrastructure-less and essentially need some kind of organization capability. But for WCS that have a well defined infrastructure, the concept of SO is relatively easy to conceive because of presence of architecture; yet it is difficult to deliver because of the lack of flexibility arising from the same architecture based nature. Therefore, relatively much less work has been done for SO in WCS. Legacy cellular systems like GSM and UMTS neither contained any noticeable SO features nor any standardization in this respect was carried out.

As highlighted above, escalating performance expectations from future WCS now means SO is not an optional but a highly needed feature. Therefore, 3GPP has considered SO as an inevitable feature of future cellular network like LTE and LTE-A [5]. This standardization support has triggered a wide interest in industry and academia to explore the new research dimension brought forth by the big bang of SO and this study embodies such an effort.

1.2 Motivation

The future of Wireless Cellular Systems (WCS) is marked by a drastic change in user behaviour triggered by the rampant growth of bandwidth hungry applications supportable by plethora of newly born smart mobile devices. While the surge in user applications is only bounded by imagination, the capacity of WCSs, that has to support these applications, is tightly bounded by fundamental physical limits. This problem is further aggravated by considering financial constraints from operator's point of view as higher capacity and QoS comes at a cost of higher capital expenditure (CAPEX) and operating expenditure (OPEX). Since users are naturally reluctant to pay proportionally higher bills for improved services,
minimizing CAPEX and OPEX in order to make the business model commercially viable is a crucial consideration for operators, while seeking to provide better QoS and capacity. As a result operators in legacy WCS are striving to reach a trade-off between providing improved services and retaining reasonable profits.

The requirement to meet the needs of both users and operators in a cost effective way has triggered research in adding SO into future WCS. This initiative is motivated mainly by the following major factors.

1. Although optimal capacity of wireless channel is known to have physical upper bounds, however there are inherently unpredictable spatio temporal dynamics associated with WCSs. Due to these dynamics, the optimal performance in terms of capacity and QoS could not be achieved with fixed and rigid design of legacy WCSs that lack flexibility to autonomously adapt with these dynamics. Therefore, due to mobile nature of users and varying nature of wireless channel, WCSs suffer from under-utilization of resources resulting in low resource efficiency at some point and over-utilization resulting in congestion and bad QoS at other point, either in time or space.

2. With increasing deployment of outdoor relays or Pico cells in seek of improved performance in terms of capacity and QoS [6], the number of nodes in future WCS are anticipated to be too large to be configured, optimised and maintained for regular operation with classic manual and field trial based approach.

3. Given the huge scale of future WCS, the classic manual approach for continual optimization required during the operational phase will not only cease to be technically viable, but the high investments to ensure Quality of Service (QoS) can be jeopardised by a simple human error, oversight or long recovery time to restore system.
1.3. Objective of Research

4. Finally, in addition to the improved performance, SO can significantly reduce the OPEX by eliminating the need for expensive skilled labour required for configuration, commissioning, optimization, maintenance, troubleshooting and recovery of WCS.

In summary, SO is the most cost effective and technically viable way of achieving and maintaining the optimal performance in future WCS. Standards for Long Term Evolution (LTE) and LTE-Advanced [7] have therefore identified self-organization as not just an optional feature but an inevitable necessity in future WCS [8]. Aforementioned facts provide motivation for this study to investigate SO in specific context of WCS.

1.3 Objective of Research

The broad objective of this study is to design self-organization in wireless cellular systems. More specifically this study aims to first develop a generic framework to design self-organization by investigating its characteristics and the underlying principles that lead to these characteristics through case studies from nature i.e. the original source of inspiration for SO. Then by applying that framework, we aim to design mechanisms/algorithms that are self-organizing in their manifestation and thus can equip the future cellular networks to cope with the different types of spatio temporal dynamics in order to improve capacity, QoS and energy efficiency.

1.4 Scope of the Thesis

In order to clarify the scope of this thesis we categorise the spatio temporal dynamics a WCS faces in four main categories as shown in the figure 1.1. It is these
Chapter 1. Introduction

Dynamics in WCS

Very Short Term: |is«c to sec
SO Example: ACM, scheduling, conventional RRM

Short Term: minutes to hours
SO Example: Electronic tilt adaption

Medium Term: hours to days
SO Example: Load balancing with coverage adaptation

Long Term: Days to Months
SO Example: Adaptive deployment

Classification of SO algorithms on the basis of their time scale of operation

Figure 1.1: Time scale classifications of WCS dynamics

dynamics that make the achievement of optimal performance objectives in WCS a challenging task. The very short term dynamics, that make the first category, have been extensively addressed in the regimes of classic radio resources management (RRM) and physical layer research. Second category consists of short term dynamics like temporary shadowing, slow user mobility and pop up hotspots. Third category mainly consists of medium term dynamics such as permanent shadowing, permanent hotspots, natural changes in user's spatial concentrations over time and resultant load imbalance in WCS. Fourth category includes long term dynamics such as change in performance objectives and their mutual priorities for operators e.g due to change of demographic or socio-economic profile of coverage area or change in traffic profile over seasons of year.

The scope of this thesis is focused on the last three types of dynamics. More specifically we propose one SO mechanism for each of these three categories of dynamics. This classification is mainly intended to indicate the immediate applications of the solutions proposed. Otherwise, each of the three solutions has potential to cope with variety of short to long time scale dynamics of WCS ranging
from minute to years. It is important to note that, very short time scale dynamics that range from micro second to seconds, usually caused by channel variation and fast fading are well dealt through classic adaptive radio resource and power management schemes. Effect of these very short time scale dynamics is averaged out while considering dynamics like user distributions over a significantly long time period, therefore, such very short scale dynamics are not included in scope of this thesis.

1.5 Main Contributions and Achievements

The major novel contributions and achievements accomplished in this thesis can be summarised as follows.

1. Identification of basic characteristics of SO systems and their underlying principles leading to these characteristics.


3. A novel framework for optimization of system wide antenna tilts (TO-BSOF i.e. Tilt Optimization through BSOF) for maximisation of spectral efficiency and throughput with focus on short term dynamics.

   (a) Formulation of aggregate system throughput optimization problem as a function of system wide antenna tilts

   (b) A novel metric to characterise user geographical distribution and hot spots.

   (c) A novel concept of center of gravity of user distribution and the proof of its usefulness.

   (d) Simplification of the problem in 3a using concepts in 3b and 3c.
Chapter 1. Introduction

(e) An analytically derived solution of simplified problem in 3d designed with in built SO on footsteps of BSOF

(f) A pragmatic algorithm to implement TO-BSOF.

(g) Performance evaluation of TO-BSOF through extensive system level simulations.

4. A Novel Framework for SO for optimal Load Balancing (LB-BSOF) to minimize blocking and maximise user satisfaction with focus on medium term dynamics for WCS.

(a) Formulation of system wide average user satisfaction maximisation problem as optimization function of radio resources allocated to cells and traffic in the cells.

(b) Analytical solution of problem in 4a.

(c) A novel concept of super cell, proof of its advantage and its use to make the solution in 4b distributed and SO by exploiting BSOF.

(d) Identification of set of use cases of LB-BSOF.

(e) An Ideal Central Controller (ICC) based algorithm to implement LB-BSOF for optimal user satisfaction.

(f) A pragmatic SO algorithm to implement LB-BSOF for near optimal user satisfaction using concept in 4c.

(g) Performance evaluation of both SO and centralised LB-BSOF through extensive system level simulations for 3 sector, 6 sector, and 6 sector relay enhanced WCS.

5. A novel Performance characterisation Framework (PCF) for WCS

(a) Three novel performance metrics to characterise spectral efficiency, service area fairness and energy efficiency aspects of WCS.
List of Publications

Journals


Chapter 1. Introduction


Peer Reviewed Conferences


1.7 Structure of Thesis

With reference to the list of contributions in section 1.5 this thesis is organised as follows.

Chapter 2: presents contribution 1 and 2.

Given all the ambiguity associated with the term self organization, this chapter first discusses out these ambiguities and thus lays down a solid foundation for rest of the work presented in this thesis. Building on a review of evaluation literature on SO in WCS and case studies of SO in nature, the key characteristics of SO are identified. These characteristics are further used to put forth a definition of SO that leaves full freedom for the diversity of techniques with potential to design SO.

Based on the case studies, and the characteristics of SO identified, a biomimetic generic framework for designing self organization (BSOF) is developed. BSOF lays out the three general steps of designing SO. These steps are: identification of SO-Objective, its transformation to simpler SO-Goal(s) and decomposition of SO-Goal(s) to basic and local SO-Functions. A unique principle of coopetition is also identified and explained as the ingredient of SO. BSOF and characteristics of SO, are the key contribution of this chapter that pave the way for the rest of the chapters.

Chapter 3: presents contribution 3.

The problem of hot spots arising from short to medium scale dynamics is addressed. Unlike previous works that addressed hot spot problem through load balancing, a novel approach of optimizing spectral efficiency at the hotspots is introduced and used. This approach has advantage that it does not necessitate handovers that are required in load balancing based hotspot relief. This advantage makes this approach suitable for even short time scale dynamics in OFDMA based WCS. A novel analytical solution namely TO-BSOF for spectral efficiency
optimization through SO of system wide antenna tilts is developed using the principle of BSOF. A novel concept of center of gravity is introduced to simplify the problem into SO-Goal. A novel concept of triplet of the sectors is then exploited to make the solution distributed and localised hence scalable and agile. Performance of TO-BSOF is evaluated through extensive system level simulations for number of heterogenous user distribution scenarios and is compared against a range of other available bench marks in terms of different performance metrics. With the proposed SO solution, upto 30% gain in spectral efficiency is observed with no significant cost in terms of signalling.

Chapter 4: presents contribution 4.

In this chapter a problem of congestion arising from medium to large time scale dynamics is addressed and a novel SO solution namely LB-BSOF is developed. The problem is formulated as global blocking minimization problem and is solved analytically with help of set of mathematical propositions. A novel concept of super cell is introduced and proved, to make the solution distributed and localised hence scalable and agile. Practical implementation and use cases of LB-BSOF are also discussed in great length. Being designed on principle of BSOF, LB-BSOF bears all the features of SO. Finally the performance of LB-BSOF is evaluated through extensive system level simulations and compared against two bench marks i.e. blocking with no load balancing mechanism and blocking with an ideal central control based load balancing. A reduction in blocking by 280% compared to no load balancing case is observed.

Chapter 5: presents contribution 5, and 6.

First three novel metrics constituting PCF are derived to capture the long term performance of WCS by incorporating the effect of the long term design features of WCS i.e. frequency reuse, number of sectors, number of relay stations per site
Chapter 1. Introduction

and the set of modulation and coding schemes. Then building on the PCF, SOFD is presented. Both PCF and SOFD are demonstrated through semi analytical results.

Chapter 6: presents a conclusive summary and the crux of achievements of this thesis. There we also discuss the directions for the future work. We discuss both the specific directions and general directions. Specific directions are based on direct extension of the work presented in this thesis whereas general directions highlight general research areas indirectly linked to our work. We also discuss a number of non deterministic resourceful problem solving techniques that can be exploited to design SO for future work.

1.8 Conclusions

In this chapter we presented the key motivation behind this study and clarified the scope of this thesis by classifying the dynamics of cellular system. The main contributions presented this thesis were listed and organisation of chapter was clarified. Having clarified the motivation and the scope of this thesis, we devote next chapter to establish basic understanding SO and its characteristics. There we also develop some key design principles of SO that will serve as a foundation for rest of the chapters in this thesis.
Chapter 2

Self Organization from Conception to Realization

2.1 Introduction

The use of the term SO in strict context of wireless communications is fairly new, a decade old only [9]. It was originally conceived through inspiration from field of natural sciences where it has been observed that certain natural systems exhibit unique self organizing behavior to obtain their desired objectives even in face of acute dynamics of their natural environment. Since its conception, designing SO into an engineering system has almost drifted from technical issue to a philosophical debate as different authors have different semantics and perceptions of the concept of SO. In this chapter, we do not aim to join this debate yet try to clarify the ambiguity associated with the term SO. This objective is achieved by explaining most pertinent existing notions followed by our notion of SO. Our notion is inline with the main stream literature, yet offers a more concise definition of SO. In order to build a clear foreground for our work, we go a step further than defining SO and identify and explain a set of characteristics that
are desirable in a mechanism or algorithm to be called self organising. This is followed by brief time line styled review of the literature that specifically focuses on self organization in WCS. We conclude this chapter by presenting our novel approach towards designing SO that will be followed in rest of the thesis.

2.2 What is Self Organization?

In this section we present a crux of different definitions of SO presented in literature so far with aim to clarify the ambiguities associated with buzz word self organisation. We present a framework consisting of set of features that fully characterises SO from view point of its practical design and applications. Building on this framework we conclude this section by providing a complementing definition of SO

2.2.1 Existing Notions

Self organisation has been defined in various fields including mathematics, biology, computer science, thermodynamics and cybernetics [10]. However it should be noted that there is still no widely accepted general definition. Even in the specific context of WCS, the semantics of this term are different in various pieces of literature. Here we present the notions of SO that are most pertinent to the context of WCS in particular and wireless communications systems in general.

In context of WCS, the term SO was first used by Spilling et al [9]. In this work authors expressed SO as a feature of network where it can detect changes and based on these changes makes intelligent decisions to minimise or maximise the effect of these changes [9]. This notion of SO is similar to conventional adaptability with the difference that instead of being applicable to particular functionality, say power control or modulation and coding, the concept of adaptability is scaled up
2.2. What is Self Organization?

to system level in a holistic way. An important observation on concept of SO presented in [9] is that it implicitly assumes a central control for detection and feedback process, to trigger the required adaptation. While this work put forth the novel concept of adaptive WCS, it hardly provides a discerning definition of SO.

Yanmaz et al. viewed self organizing system as a system in which entities work cooperatively in response to changes in the environment in order to achieve certain goals [11]. This notion of self organization is more granular as it characterises behaviour of individual entities of system and identifies that these individual behaviours should emerge into a system level self organizing behaviour. With this approach towards SO authors in [11] tend to define SO on the basis of how it can be designed in system while considering it as a feature of a system as a whole. Furthermore, it should be noted that the need for cooperation among entities without putting a bound on the scope of this cooperation has been integrated in this definition of SO. The definition in [11] is more plausible as it does not simply assume a central control as in [9] and builds on the inspiration from SO in some biological systems, but the cooperation assumed here might require extensive signalling among the entities of the system. This kind of explicit signalling is not observed in biological self organising systems and neither it is a desirable phenomenon in context of cellular system.

Prehofer et al. in [4] provides a more generic definition of SO system as follows: "A system is self-organized if it is organized without any external or central dedicated control entity". The definition of self organization provided in [4] not only avoids need for central or external control but in this paper authors further argue that in order to be called SO, system should have more than just distributed control. It emphasises that in self-organized systems, the execution of rather simple behaviours at the microscopic level leads to a coherent behaviour of the overall system. While this notion is closest embodiment of SO in natural systems and it is one
of the most comprehensive descriptions of SO in literature pertinent to WCS, it also approaches SO as holistic feature of a monolithic system. Furthermore, authors in [4] focus directly on the question of how a behaviour similar to SO in natural system can be designed in communication systems without engaging deep in defining SO more precisely.

In summary, though there are various notions of SO in literature as discussed above, yet to the best of our knowledge, there is no generally accepted definition of SO that can be used to classify systems or their functionalities as self organizing or non self organizing because of gaps and differences among existing notions of SO and its inherent overlapping with conventional adaptation. In next two sections we provide our notion of self organization that attempts to fill this gap and bridges up the differences in existing notions while clarifying the boundary between adaptation and SO.

### 2.2.2 Some Refinements of Notion of Self Organization

This section presents some refinements of the existing notion of SO to clarify some of the ambiguities associated with it. These refinements should ultimately help us reach clearer definition of SO. There are following three aspects of notion of SO that need to be clarified.

Firstly, SO should not be considered as a holistic feature of a monolithic system. Rather considering SO as feature of an individual functionality of a system is more logical approach. This refinement is in line with the perspective explained in [12] and [13] that wireless systems with some sort of SO do not form a new class of wireless systems rather SO should be viewed as a feature of these systems. The tendency to approach SO as feature of a system as a whole comes from the fact, that in natural systems that are source of inspiration for SO, the functionalities that have SO become very prominent e.g. V-shape formation of birds while
2.2. What is Self Organization?

flocking, shoaling of fish, synchronous flashing of fire flies. But the same systems, namely group of birds, fish or fire flies, might have many other functionalities e.g. searching for food, or breeding etc that might not be self organizing. Since these functionalities remain in background as they do not emerge as collective prominent behaviour so we tend to neglect them, and tend to classify the whole system as SO or non-self organizing.

Secondly, contrary to viewpoint in [11] and [4], SO need not be defined based on how it is achieved. Rather a definition of SO on the basis of what it can achieve is more relevant in engineering systems. In other words, although, the emergence of the system wide coherent behaviour as result of local behaviour is a key observation in the design of natural systems with SO, but it should not be seen as defining characteristic of SO. The first rational behind this viewpoint is that there are examples in nature, like homeostatic operational control in living beings, where such an emergence is not present or is hidden from the observer but still SO exists [13]. Therefore, in order to have sustainable definition of SO it should be defined on the basis of its performance characteristics rather than on the basis of its design. The approach towards designing SO should be left open to invite a variety of different inspirations from natural as well as non natural systems. The second rational behind the perspective of not using design based characterisation of SO is the fact that there are many non biological systems that have been observed to have SO, e.g economy in a free market is one such example. The success of game theory to devise adaptive and some times truly SO algorithms, is proof of this fact. We will further explain another important rational behind this argument in section 2.6.2 on design approaches towards SO.

Finally, SO should be clearly distinguished from adaptiveness, but at the same time, it should not be essentially considered as only arising from high degree of intelligence in the system. Though, assuming SO as an outcome of intelligence is very intuitive notion, but it is equally vague and difficult to design from practical
Chapter 2. Self Organization from Conception to Realization

Figure 2.1: An example of group flight of common cranes. Flock of common cranes is SO system in nature [14]

point of view, due to all the ambiguities associated with intelligence itself. We rather suggest, from a viewpoint of design and operation of system, it is more pragmatic to think on the line that intelligence actually can emerge from SO. In order to support this argument and distinguish SO form adaptiveness, in next section we present two well-known case studies of perfect self organizing systems in nature. We will finally conclude this discussion with a more sustainable definition of SO by building on these case studies and refinements of existing notions above.

2.2.3 Case Studies of Self Organization in Nature

2.2.3.1 Flock of Common Cranes

A flock of common cranes adapts its flight attributes and flies such that the average flight efficiency of the whole flock is maximised by minimising the average air drag each bird faces by up to 70% compared to individual bird flight efficiency [15] by dynamically maintaining group flight optimal delta formation during flight.
2.2. What is Self Organization?

Figure 2.2: An example of shoaling behaviour of fish. A school of fish is SO system in nature.

Most importantly, it does so:

• without a leader and without global (flock-wide) information exchange or coordination among all birds of the flock i.e. no explicit global signalling and no central control (Scalability)

• without swaying from the long term direction of flight or breaking apart even in face of changing wind and weather conditions. (Stability)

• without loosing their ability of acutely execute individual as well as collective maneuvering to avoid predator attacks and large hurdles (Agility)

2.2.3.2 Shoal of Fish

A shoal fish (figure 2.2) dynamically adapts its shape with environment to scare away or avoid large predators [16] and supposedly gain some hydrodynamic effi-
ciency \[17\] and it does so:

- without a leader and without global information exchange or coordination among all fish of the school i.e. no global signalling, no central control (Scalability)

- without swaying from the long term direction of swim and breaking apart even in face of intrusions. (Stability)

- without losing their ability of fast response to changes in their operational environment (Agility)

In next section we use the key observations in these case studies to present a definition of self organization that is inline with our refined notion of SO discussed in section 2.2.2

### 2.2.4 Definition of Self Organization

Both of the systems described in above section 2.2.3 behave in perfectly self-organising manner i.e. they are adaptive and autonomous. As result of this intrinsic self organization built in them, they manifest certain degree of intelligence in their behaviour. But it should be noted that no single member of the system is intelligent enough individually to control or lead the whole system to result into kind of self organization these systems exercise. Neither all of the members combine their aggregate intelligence to achieve self organization as a whole as there is no central authority or global communication among all of them. Hence, it can be inferred from close observation of these and other similar natural systems that it is self organization that emerges from certain characteristics of the system that in turn may manifest some degree of intelligence. Based on the observations in nature, these characteristics of self-organising system can be identified
to be scalability, stability and agility as categorically highlighted in case studies in section 2.2.3. Hence without engaging into the details of its design process, a complementing definition of SO built on the refinements in section 2.2.2 would be:

'An adaptive functionality in a system is said to be self-organising if it is scalable, stable and agile enough to maintain its desired objective(s) in face of all potential dynamics of its operational environment autonomously'.

To translate this definition from context of biological systems into the exact context of WCS, next section explains these three essential characteristics of self organization i.e. scalability, stability and agility, specifically in context of WCS. These characteristics can be used to not only establish a more concrete and technically achievable definition of SO but can also be used in assessing the degree of SO in solutions presented in literature.

2.3 Characteristics of Self Organization

2.3.1 Scalability

Scalability means the complexity of a solution should not increase unboundedly along with the scale or size of the problem. The characteristic of scalability in natural systems results from the fact that global behaviours mostly emerge only from local and simple behaviours. The same approach can be followed in engineering systems to ensure scalability. More precisely, in order to ensure that an adaptive algorithm or solution is scalable, following two main factors should be considered in the design process:

- Minimal complexity: the algorithm should be austere in terms of time, space and other hardware resources required to implement it.
• Local cooperation: algorithm should not require global cooperation or signalling, rather local coordination should be relied upon where possible. This would reduce any overheads because if cooperation among all nodes is required for implementation of an algorithm, its overheads will increase as the number of nodes increases in the system making it unscalable.

A simplistic way of describing scalability in context of SO is through condition below.

\[
\lim_{n \to N} \frac{d}{dn} O(n) \leq C
\]

(2.1)

Where \( O \) represents implementation complexity of an algorithm as function of number of nodes \( n \) over which the algorithm needs control or coordination. \( N \) is the total number of nodes in the system over which the objective of the algorithm is to be achieved. \( C \) is a constant that can be zero for perfectly scalable solution. For example, consider an adaptive algorithm that changes the antenna tilts of all nodes in the system with an objective of system wide load balancing. If the algorithm requires complex coordination among all nodes centrally, implementation complexity of this adaptive algorithm increases as the number of base stations increases in the system, making it less scalable hence less self organizing. In order to be perfectly scalable this algorithm should ideally operate over \( m \) nodes independently such that \( m << N \) nodes still achieving the desired system wide objective among \( N \) nodes. With this sort of distributive design perfect scalability i.e. \( C = 0 \) can be achieved in large systems. Although with advent of unfathomable computational power, scalability has become relative notion and even centralised complex systems can be afforded. Nevertheless, in the context of pure SO, scalability achieved through both minimal complexity and local control are highly desirable as it helps to achieve other two characteristics of SO as discussed below.
2.3.2 Stability

Stability has different precise meanings in different contexts but in this particular context of SO we use a simplistic but generic definition of stability i.e. an algorithm or adaptation mechanism is stable if it has finite number of states and finite time to traverse all its states. Mathematically, we can write condition for stability as follows

\[
\lim_{t \to T_b} \sum_{s_m \in S_d} P(s_i(t) \rightarrow s_m(t)) = 1, \quad \forall s_i \in S
\]

\[
T(s_i(t) \rightarrow s_m(t)) \leq T_b, \quad \forall s_m \in S_d
\]

where \( S \) is set of all possible states in algorithm and \( S_d \) is set of desired states only such that \( S_d \subseteq S \) and \( s_i \) is arbitrary current state such that \( s_i \in S \) and \( s_m \) is an arbitrary future state such that \( s_m \in S \). The symbol \( \rightarrow \) represents transition from one state to other, \( P \) and \( T \) represent probability and times required for these transitions. \( T_b \) is an arbitrary bounded number. First condition ensures that system does not get stuck in a given state i.e. system is robust and does not have a failure state or point of no return or chaos in it. This condition also means system should have some kind of self healing capability as well, if a disaster or unusual operational environment is one of the possible anticipated states in which system might have to operate, it should ultimately recover back to one of the desired states in finite time. Second condition ensures system does not have infinite oscillations and ultimately reaches to stable state in a bounded time while transiting form one state to other. Designing stable algorithms is an issue very specific to the nature of the algorithms and is separate field of research in itself. For the scope of this study this simple notion of stability is concise enough as through first condition it ensure robustness and through second it ensures system does not have ping pong effect. Although robustness is fundamental requirement of SO but it need not be considered as separate criterion for SO, rather it can be considered as part of stability. It is important to mention that, perfect stability also requires that...
system should be elastic and self healing. That in turn, generally means that single point failures or chaotic design should be avoided that mostly results from a centralised control.

### 2.3.3 Agility

An adaptive system can be scalable and stable but still not perfectly self organizing as it might be sluggish in its adaptation. Agility is an other key characteristic of self organizing systems (as observed in the case studies in above). Agility describes how supple or acutely responsive an algorithm is in its adaptation to the changes in its operational environment. That is, in order to be self organizing, algorithm should not only have capability to adapt to cope with its changing environment (stability), it should also not be sluggish in its adaptation (agility). Furthermore, the algorithm must not be over reactive to temporary changes in the system to prevent oscillation between states as discussed in subsection on stability. Here we present a simple metric to evaluate the degree of agility of an adaptive system.

\[
A = \frac{1}{S} \sum_{s=1}^{S} \left( \frac{t_{e+1}^s - t_{a+1}^s}{t_{e+1}^s - t_a^s} \right) \times 100
\]  \hspace{1cm} (2.4)

where \(S\) is the total number of possible states to which the operational environment of the algorithm can change, and it also represents the corresponding stages of the algorithm. The time \(t\) represents the duration associated to switching to a particular state while superscripts \(a\) and \(e\) represent the algorithm and environment respectively. Thus \(t_a^s\) and \(t_{e+1}^s\) are times at which the environment is in \(s^{th}\) and \((s+1)^{th}\) state respectively and \(t_{e+1}^s\) is the time at which algorithm reaches to \((s+1)^{th}\) state in response to shift of environment to \((s+1)^{th}\) from \(s^{th}\) state.

It should be noted that \(A=0\) means system is not adaptive at all and \(A > 100\%\) means system is over agile i.e. it is not stable and can have oscillations. For a stable adaptive systems values of \(A\) vary from 0\% to 100\% with 100\% being
a perfectly agile solution. Perfect agility is a tough condition to be met in real systems because of the feedback, processing and decision making, and actuation delays involved in any physical system. To explain this measure of agility we consider an example of power control algorithms. Consider three different adaptive power control algorithms X, Y, and Z that adjust user uplink power with respect to its distance from the BS. We assume all three algorithms satisfy the conditions of scalability and stability. We consider an operating environment of the power control algorithm that changes through following 4 states i.e. user changes its location relative to BS such that its distance acquires values given in the set \( d = [600m, 500m, 400m, 500m] \) at times \( T^e = [0,2,5,6] \) in seconds. If the algorithms X, Y and Z switch the user power to appropriate power levels for each user locations at times \( T_X = [0,3,6,9] \), \( T_Y = [0,2.1,5.1,8] \) and \( T_Z = [0,2.1,5.1,6.5] \) in seconds respectively. Then the level of agility for the three algorithms calculated using the metric in equation 2.4 are 55%, 75% and 86% respectively. Depending on the desired objective a threshold is set by the operator of a certain level of agility desired. Thus an acceptable level of agility is essential for an algorithm to be called SO. Finally, it is worth mentioning that distributive control is a key to enhance agility as it reduces the delays incurred due to feedback to central location, processing of feedback, decision making and propagation of the decisions back to the point of actuation.

2.4 Distinction between Self Organization and Adaptability

The three key characteristics of SO systems discussed in last section i.e. stability, scalability and agility – observed in all natural SO systems and desirable in almost all engineering systems – are distinctions of SO over classic adaptability. Here we further clarify this distinction with the following example. Consider a system has
adaptive modulation and coding algorithm that is scalable and stable but not agile may be because the feedback delay involved is longer than coherence time of channel due to high user mobility etc. This algorithm should not be called self organizing despite the fact that it is still adaptive. On the other hand if there is another adaptive coding and modulation algorithm that is stable and at the same time agile enough but not scalable may be because it requires excessive signalling overhead that increases its implementation complexity as number of users in the system increase. This adaptive algorithm is not truly self organizing either. But if same adaptive coding algorithm is modified to be implemented without excessive signalling may by using only partial or imperfect feedback, and still remains agile and stable according to definitions provided above, it becomes self organizing. The only adaptiveness complemented with scalability, stability and agility should be called SO. Minimal complexity, only local and minimal coordination and distributive control are the key players in achieving self organization.

2.5 Self Organization in Wireless Cellular Systems: A Brief Review

In order to set the necessary background of our work, this section provides only a brief overview of pioneering papers in open literature, major research projects dedicated to SO and how these works have contributed to evolving standards and activities in 3GPP. A very brief time line based style is taken here with aim to present summary of evolution of literature on SO in WCS. A detailed and comprehensive review of the state of the art on SO has been published in IEEE surveys and tutorials [3].
2.5. Self Organization in Wireless Cellular Systems: A Brief Review

2.5.1 Individual Works

Although much work has been done on SO for ad hoc, mesh and sensor networks but only a few works delved in SO in WCS. In regime of WCS, the term self organization was first used by authors in [18] in a focused context of adaptiveness in the power control in GSM. Both [18] and [19] presented a simple adaptive power control algorithm to show how this kind of self organization can mitigate the effect of BS positioning error. Soft frequency hopping was proposed as a self organising mechanism in [20] to cope with high interference in Pico cell environment in GSM. In [9] the authors provide first system wide view of self organisation in cellular networks by listing a number of adaptive mechanisms like intelligent relaying, dynamic cell sizing, BS bunching and situation awareness. Furthermore [9] also discusses the expected gains in performance, capacity and cost that can be achieved through such mechanisms. However, no framework or design strategy to enable these techniques for a system wide self organization is provided in [9].

The concept of self organisation is applied at the scheduling management level in [21]. The basic idea in [21] is to divide the cells with similar traffic requirements into clusters by using neural self organising maps and use appropriate scheduling scheme for each cluster. Significant improvement in performance using this cluster based scheduling scheme is shown compared to scenario where same scheduling scheme is used in whole network. In [21] the algorithm does not require global coordination after the clusters are formed. However, in the dynamic process of making clusters, the large amount of global coordination required, the convergence time and the complexity of clustering process itself, especially in large network, makes this algorithm devoid of the main features of SO like agility and scalability.

Authors in [22] and [4] attempt to provide a generic design strategy for design-
ing self organization into cellular networks and a general communication network respectively. Authors in [22] borrow the concepts of small world and scale free world from mesh and ad hoc networks. It is suggested that if these concepts can be applied to WCS with suitable modifications, as WCS does not hold these two properties intrinsically, SO can be achieved into WCS as well. However, no further description of the required modifications are suggested in [22], rendering the reported example of WCS effectively an example of ad hoc self organising network. In [4] authors postulate four essential paradigms that should be considered in designing process of a self organising system. While [22] and [4] open up new horizons on design of SO, in general, they just held off from demonstrating the proposed design principles through their pragmatic application by providing a self organising solution to a specific problem in cellular systems.

Preliminary idea of using relay station as self organising agent or more specifically self healing agent proposed in [22] is further extended in [11]. Authors in [11] determine the optimal number and topology of relay stations placement required for cost efficient QoS aware coverage in the area of interest. A decentralised medium access protocol is proposed in [23] and authors show that the proposed MAC protocol is capable to operate in worst case scenario of no frequency and BS station location planning hence it can pave the way towards self organizing or more precisely self configuring cellular systems essentially by eliminating the need for pre-planned deployment.

Potential challenges to implement multiple self organising functionalities in future cellular systems are identified in [24]. Authors in [24] conclude that a complex coordination plane is inevitable to enable system with multiple self organizing functionalities in order to avoid the conflicts among the objectives targeted by these functionalities. These conflicts arise because the sets of parameters influencing the major optimization objectives like optimization of coverage, capacity, spectral efficiency or energy efficiency, are not mutually exclusive. It is to be
2.5. Self Organization in Wireless Cellular Systems: A Brief Review

noted that a centralised control plane as suggested in [24] may come at cost of low scalability due to high signalling overhead thus diminishing the most desirable features of self organization.

In [25] authors present self organising algorithms and their expected gains for three different aspects of cellular network namely Handover Parameter Optimisation, Protocol Stack and Topology Self-configuration and Knowledge based Proactive Context Handling. In [26] authors propose to use self organization as tool for inter and intra operator joint radio resource management and provide expected gains based on preliminary results. A self organising power and spectrum allocation algorithm that can operate in distributed manner for femto cells based on the user reported CQIs is presented in [27]. The proposed spectrum allocation scheme in [27] is shown to ameliorate the spectral efficiency compared to a random spectrum allocation.

In [28] authors present a self organising framework for joint optimization of spectral efficiency, QoS and energy efficiency by building on performance metrics for these measures derived in [29]. The solution space for optimization problem is derived semi analytically and a bimodal utility function is proposed that enables switching to the optimal state in the solution space based on current system requirements in spectral efficiency, QoS and energy efficiency.

In summary, work on self organization in cellular networks can be broadly classified into three eras. In first era works like [18,21] brought forth the need for the self organization in cellular network and presented a number of adaptive algorithms under the umbrella of SO but did not attempt to present an explicit definition of SO and design strategy to build SO into cellular systems. In second era of SO in cellular networks works like [22] and [4] embarked on holistic approach towards designing self organization and postulated a set of germane principles and paradigms to be considered in the design process of self organising system but stood short of demonstrating these principles through a pragmatic application.
to a specific problem in the cellular systems. Third era can be marked by the works like [11, 28] that ventured on designing self organizing solutions for various problems in cellular system but in general the solutions presented in these works yield the useful level of adaptability, however, mostly at the cost of compromise on one or other feature of self organization i.e. scalability, stability or agility.

### 2.5.2 Major Research Projects

The interesting research problems highlighted in the individual works and increased understanding of dire need for self organisation in future WCSs has prompted research in a number of large research projects in recent years dedicated to investigating, analysing and evaluating the possible deployment of self organisation functionality in WCS. Here we present a summary of the aims and achievements of the major research projects on SO.

European Celtic Gandalf project [30] identified the heterogeneous nature and increased complexity of future wireless networks, and studied SO as potential solution to overcome the complexity expected from concurrent operation of 2G, 2.5G, 3G and future network solutions. They evaluated the effect of automation in network management and joint radio resource management in WLAN/UMTS networks with automated fault troubleshooting and diagnosis. The European “end-to-end efficiency (E³)” project [31] focused on integrating current and future heterogeneous wireless systems into cognitive systems for self-management, self-optimization and self-repair (described as self-X systems). Focussing on autonomous cognitive radio functionalities, feasibility tests were performed for various schemes including autonomous Radio Access Technology (RAT) selection, acquiring and learning user information, self configuration protocols and awareness signalling among the self organising network nodes. The tests validated that the proposed autonomous schemes were more efficient and feasible solutions for
2.5. Self Organization in Wireless Cellular Systems: A Brief Review

Another major project dedicated to SO is the European SOCRATES project [32] which has identified main objectives for which self organization is required in WCS. These objectives are classified in the form of set of use cases. It is further concluded that most of the objectives, SO has to deal with in WCS, are not independent from each other as the parameters controlling them are not mutually exclusive making joint optimization difficult if not impossible. They have also gone in great details to describe framework for major tasks involved in self organised wireless networks which include self-configuration, self-optimization and self-healing functionalities in future radio access networks.

2.5.3 Standardization Activity on Self Organization

With the requirements of operators identified and activities from various research projects validated and ready for deployment, it was necessary to have a common standard for a unified performance evaluation, compatible interfaces and common operating procedures among all vendors and operators.

A consortium of major mobile communication vendors came together under the working group of Next Generation Mobile Network (NGMN) and agreed that self organizing functionality was a solution towards improving operational efficiency of future networks [8]. Future network architectures are flat and simplified in order to reduce the complexity. Operators also would have to reduce the operational effort in order to minimise the cost involved in deploying such networks. The need for automation in some of the operation and maintenance (O&M) modules in self configuring and self optimising would lead to systematic organised behaviour that would improve network performance, be more energy efficient and pose a good business case for operators. The 3rd Generation Partnership Project (3GPP) have set out working groups aimed at coming up with specification for
Figure 2.3: An illustration of the possible approaches towards design of SO [3]

agreed descriptions of use cases and solutions with emphasis on the interaction of self optimization, self configuration and self healing. The main documents describe concepts and requirements for self organizing networks [33] as well as specific documents on self establishment of eNodeBs [34], Automatic neighbour management [35] self optimization [36] and self healing [7] are also produced and released.

2.6 Designing Self Organization in Wireless Cellular Systems

Since the conception of the idea of SO and subsequent realisation of its benefits in WCS a number of divers approaches starting from artificial intelligence based learning techniques [37-39] and evolutionary heuristics [40,41] to multi objective optimization [28] have been taken in literature to design SO. A detailed account of the possible design approaches towards SO is presented in [3] and a tree diagram illustrating the taxonomy of key approaches is reproduced in figure 2.3 for quick
2.6. Designing Self Organization in Wireless Cellular Systems

While these approaches are very useful to find solutions to specific problems, a generic framework to design SO is missing in the literature. A most relevant attempt has been made by authors [4] who identify four paradigms for design of SO. As already been discussed above in section 2.5.1 this work provides very generic directions and does not provide an actually pragmatic framework for design of SO. A most natural approach to devise such framework would be to mimic the design of SO in natural systems i.e. through Biomimetics. In the following subsections we briefly explain the main concepts of Biomimetics and exploit these concepts to devise a generic framework for designing SO in next section.

2.6.1 Biomimetics: A Bio Inspired Design Approach

Nature contains plethora of enormously complex systems that are absolutely perfect in their design and operation as highlighted in the previous sections. Biomimetic is a recently evolved branch of science that investigates such natural systems with aim to exploit their working principles for improvement in design and operation of man made systems.

In nature, there are myriad of examples of self-organizing behavior, for example, ants colonies finding shortest routes to food sources, termites collectively building complex constructions without using a blueprint, fish shoals organizing themselves without a leader, and swarms of fireflies in south-east Asia synchronously emitting light flashes. The fact that SO is originally a bio inspired phenomenon and the abundance of perfect SO in biological systems makes Biomimetics a perfect paradigm for investigating the constituents and working principles of SO with aim to design SO in engineering systems through these inspirations. There are two different approaches in Biomimetics: the direct and the indirect approach. We use both of them in this thesis and therefore explain them briefly below.
2.6.1.1 Direct Approach of Biomimetics

In the direct (or top-down) approach, an engineering problem is tackled by looking for natural systems solving an equivalent problem. The biological solution and its principles are then analyzed and re-built in a technical application. Examples of the direct approach are the design of aeroplane wings by observing the gliding flight of birds as it was done by Otto Lilienthal in the 19th century, or, after a closer analysis of the up-bent feathers at the wing tips of several birds, the refinement of aeroplane wings by turbulence-reducing and thus fuel-saving winglets [42].

2.6.1.2 Indirect Approach of Biomimetics

In contrast, the indirect (or bottom-up) approach of bio-inspired design involves first the derivation of principles by analyzing natural systems. This step is done in a basic research effort that is not yet targeted at a specific application. The principle is then abstracted from its biological context and used in particular technical applications where they could be suitable. Examples of the indirect approach are the concept of artificial neural networks or the concept of ant foraging behavior being applied to mesh network packet routing.

2.6.2 Considerations in Use of Biomimetics

Although, biomimetics is a rich and promising paradigm for design of SO in particular, however, there are some consideration and implications that must be apprehended while applying biomimetics, particularly the direct approach of Biomimetics to engineering systems. These limitations mainly arise due to the notable differences between biological and engineering systems: These differences can be categorized in three main types.
Firstly, in many biological systems, especially for lower animals, there is no real counterpart to what we call softwares in engineering systems i.e. a software update is not possible. For example, protozoa have no mechanism to learn and circulate a new behavior during lifetime only. Such updates take place over generations through genetical changes only. That is much slower process than software update in engineering systems.

Secondly, while the design of an engineering system is only limited by technical and financial constraints, all living beings on other hand follow more or less a general blueprint instead of having no obvious design constraints, particularly in the case of a given class such as mammals, the design is even more restricted. Therefore, it is not possible to achieve an intended specific and sophisticated system design merely by some evolutionary algorithm. For example, evolution have never been able to design a wheeled vehicle-like animal or other rotating machines.

Thirdly and most importantly, some functionalities of biological systems are hard to mimic and it is not technically feasible or even possible to achieve such functionality in engineering system may be because of lack of their understanding or because of technological limitations. Examples of such immimicable functionalities so far are the intelligence of human brain, or 100% efficiency of chemical energy to electromagnetic energy conversion process i.e. light in fire flies.

Thus, Biomimetics instead of all its potential might not be an appropriate approach when: (i) the biological solution is too difficult to rebuild with technical means or (ii) a technical solution can be more efficient by taking advantage of mechanisms that cannot be found in the biological paradigm. A mistake to be avoided is to stick to biological solutions just because of their seeming elegance.
Chapter 2. Self Organization from Conception to Realization

2.7 A Biomimetic Self Organization Framework (BSOF)

Myriad of systems in nature exhibit perfect self organization e.g. school of shoaling fish, swarming insects, herding sheep, synchronously flashing fire flies, and flocking birds to name a few. This provides us with an opportunity to devise and extract the generic principles of SO from nature using the indirect biomimetics as explained in section 2.6.1.2. Here we take one specific case study to investigate the underlying principles of self organization in such system with aim to come up with a generic design and operational framework for self organization.

2.7.1 Revisiting SO in Nature: Flock of Common Cranes

As explained in section 2.2.3.1 flock of common cranes is one of the myriad of perfect examples of SO system in nature. We reconsider the case study of flock of common cranes to further delve into their SO group flight phenomenon as shown in figure 2.4. It has been investigated that each birds individually executes and maintains a certain set of simple flight attributes such that the flock is always in near V-formation that happens to be the optimal formation for group flight efficiency [43]. Somehow, for common cranes, nature has solved the complex problem of group flight efficiency optimization to a much simpler problem of maintaining a V-formation. Nurture on the other hand has taught them, how to control their own flight attributes with reference to their immediate i.e. line of sight neighbours to maintain the V-formation while flying. This attributes have been identified to be barely three i.e. cohesion, separation and alignment [44]

1. Separation: Each bird tries to avoid crowding neighbors

1Nurture here means the process of bringing up, training given by parents/peers or learned independently etc
2.7. A Biomimetic Self Organization Framework (BSOF)

A flock of common cranes optimises its flight efficiency by nearly 70% through self organization i.e. by maintaining near V-formation through simple individual actions of cohesion, separation and alignment executed by each bird.

2. Alignment: Each bird tries to steer towards average heading of neighbors.

3. Cohesion: Each bird tries to steer towards average position of neighbors.

Figure 2.4 and 2.5 explains how cohesion, separation and alignment executed by each bird during flight results into an overall V-formation and hence flight efficiency optimization.
2.7.2 Constituents of Self Organization

By inspection of the above example, three main components of SO can be identified:

1. A specific objective e.g. maximization of flight efficiency.

2. A goal which is effectively a much simpler manifestation of the same objective e.g. formation of V shape.

3. A small set of simple functions which are performed by the entities of the system (mostly independently or semi independently) to achieve that goal e.g. separation, alignment and cohesion in this case.

In order to put our future discussion in a consistent context we will refer to these three identified components of a SO as SO-Objective, SO-Goal and SO-Function respectively.
2.7. A Biomimetic Self Organization Framework (BSOF)

2.7.2.1 Relationship Between SO-Objective and SO-Goal

A pivotal observation to be made here is that in a system with SO, the complex
SO-Objective (e.g. the maximisation of flight efficiency in this case study) is not
dealt with as it is by each bird, rather it is first mapped to a much simpler goal
having equivalent semantics (i.e. flying in V shape), as it is optimal formation
to minimize aerodynamic drag and hence maximise group flight efficiency [15].
Thus a crucial step to achieve SO is translating the complex SO-Objective into
a simpler SO-Goal. In this case study, nature has done this job. Next equally
important step in design of SO is design of such simple SO-Functions that can
achieve the SO-Goal(s).

2.7.2.2 Designing SO-Functions

Simplicity of SO-Functions is one of the basic properties for achieving SO and
its associated benefits. Other than that following general characteristics of SO-
Functions are worth noting in all natural systems with SO:

1. SO-Functions do not require any discriminative abilities e.g. each bird in
   the flock can perform all the three SO-Functions i.e. separation, alignment
   and cohesion.

2. SO-Functions do not require explicit or global co-ordination among entities
   of system e.g. in bird flock example separation, alignment and cohesion are
   performed by each bird by relying on its local observation only.

3. SO-Functions are not based on pan system actions.

So far, we have explicated three basic constituents of SO i.e. SO-Objective, SO-
Goal, and SO-Functions and have figured out the characteristics of each of them.
Next, we need to determine the key operational principle of SO i.e. the nature
of interaction of entities of the system that ensures achievement of a common SO-Goal. The full understanding of this operational principle is crucial in designing appropriate SO-Functions in a real world engineering system. Following subsection investigates the type of mutual behavior among entities of the natural system with SO.

2.7.2.3 Coopetition: An Interesting Principle Behind SO-Functions

A probe into the mutual behavior among the birds, in the example of bird flock, deciphers that SO-Functions executed by each bird are neither purely cooperative towards other birds nor purely competitive. A pure cooperative behavior among birds will require explicit communication which is not the case in the bird flock. On the other hand, a pure competitive behavior among birds will result in a conflict of interests and will not result in common SO-Goal e.g. instead of single well maintained V-shape multiple independent subgroups with different shape can result or all birds might flight in straight line in parallel to each other as result of pure competition. This implies that a system may not evolve to SO if the individual entities of the system behave either in pure cooperative fashion or pure competitive fashion.

A cautious inquisition of the above case study shows that the three SO-Functions enacted by each bird are a well composed combination of both cooperative and competitive behavior towards other birds. In fact SO-Function separation exhibits a kind of cooperative behavior towards other birds as by executing it each bird is being friendly to other by giving it more space. The SO-Function Alignment has dominantly competitive nature as by executing it each bird is trying to reach the same destination as its peers in a rush to avail the resources (food etc) available at that destination. Whereas, SO-Function cohesion is an intricate amalgam of both cooperation and competition as by executing it birds tend to push into their neighbor birds but simultaneously maintaining a threshold sepa-
2.7. A Biomimetic Self Organization Framework (BSOF)

In essence, SO-Functions executed by the entities of a SO system are such that their mutual behavior is neither pure competition nor cooperation rather it is a judicious combination of both of these extreme attitudes. The most suitable term to embody this type of behavior is coopetition: a neologism introduced in the field of economics to represent cooperative competition [45]. Hence, it is neither competition nor cooperation but coopetition among entities in the system which can result into SO behavior on system level.

2.7.3 A Generic Biomimetic Self Organization Framework (BSOF)

Building on all these findings through analysis of SO in natural systems, we can infer a general framework to design SO in a system which is based on three steps:
1. Identification of SO-Objective

2. Mapping of SO-Objective into simple SO-Goal

3. Identification of functionalities bearing the characteristics of SO-Functions and executable under the principle of coopetition to achieve the SO-Goal.

These three steps are illustrated in figure 2.6. In summary, in order to design SO, we need to do the job of both nature and nurture. That is, transformation of complex objective into much simpler probably still global but locally achievable goal has to be accomplished. This should be followed by the design of such actions of individual entities of the system that do not require global coordination but when executed, collectively can achieve the goal. By following these steps, we can design self-organising solution for that particular objective. We call this generic framework BSOF i.e. Biomimetic Self Organization Framework. It is important to mention here that, for transformation of SO-Objective to SO-Goal, SO-Goal to SO-Function or, from SO-Objective to directly SO-Function; any of suitable approaches discussed in beginning of section 2.6 and illustrated in figure 2.3 can be used. In this thesis we mainly use optimization approaches highlighted in figure 2.3 with shaded boxes.

2.8 Conclusions

In this chapter we presented brief overview of existing notions of self-organization (SO) that are most pertinent to the context of wireless cellular systems. We identified the need for more concise definition of SO and provided a definition that characterises SO on basis of what SO can achieve, unlike existing notions that characterise SO on basis of how SO can achieve. The necessary conditions for adaptiveness to be called SO were also iden-
tified and explained in detail through examples. Based on the established notion of SO, a brief review of the state of the art was presented to build foreground of our work.

Having established the definitions of SO and the relevant background, we focused on the design of SO and identified that Biomimetics is the most rich paradigm for design of SO. We explained the two main approaches of Biomimetics and also highlighted its implications that need to be apprehended when using biomimetics. With this understanding of potential and limitation of Biomimetics, we devise a generic framework for design of self organization called BSOF through extensive case study from nature.

In next chapter we apply BSOF with the help of analytical approaches also highlighted in this chapter to have most viable solutions to the problems we address in this thesis. More specifically, in subsequent chapter BSOF is applied to a set of problems and the intermediate steps of the BSOF are solved with the conventional analytical tools suitable to the nature of problem.
Chapter 3

Spectral Efficiency Optimization through SO of Antenna Tilts

3.1 Introduction

Spatio-temporally unpredictable traffic hotspots are a persistent problem in WCS that are a frequent source of low user satisfaction. In this chapter we present a novel SO solution to this problem by applying BSOF. This solution achieves a holistic objective of throughput enhancement through system-wide BS antenna tilt optimization by relying on local coordination only. The fairness among users is also considered in the design of this solution. The potential gain of the solution is demonstrated for two different scenarios of pop up hotspots based user distributions. This solution is agile enough to cope with short term dynamics identified in section 1.1 and thus can be used for medium and long term dynamics as well. The complex system-wide objective (i.e. SO-Objective) of throughput maximisation (over optimization variables of system-wide antenna tilts) is transformed analytically into much simpler optimization goal (i.e. SO-Goal). This simplification is achieved by designing symmetry into the the system model and exploit-
Chapter 3. Spectral Efficiency Optimization through SO of Antenna Tilting

ing this symmetry in problem formulation. The SO-Goal is then further decomposed into locally executable simple tilt optimization functions (i.e. SO-Functions). These local optimization functions are then solved through sequential quadratic programming. The solution obtained is implementable through a fully scalable algorithm. We name this solution TO-BSOF i.e. Tilt Optimization through BSOF. The performance gain of the TO-BSOF is evaluated through system level simulations for OFDMA based general WCS like LTE. Numerical results show that, compared to off-line fixed tilting options, over 30% gain in average user throughput can be achieved through TO-BSOF without resorting to global coordination or central control.

This chapter is organized as follows. Section 3.2 presents the necessary background and brief review of relevant works that have focused on tilt optimization or adaptation. The system model and assumptions used for problem formulation are presented in section 3.3. In section 3.4 we apply BSOF to our problem of traffic hotspots in WCS to develop a SO solution. In section 3.5 we evaluate the developed solution through system level simulations for a general OFDMA based WCS. In section 3.6 the implementation aspects of the developed solution are discussed to establish its practicality. Finally, in section 3.7 we present key conclusions of the contributions presented in this chapter.

3.2 Background and Related Work

User distribution in real world live WCS is not uniform in contrast to a uniform user distribution often considered during WCS planning and design phase as well as in simulation based evaluations. In real world, high user densities intermittently occur near points of attractions e.g. theaters, shopping malls, event venues, beaches and office complexes etc. This re-
3.2. Background and Related Work

...suits in pop up traffic hotspots at random times and at random locations in the coverage area. These traffic hotspots become a source of low user satisfaction due to three major problem they cause:

Firstly, when hotspots occur at a location the radio resources available at BS serving that location become insufficient and hence call blocking increases in that cell. This type of blocking will be called hard blocking to differentiate it from soft blocking. Soft blocking occurs when sufficient radio channels are available to accommodate the call request by a user, but interference on these channels is too high. In this case that call can not be initiated and sustained with minimum QoS requirement and hence the user call request is not accepted i.e. blocked.\(^1\)

Secondly, uneven traffic distribution resulted from pop up hotspots at some locations ultimately causes under utilization of spectrum at other locations in the system, and hence system throughput is degraded.

Thirdly, hotspot can cause severe performance degradation, if they occur in areas where interference or shadowing is high. In such scenarios only low order modulation and coding schemes can be used by the system to provide service to the users in these hot spots. Therefore, such hotspots cause a significant proportion of user population to be served with low spectral efficiency. This in turn deteriorates the QoS and system wide throughput in WCS.

The first and second problems caused by uneven traffic distribution will be dealt in detail in chapter 4. In this chapter our focus remains on resolving mainly the third problem caused by hotspots.

The problems caused by the pop up hotspots in general are hard to be designed out during the WCS deployment phase because of their unpre-

\(^1\)Hard blocking caused by uneven traffic distribution will be dealt in next chapter where both terms i.e. hard blocking and soft blocking, will be used frequently.
dictability both in time and space. An over safe design to accommodate potential pop up hotspots is not feasible option either, as it can be expensive and may result in under utilization of resources itself. To this end, in this chapter we provide a SO solution to improve the average spectral efficiency and hence throughput in the WCS in face of hot spots. The basic idea is to enhance the Signal to Interference Ration (SIR) dynamically at locations of pop up hotspots, through optimization of system wide antenna tilts, but without sacrificing average user throughput in the system.

There have been several studies on tilt optimization\textsuperscript{2} most of which focused on improving the coverage and capacity in general for GSM [46], CDMA [47-50], HSDPA [51] and LTE [52-56] based WCS. The approaches taken in all these works can be divided in two main categories.

In first category fall the schemes that use antenna tilt basically as an interference reduction tool. Such schemes aim to adapt antenna tilt to minimise interference and thus improve capacity. These type of schemes are mostly investigated in context of the CDMA based WCS to control inter cell interference [46,48-50]. However, works in [52,54,55] specifically investigate impact of antenna tilt on interference and capacity in 3GPP LTE based WCS as well. Although both, the works in [46,48-50] and [52,54,55] use the same basic approach of tilt adaptation to reduce interference, but the difference between them lies on their modeling of the interference. This is because interference generation phenomenon in CDMA based WCS is fundamentally different than that in OFDMA based WCS. In CDMA based WCS a full frequency reuse is always used. Furthermore, in CDMA based WCS all users in the system transmit on the whole of the spectrum. Whereas in OFDM/OFDMA based system, bandwidth is partitioned into sub carriers

\textsuperscript{2}A detailed survey of works on tilt optimization has been published in [3], here only a brief overview and discussion on most relevant works is being presented.
and within a cell, subcarrier are not reused. Therefore unlike CDMA based WCS, in OFDMA based WCS intra cell interference does not exist in general. However, in OFDMA based WCS subcarrier are reused in the other cells giving rise to inter cell interference, but on the reused subcarrier only. Therefore, inter-cell interference in OFDMA bases WCS is also different as it is not spread over whole bandwidth, unlike in CDMA based WCS.

Second category consist of schemes that exploit antenna tilt as a tool to control effective coverage and thus control the load in the cell to achieve load balancing either in the CDMA based WCS [47,51,57] or OFDMA based WCS [53].

The most relevant to the scope of this chapter are works presented in [47] and [57]. Both [47] and [57] deal with hotspot through antenna tilting but for CDMA and HSDPA respectively unlike our work where we address this problem for OFDMA based WCS. In addition to the differences discussed above, unlike the soft handover in CDMA based WCS, in OFDMA based WCS the handover is hard handover i.e. this may involve a change of carrier frequency. This change of carrier frequency increases the complexity and overheads associated with handovers compared to the soft handovers in CDMA based WCS. Since pop hotspots are short term dynamics (see section 1.4), therefore, the compensating mechanisms for them, that require excessive handover among cells do not remain an attractive solution for OFDMA based WCS. The tilting mechanism proposed in both [47] and [57] are based on the basic idea of tilting down the overloaded cell antenna to reduce its effective coverage area in order to shift its load to neighbouring cells. Such tilting mechanisms are shown to rely on excessive handovers from overloaded cell to adjacent cells [47]. As discussed above, such handovers

\[ ^{3}\text{Load balancing is the scope of our next chapter and these works will also be discussed in that chapter.} \]
The approach we take in this chapter, is novel and fundamentally different from the two approaches explained above and used in the works so far. Although we use antenna tilts to deal with hotspot as used in [47,57] but unlike both of these works that focus on the second of the three main problems caused by hotspots, as discussed above, our work focuses on the third problem caused by hotspots. In other words, we do not seek hotspot relief through load balancing achieved by antenna down tilting of over loaded sectors and thus triggering handovers from that cells to neighbour cells. Neither we present a scheme to tilt down antennas for interference minimization in the system in general. Rather we introduce a novel concept of traffic's Center of Gravity (CG) to represent the user geographical distribution in a cell, wether uniform or hotspot based. By building on this concept we then develop a unique scalable, stable and agile mechanism where a pre determined set of neighbouring cells jointly optimize their tilts to focus their antenna gains at the CG's e.g. hot spots in those cells. This antenna tilt optimization is performed in distributed manner but on system wide scale. In addition to the cells containing hotspots, this system wide optimization process also includes the cells that do not have clear hotspots or have uniform user distribution, by determining their CG. Thus this mechanism can improve the overall spectral efficiency of WCS in face of a realistic non homogenous dynamic traffic distribution. Unlike the hot spot relief approaches used in [47,57] our approach does not achieve hot spot relief by transferring load to nearby under loaded cells, rather it relieves congestion by enhancing the spectral efficiency at hotspots, therefore, it does not necessitates handovers. The intrinsic SO nature of this solution enabled by

overs can be manageable softly in CDMA, but in OFDMA based networks like LTE and LTE-A they are hard handovers and undermine the practicality of this approach.
3.3. System Model and Assumptions

Assumptions and Nomenclature:

The analysis and the simulation results in this chapter will be focused on the down link of WCS for sake of clarity and conciseness. However, most of the conclusions drawn are general and remain equally valid for uplink as well.

By throughput we means bandwidth normalised throughput which is analogous to spectral efficiency.

It is assumed that all users devices have omnidirectional antennas with $0dB$ gain.

An interference limited scenario is assumed i.e. noise is negligible compared to interference.

The term sector is used in the same meanings as cell.

The subscripts always denote association receiving user and post scripts denote association with the transmitting node until unless specified otherwise.

System Model:

We consider a sectorized multi cellular network with each base station having three sectors as shown in figure 3.1. Let $N$ denote the set of points corresponding to the transmission antenna location of all sectors and $K$ denote the set of points representing location of all users in the system. The geometric Signal to Interference Ratio i.e. SIR perceived by a user at a
Figure 3.1: System model for problem formulation. Small (red) circles show location of hot spots in some sectors while others have uniform user distribution.

location $k$ being served by $n^{th}$ sector can be given as

$$\gamma_k^n = \frac{P^n G^n_k \alpha (d^n_k)^{-\beta}}{\sum_{m \in N \setminus n} (P^m G^m_k \alpha (d^m_k)^{-\beta})} \quad m, n \in N, k \in K$$  

(3.1)

where $P^n$ and $P^m$ is transmission power of the sector located at points $n$ and $m$ respectively, $d^n_k$ is distance between the transmitting sector located at point $n$ and receiving user with located at point $k$. $\alpha$ and $\beta$ are pathloss model coefficient and exponents respectively. The notation "\" means set subtraction i.e. $\forall m \in N \setminus n$ means all elements in $N$ except $n$. $G^n_k$ and $G^m_k$ are antenna gains from sectors located at points $n$ and $m$ to user located at point $k$. For 3GPP LTE and LTE-A it can be modelled as proposed in [53]
and can be written in dB as:

\[
G_k^n = \lambda_v \left( G_{max} - \min \left( 12 \left( \frac{\theta_v^n - \theta_{tilt}^n}{B_v} \right)^2, A_{max} \right) \right) + \\
\lambda_h \left( G_{max} - \min \left( 12 \left( \frac{\phi_h^n - \phi_{tilt}^n}{B_h} \right)^2, A_{max} \right) \right), \forall n \in N \tag{3.2}
\]

where \( \theta_v^n \) is the vertical angle in degrees from \( k^{th} \) location of user to \( n^{th} \) sector \( \theta_{tilt}^n \) is the tilt angle of the \( n^{th} \) sector\(^4\) as shown in figure 3.1. The \( \phi^n_h \) is horizontal angle in degrees with similar meanings of subscript and postscript. Subscripts \( h, a \) and \( v \) denote horizontal, azimuth and vertical respectively. Thus \( B_h \) and \( B_v \) represents horizontal and vertical beamwidths of the antenna respectively, and \( \lambda_h \) and \( \lambda_v \) represent weighting factors for the horizontal and vertical beam pattern of the antenna in 3D antenna model [53], respectively. \( G_{max} \) and \( A_{max} \) denote the maximum antenna gain at the boresight of the antenna and maximum antenna attenuation at the sides and back of the boresight of the antenna respectively, in dB.

It is important to note that \( G_{max} \) and \( A_{max} \) are same for horizontal and vertical radiation pattern (which is usually the case with commercial WCS antennas), therefore no subscript \( v \) and \( h \) are associated to them. In order to substitute in (3.1), the antenna gain in (3.2) can be written in linear form as follows:

\[
G_k^n = 10^{0.1 \left( \lambda_v \left( G_{max} - \min \left( 12 \left( \frac{\theta_v^n - \theta_{tilt}^n}{B_v} \right)^2, A_{max} \right) \right) + \lambda_h \left( G_{max} - \min \left( 12 \left( \frac{\phi_h^n - \phi_{tilt}^n}{B_h} \right)^2, A_{max} \right) \right) \right)} \tag{3.3}
\]

For sake of mathematical tractability we can neglect the maximum attenuation factor \( A_{max} \) in (3.3) i.e. we assume infinite attenuation. Without this finite level clamping factor \( A_{max} \), antennas model will have continuously increasing attenuation. This attenuation will increase as function of

\(^4\)By \( n^{th} \) sector or \( k^{th} \) user we will mean the sectors or users located at points \( n \) and \( k \) respectively
angular distance from their boresight i.e. \((\phi_k^n - \phi_a^n)\) or \((\theta_k^n - \theta_a^n)\), instead of flat attenuation \(A_{\text{max}}\) after certain angular distance from boresight. For example the horizontal angular distance i.e. \((\phi_k^n - \phi_a^n)\) beyond which the flat attenuation occurs in (3.2) can be seen to be 90° in case of a three sector antenna with 70° beam width. Thus the clamping factor plays a role in only determining antenna attenuation for users very far away from boresight of the antenna. Therefore, in sectorized deployment with hexagonal topology the effect of this simplification is even more negligible. Furthermore, this assumption actually brings the antenna model in (3.2) closer to real antennas as real antennas do not have flat attenuation [58].

Without loss of generality we can assume maximum gain \((G_{\text{max}})\) of 0 dB. Thus by neglecting clamping effect i.e. putting \(A_{\text{max}} = \infty\) and \(G_{\text{max}} = 0\) dB in (3.3), it can be simplified as:

\[
G_k^n = 10^{-1.2 \left( \lambda_u \left( \frac{\phi_k^n - \phi_a^n}{B_u} \right)^2 + \lambda_h \left( \frac{\theta_k^n - \theta_a^n}{B_h} \right)^2 \right)}
\]  

(3.4)

We assume that all the base stations transmit with same power.\(^6\) For such scenario, by using (3.4) in (3.1) the SIR at location of \(k^{th}\) user can be determined as:

\[
\tau_k^n = \frac{\alpha(d_k^n)^{-\beta} 10^{-1.2 \left( \lambda_u \left( \frac{\phi_k^n - \phi_a^n}{B_u} \right)^2 + \lambda_h \left( \frac{\theta_k^n - \theta_a^n}{B_h} \right)^2 \right)}}{\sum_{m \in N \setminus n} \left( \alpha(d_m^n)^{-\beta} 10^{-1.2 \left( \lambda_u \left( \frac{\phi_m^n - \phi_a^n}{B_u} \right)^2 + \lambda_h \left( \frac{\theta_m^n - \theta_a^n}{B_h} \right)^2 \right)} \right)}
\]  

(3.5)

For the sake of simplicity of expression we use following substitutions:

\[
c_k = \frac{B_u^2 \lambda_h}{\lambda_u} \left( \frac{\phi_k^n - \phi_a^n}{B_h} \right)^2
\]  

(3.6)

\(^{\text{6}}\)Both of these assumptions i.e. \(A_{\text{max}} = \infty\) and \(G_{\text{max}} = 0\) dB make the problem analytically traceable. However, in system level simulation results, these assumptions are not used.

\(^{\text{6}}\)This assumption is inline with OFDMA based WCS e.g. LTE where no power control is applied on down link as adaptive modulation and coding scheme is used instead.
3.4 Hot Spot Traffic Relief by Tilt Optimization through BSOF(TO-BSOF)

\[ h_k^m = \alpha (d_k^m)^{-\gamma} \]  \hspace{1cm} (3.7)
\[ \mu = \frac{-1.2 \lambda_w}{B^2_g} \]  \hspace{1cm} (3.8)

Using the substitutions in (3.6)-(3.8), the SIR in (3.5) can be written as

\[ \gamma_k^n = \frac{h_k^n 10^\mu((\theta_k^n - \theta_{\text{tilt}})^2 + \alpha)}{\sum_{\gamma_m \in \mathcal{N} \setminus \mathcal{N}_s} (h_k^n 10^\mu((\theta_k^n - \theta_{\text{tilt}})^2 + \alpha))} \]  \hspace{1cm} (3.9)

Note that it can be seen that \( \gamma_k^n \) is function of vector of tilt angles of all sectors i.e. \( \theta_{\text{tilt}}^N = [\theta_{\text{tilt}}^1, \theta_{\text{tilt}}^2, \theta_{\text{tilt}}^3 ... \theta_{\text{tilt}}^N] \) where \( N = |\mathcal{N}| \), but for sake of simplicity of expression we will show this dependency only where necessary.

We consider scenario with traffic hot spots that makes the user distribution non uniform. Let \( S \) denote the location points of these randomly located hotspots in the system (represented as small circles in figure 3.1) and \( \mathcal{N}_S \) denote the locations of sectors that contain these hotspots such that \( N \subseteq \mathcal{N}_S \). Given the small sector size we safely assume that a sector at most can have one hot spot within it at random location i.e. \( |\mathcal{N}_S| \leq |\mathcal{N}| \). This hotspot can contain a significant percentage of total users in that sector as shown in figure 3.1.

3.4 Hot Spot Traffic Relief by Tilt Optimization through BSOF(TO-BSOF)

In this section we follow the steps of BSOF to find the solution for hot spot traffic relief. First the problem is formulated using the system model i.e. SO-Objective is identified. In next subsections the SO-Objective is transformed into SO-Goal. In following section SO-Functions are determined and a solution is derived using sequential quadratic programming to enable execution of these SO-Functions.
3.4.1 Problem Formulation: Identifying SO-Objective

Fortunately, for this particular problem we are able to use direct biomimetic approach, i.e. one to one mapping to the case study presented section 2.7.1 and our problem can be established. To this end, cellular system can be considered analogous to flock of common cranes with BS being analogous to common cranes, operation of system being analogous to flight of flock, spectral efficiency of cellular system being analogous to flight efficiency of flock, and interference being analogous to air drag. Using this analogy of case study presented section 2.7.1 our problem can be described as follows:

A system(flock) of BS(common cranes) as shown in figure 3.1 wants to control its tilts(flight attributes) and operate(fly) such that the average spectral(flight) efficiency on the hot spots in the whole system (flock) is maximised by minimising the average interference(air drag) each hot spot location receives (bird faces). Furthermore, it should be done:

- without a central control entity (no leader BS) i.e. scalability,
- without global information exchange or coordination among the all BS (birds) of the system (flock) i.e. agility
- without sacrificing the average user throughput of users in sectors without hot spots (without swaying from the long term direction of flight) i.e. stability

Thus, in terms of BSOF our SO-Objective is to maximize the throughput \( \eta \) at all the hot spots altogether such that the average throughput of all the users does not fall below that prior to the optimization. Mathematically:

\[
\max \eta (\theta_{\text{tilt}}) = \max_{\theta_{\text{tilt}}} \sum_{\gamma_{\text{hit}}} \log_2 (1 + \gamma_{\text{hit}} (\theta_{\text{tilt}}))
\]  

subject to:

\[
\pi (\gamma_{\text{hit}}) \geq \pi (\gamma_{\text{hit}})
\]
\( \pi \) is function that returns the average per user throughput at the locations of those users who do not belong to sectors containing containing hot spots i.e sectors with uniform user distribution. It can be written as:

\[
\pi(z) = \frac{1}{|\mathcal{W} \setminus \mathcal{W}_s|} \sum_{\nu \in \mathcal{W} \setminus \mathcal{W}_s} \left( \frac{1}{|\mathcal{K}^n|} \sum_{\nu \in \mathcal{K}^n} \log_2(1 + z) \right)
\]  

(3.12)

using (3.12) in (3.11) the condition can be written as:

\[
\frac{1}{|\mathcal{W} \setminus \mathcal{W}_s|} \sum_{\nu \in \mathcal{W} \setminus \mathcal{W}_s} \left( \frac{1}{|\mathcal{K}^n|} \sum_{\nu \in \mathcal{K}^n} \log_2(1 + \tilde{\gamma}_{k}^n) \right)
\]

\[\geq \frac{1}{|\mathcal{W} \setminus \mathcal{W}_s|} \sum_{\nu \in \mathcal{W} \setminus \mathcal{W}_s} \left( \frac{1}{|\mathcal{K}^n|} \sum_{\nu \in \mathcal{K}^n} \log_2(1 + \gamma_k^n) \right)
\]  

(3.13)

and \( \tilde{\gamma}_{k}^n \) is post tilting SIR, where i.e. SIR obtained after the sector antennas are tilted at optimal angles obtained from (3.10) and is given as:

\[
\tilde{\gamma}_{k}^n = \frac{h_k^n 10^{\alpha_0((\theta_k^n - \theta_m^n)^2 + \varphi)}}{\sum_{\nu \in \mathcal{N} \setminus \mathcal{W}} \left( h_k^n 10^{\alpha_0((\theta_k^n - \theta_m^n)^2 + \varphi)} \right)}
\]

(3.14)

where \( \gamma_k^n \) is the geometric SIR at the center of hot spot such that:

\[
\gamma_k^n = \frac{1}{|\mathcal{K}_s^n|} \sum_{\nu \in \mathcal{K}_s^n} \tilde{\gamma}_{k}^n
\]

(3.15)

where \( \mathcal{K}^n \) is set of location of users in \( n \)th sector, \( \mathcal{K}_s^n \) set of location of users in the hot spot present at location \( s \) in sector represented by \( n \) where \( \mathcal{K}_s^n \subseteq \mathcal{K}^n \). It is to be noted that the condition in (3.13) ensures that system wide tilting is such that it does not reduce the average user throughput in the sectors with no hot spot while improving throughput of users that are located in the hot spot.

### 3.4.2 Transforming SO-Objective into SO-Goals

The next step of BSOF is to transform the problem in (3.10)–(3.11) into SO-Goal that is a simpler manifestation of this complicated objective. The
main reason for the complexity of the SO-Objective is the constraint in (3.13). This constraint arises from the lack of symmetry in the problem scenario. Since in real scenario not all the sectors will have hot spot as they might have uniform traffic distribution hence \(|\mathcal{S}_s| \leq |\mathcal{S}|\), so we need this constraint in (3.11) to ensure a system wide throughput optimization instead of optimization for only those sectors that have hot spots. We present following proposition to dissolve this complexity in order to transform the SO-Objective into a suitable SO-Goal as first step of BSOF for this problem.

Before further analysis we define a metric \( \rho^n = \frac{A^n_s|s^n|}{A^n} \) that characterises user density i.e. user geographical distribution in sector \( n \), where \( A^n_s \) is area of hot spot with center at location \( s \) in sector \( n \) and \( A^n \) is whole area of the same sector.

Figure 3.2 plots \( \rho \) as function of 'area of hot spot as percentage of total sector area' and 'percentage of total users confined in that hot spot'. The region separated out by arrows (right bottom side of the graph) shows the scenarios with user geographical distributions have sever hot spots i.e. a large fraction of total user population is concentrated in the hot spots. The two cases indicated by markers inside this region, indicate two different arbitrarily chosen instances of hotspot based user geographical distribution. These cases are used for evaluation of performance of TO-BSOF. Case 1 represent a scenario, where 50% of users are concentrated within the hot spots that have area 10% of the total cell area. Second case shows a scenario where hotspots are even more sever as 80% of user population is concentrated within randomly located hot spots of same area.

We are interested in a natural kind of user distribution where some sectors have uniform user distribution and others have hot spots in them as shown in figure 3.1. For such scenario \( 0 \leq \rho^n \leq 1 \) where \( \rho^n = 1 \) characterises a
Figure 3.2: $\rho$ is metric defined to characterise user geographical distribution. It is plotted as function of hot spot radius and percentage of total users in the sector within that radius. Mesh region represents scenario where $0 < \rho < 1$. 

Region where $\rho << 1$

i.e. severe hotspots

Area of the hot spot as percentage of sector area

Users in hot spot as percentage of total users in sector
Figure 3.3: Illustration of SO-Goal. Stars show Center of Gravities (CG's) of sectors that do not have clear hot spots. CG are determined through proposition 1 or 2. Circles show hot spot locations. Compared to system model diagram that also represented SO-Objective, here symmetry is designed into the system by introducing the concept of CG. Thus SO-Objective is transformed into much simpler and SO-Goal by dissolving the condition within the objective function.
sector with no hot spot i.e. perfectly uniform user distribution and \( \rho^n = 0 \) represents a scenario where all the users in the sector are concentrated on same location or equally a single user is using all resources. Thus sector with \( \rho << 1 \) has a hot spot and its CG can be assumed to be the point \( s \) i.e. geographical center of the hot spot. In order to determine the center of gravity of sectors with no clear hot spot, we present following propositions:

**Proposition 1.** for

\[
\rho^n = \frac{A^n|K^n|}{A^n|K^n_2|} = 1
\]  

(3.16)

\[
\frac{1}{|K^n|} \sum_{\forall k \in K^n} \log_2(1 + \gamma^n_k) \geq \frac{1}{|K^n|} \sum_{\forall k \in K^n} \log_2(1 + \gamma^n_k)
\]  

(3.17)

if

\[
\int \int \left( (\theta^n_{x,y} - \tilde{\theta}^n_{x,y}) \gamma^n_{x,y} \right) dx dy = 0
\]  

(3.18)

Integral in (3.18) is surface integral over whole area of the sector \( n \) and \((x,y)\) denotes the coordinates of given point \( k \) in the sector.

**Proof.** Proof of proposition 1 is provided in Appendix A

There may be sectors that do not have clear hot spot to have straightforward CG but at the same time their distribution can not be approximated with perfectly uniform user distribution. For such possible though rare scenario we present following proposition

**Proposition 2.** for

\[
\rho^n = \frac{A^n|K^n|}{A^n|K^n_2|} \leq 1
\]  

(3.19)

\[
\frac{1}{|K^n|} \sum_{\forall k \in K^n} \log_2(1 + \tilde{\gamma}^n_k) \geq \frac{1}{|K^n|} \sum_{\forall k \in K^n} \log_2(1 + \gamma^n_k)
\]  

(3.20)

if

\[
\sum_{k=1}^{K^n} \left( (\theta^n_k - \tilde{\theta}^n_k) \gamma^n_k \right) = 0
\]  

(3.21)
Proof. Proof of proposition 2 is part of proof of proposition 1 and is provided in Appendix A.

Now we show how we can use propositions 1 and 2 to reduce the complexity of our problem by dissolving the condition (3.13) into the objective function itself. Note that (3.18) and (3.21) return optimal tilt $\hat{\theta}_{\text{opt}}$ for $n^{th}$ sector dependent on the user distribution, for given tilts of neighbouring sectors. This optimal tilt angle can be translated into the locus or set of certain points $\mathcal{F}$ in the sector such that

$$d_n = H_n \times \tan(\hat{\theta}_{\text{opt}})$$  \hspace{1cm} (3.22)$$

$$\mathcal{F} = \{ f | f \in \mathcal{K}, d(f \leftrightarrow n) = d_n \}$$ \hspace{1cm} (3.23)$$

where $H_n$ is height $n^{th}$ sector antenna i.e. BS height and $d(f \leftrightarrow n)$ denotes distance between the location of the sector $n$ and a user represented by location $f$ respectively. Note that according to proposition 1 and 2 a tilt angle optimized for any of points $f^n \in \mathcal{F}$ optimises average throughput of users in a sector $n$. We call this point CG of the sector $n$. Therefore, the CG's for sectors with uniform or not uniform but hot spot less user distribution can be determined through proposition 1 and 2 respectively. It is important to mention that proposition 1 and 2 yields a locus of points as expressed with (3.22) and (3.23). Any point on this locus can be considered as CG as illustrated in figure 3.4.

Thus for all sectors with no traffic hot spots we can define a set of CG's $\mathcal{G}$ that can be determined using (3.22) and (3.23) such that.

$$\mathcal{G} = \bigcup_{n=1}^{W-S} f^n, \quad f^n \in \mathcal{F}$$ \hspace{1cm} (3.24)$$

Hence, using proposition 1 and 2, sectors that do not have a hot spot can
3.4. Hot Spot Traffic Relief by Tilt Optimization through BSOF(TO-BSOF)

still be considered to have virtual hotspot such that $|\mathcal{N}| = |S \cup \mathcal{G}|$. Since

$$\frac{1}{|\mathcal{K}|} \sum_{vk \in \mathcal{K}} \log_2 (1 + \gamma_{vn}^p) \geq \frac{1}{|\mathcal{K}|} \sum_{vk \in \mathcal{K}} \log_2 (1 + \gamma_{vn}^p) \implies$$

$$\frac{1}{|\mathcal{N}\setminus S|} \sum_{vn \in \mathcal{N}\setminus S} \left( \frac{1}{|\mathcal{K}|} \sum_{vk \in \mathcal{K}} \log_2 (1 + \gamma_{vn}^p) \right) \geq \frac{1}{|\mathcal{N}\setminus S|} \sum_{vn \in \mathcal{N}\setminus S} \left( \frac{1}{|\mathcal{K}|} \sum_{vk \in \mathcal{K}} \log_2 (1 + \gamma_{vn}^p) \right) \quad (3.25)$$

Based on arguments presented above through (3.17)–(3.25) the objective in (3.10) with condition (3.11) dissolved in it can be written as

$$\max_{\theta_{\text{tilt}}} \eta \left( \theta_{\text{tilt}}^N \right) = \max_{\theta_{\text{tilt}}} \sum_{vp \in \mathcal{P}} \log_2 \left( 1 + \gamma_{vp}^u \left( \theta_{\text{tilt}}^N \right) \right) \quad (3.26)$$

where $\mathcal{P} = S \cup \mathcal{G}$. This is shown in figure 3.3 where circles represent hot spots i.e. $S$ and stars represent centers of gravity or virtual hot spots in sectors with uniform user distribution i.e. $\mathcal{G}$. In next subsection we proceed to next step of BSOF i.e. decomposition of SO-Goal into SO-Functions.
3.4.3 Decomposing SO-Goal to Design SO-Functions

While SO-Goal in (3.26) is much simpler than the SO-Objective in (3.10)-(3.11) it is still a non-linear non-convex large scale optimization problem (see Appendix D) whose solution would require global cooperation. As discussed in section 2.7.2.3 on coopetition, in addition to cooperation, competition is also desired element of SO as it can bring automaticity and decreases complexity. While designing SO-Functions to achieve this SO-Goal this fact can help us dissolve the remaining complexity. There are two main reasons of complexity in this SO-Goal; firstly it is large vector optimization over vector $\theta_{tilt}^n$ that has as many components as number of sectors in the system i.e. $|\mathcal{N}|$ which prevents scalability. Secondly the mutual coupling between these variables that requires global cooperation. In order to disintegrate this simplified but still global SO-Goal into local SO-Functions we present following propositions.

**Proposition 3.** The aggregate throughput at the CG's in the system when interference from only two immediate neighbouring sectors is considered, will be greater or equal to the aggregate throughput at the same points when interference from all the sectors is considered. Mathematically:

$$\hat{\eta} \geq \eta \quad (3.27)$$

if

$$\hat{\eta} = \sum_{p \in \mathcal{P}} \log_2 (1 + \hat{\gamma}_p (\theta_{tilt}^n)) \quad (3.28)$$

where

$$\hat{\gamma}_p (\theta_{tilt}^n) = \frac{h_p n^{10^\mu (\phi_{p}^n - \phi_{tilt}^n)^2 + c_p}}{\sum_{\forall t \in \mathcal{T} \setminus n} h_p n^{10^\mu ((\phi_{t}^n - \phi_{tilt}^n)^2 + c_t)}} \quad n, t \in \mathcal{T}, \mathcal{T} \subset \mathcal{N} \quad (3.29)$$
3.4. Hot Spot Traffic Relief by Tilt Optimization through BSOF(TO-BSOF)

\( n \) here represents a sector in which point \( p \) lies and \( T \) is set of \( n \)th and the two other most interfering sectors adjacent to \( n \)th sector all mutually facing each other and termed as triplet (as illustrated in figure 3.5 by dashed red lines) and \( \theta_{n}^{T} \) is vector of \( T \) tilt angles of sectors within the triplet in which point \( p \) lies such that \( |T| = T \).

**Proof.** Proposition 3 is quite intuitive and order to prove it, we actually need to show that,

\[
\gamma_{p}^{n} (\theta_{n}^{T}) \leq \gamma_{p}^{n} (\theta_{n}^{T}), \quad \forall p \in \mathcal{P}
\]  

(3.30)

i.e.

\[
\frac{h_{p}^{n} 10^{\mu((\theta_{p}^{n} - \theta_{n}^{T})^2 + \phi_{n}^{m})}}{\sum_{\forall m \in \mathcal{N} \setminus n} (h_{k}^{m} 10^{\mu((\theta_{p}^{m} - \theta_{n}^{T})^2 + \phi_{n}^{m})})} \leq \frac{h_{p}^{n} 10^{\mu((\theta_{p}^{n} - \theta_{n}^{T})^2 + \phi_{n}^{m})}}{\sum_{\forall t \in \mathcal{T} \setminus n} (h_{p}^{t} 10^{\mu((\theta_{p}^{t} - \theta_{n}^{T})^2 + \phi_{n}^{m})})}, \forall p \in \mathcal{P}
\]  

(3.31)

by inverting the both sides

\[
\sum_{\forall m \in \mathcal{N} \setminus n} (h_{k}^{m} 10^{\mu((\theta_{p}^{m} - \theta_{n}^{T})^2 + \phi_{n}^{m})}) \geq \sum_{\forall t \in \mathcal{T} \setminus n} (h_{p}^{t} 10^{\mu((\theta_{p}^{t} - \theta_{n}^{T})^2 + \phi_{n}^{m})}), \forall p \in \mathcal{P}
\]  

(3.32)

By opening the left hand side

\[
\sum_{\forall t \in \mathcal{T} \setminus n} (h_{p}^{t} 10^{\mu((\theta_{p}^{t} - \theta_{n}^{T})^2 + \phi_{n}^{m})}) + \sum_{\forall m \in \mathcal{N} \setminus T} (h_{k}^{m} 10^{\mu((\theta_{p}^{m} - \theta_{n}^{T})^2 + \phi_{n}^{m})})
\]

\[
\sum_{\forall m \in \mathcal{N} \setminus n} (h_{k}^{m} 10^{\mu((\theta_{p}^{m} - \theta_{n}^{T})^2 + \phi_{n}^{m})})
\]

since

\[
\sum_{\forall m \in \mathcal{N} \setminus T} (h_{k}^{m} 10^{\mu((\theta_{p}^{m} - \theta_{n}^{T})^2 + \phi_{n}^{m})}) \geq 0
\]  

(3.33)

Hence the proposition in (3.30) is true. Since \( \eta \) is monotonically increasing function of \( \gamma \), hence, \( \gamma_{p}^{n} \leq \gamma_{p}^{n}, \forall p \in \mathcal{P} \) implies that \( \eta \geq \eta \).
Proposition 4. As $\beta$ and the cell radius grows large $\hat{\eta}$ becomes closer approximation of $\eta$ i.e. for large value of $\beta$ and large cell sizes, $\hat{\eta} \approx \eta$.

Proof. Proof of proposition 1 is actually a corrolary of proposition 3, and can be easily proved by putting large values of $\beta$ and $d$ in (3.33). □

Proposition 5. $\hat{\eta}_{q,\text{max}} = \hat{\eta}_{\text{max}}$, where

$$\hat{\eta}_{\text{max}} = \max_{\theta_{\text{tilt}}^N} \hat{\eta} \left( \theta_{\text{tilt}}^N \right) = \max_{\theta_{\text{tilt}}^N} \sum_{p \in P} \log_2 \left( 1 + \hat{\gamma}_p \left( \theta_{\text{tilt}}^N \right) \right)$$

where $\hat{\gamma}_p$ is the approximate SIR at point $p$ representing CG determined through (3.29) and

$$\hat{\eta}_{q,\text{max}} = \sum_{q \in \mathcal{Q}} \hat{\eta}_{q,\text{max}}$$

where

$$\hat{\eta}_{q,\text{max}} = \max_{\theta_{\text{tilt}}^q} \sum_{p \in \mathcal{P}_q} \log_2 \left( 1 + \hat{\gamma}_p \left( \theta_{\text{tilt}}^q \right) \right), \quad \mathcal{P}_q \subset \mathcal{P}, \mathcal{T}_q \subset \mathcal{Q}, \forall q \in \mathcal{Q}$$

where $\mathcal{T}_q$ is the $q$th set of triplets as explained in proposition 3 and illustrated in figure 3.5 such that $|\mathcal{P}_q| = |\mathcal{T}_q| = 3, \forall q \in \mathcal{Q}$, $\theta_{\text{tilt}}^q$ is vector of tilt angels of sectors within $q$th triplet and

$$\mathcal{P}_q \cap \mathcal{P}_{q'} = \emptyset \quad \text{and} \quad \mathcal{T}_q \cap \mathcal{T}_{q'} = \emptyset \quad \forall q \neq q', \forall q, q' \in \mathcal{Q}$$

$\mathcal{Q}$ is set of all such triplets, such that $\mathcal{T}_q \subset \mathcal{N}$ and $|\mathcal{Q}| = \frac{|\mathcal{N}|}{|\mathcal{T}_q|}$ is the total number of triplets in the system.

Proof. Since $|\mathcal{Q}| \times |\mathcal{T}_q| = |\mathcal{Q}| \times |\mathcal{P}_q| = |\mathcal{N}| = |\mathcal{P}|$ so the (3.34) can be written in the following for

$$\hat{\eta}_{\text{max}} = \max_{\theta_{\text{tilt}}^N} \left\{ \sum_{p \in \mathcal{P}_1} \log_2 \left( 1 + \hat{\gamma}_p \left( \theta_{\text{tilt}}^1 \right) \right) + \sum_{p \in \mathcal{P}_2} \log_2 \left( 1 + \hat{\gamma}_p \left( \theta_{\text{tilt}}^2 \right) \right) + \sum_{p \in \mathcal{P}_3} \log_2 \left( 1 + \hat{\gamma}_p \left( \theta_{\text{tilt}}^3 \right) \right) + \cdots + \sum_{p \in \mathcal{P}_q} \log_2 \left( 1 + \hat{\gamma}_p \left( \theta_{\text{tilt}}^q \right) \right) \right\}$$

(3.38)
where $|Q| = Q$. According to (3.37) $P_q \cap P_{q'} = T_q \cap T_{q'} = \emptyset, \forall q \neq q'$ where $q, q' \in Q$ i.e. all the terms in the above series are disjoint that implies that they are not mutually dependent, therefore the maximization can be performed on the individual terms of the series, so (3.38) can be written as

$$
\hat{n}_{\text{max}} = \max_{\theta_{\text{init}}^n} \sum_{p \in P_1} \log_2 \left(1 + \gamma_p^n \left(\theta_{\text{init}}^n\right)n\right) + \max_{\theta_{\text{init}}^n} \sum_{p \in P_2} \log_2 \left(1 + \gamma_p^n \left(\theta_{\text{init}}^n\right)n\right)
$$

$$
= \max_{\theta_{\text{init}}^n} \sum_{p \in P_1} \log_2 \left(1 + \gamma_p^n \left(\theta_{\text{init}}^n\right)n\right) + \max_{\theta_{\text{init}}^n} \sum_{p \in P_2} \log_2 \left(1 + \gamma_p^n \left(\theta_{\text{init}}^n\right)n\right)
$$

(3.39)

closing the summation

$$
\max \hat{n} = \max_{\theta_{\text{init}}^n} \sum_{p \in Q} \log_2 \left(1 + \gamma_p^n \left(\theta_{\text{init}}^n\right)n\right)
$$

(3.40)

Hence proved

$$
\hat{n}_{\text{max}} = \hat{n}_{Q,\text{max}} = \sum_{\theta_{\text{init}}^n} \hat{n}_{k,\text{max}}
$$

(3.41)

Each term inside the summation in (3.41) represents the SO-Function to be executed by each triplet. Figure 3.5 shows how SO-Functions can be executed individually in each triplet. In natural SO system nature or nurture does the job of teaching execution of SO-functions (figure 2.4 and 2.5). In case of engineering systems, the system designer has to do the this task. This can be done using any of the suitable methodologies listed in figure 2.3. Next subsection presents an optimization based methodology to solve the SO-Function derived in this section.
Figure 3.5: Red lines show a triplet and $S$ represent the location of hotspots or center of gravity in each sector. In order to achieve SO-Goal, each triplet performs its SO-Function independently from other triplets (competition), whereas within each triplet all three sectors cooperate with each other to jointly optimize and maintain their tilts with respect to the three center of CG's in that triplet.
3.4.4 Solving the SO-Function: Tilt Optimization Problem within a Triplet

A single SO-Function to be executed within $q^{th}$ triplet is given as

$$\max_{\theta_{t,u}^q} \sum_{\nu \in C_{\nu}} \log_2 \left( 1 + \gamma_\nu^q \right) \quad \text{(3.42)}$$

To enable execution of these SO-Functions in each triplet, the optimization problem in (3.42) needs to be solved. Below we present a solution methodology for the SO-Functions.

If $C$ is the total achievable throughput at the CG in a given triplet (subscript $q$ dropped for simplicity of expression), then

$$C = \log_2 \left( 1 + \gamma_1^q \right) + \log_2 \left( 1 + \gamma_2^q \right) + \log_2 \left( 1 + \gamma_3^q \right) \quad \text{(3.43)}$$

where subscript denote sector and subscripts denote CG's within given triplet as shown in figure 3.5

$$C = \log_2 \left( 1 + \left( \frac{h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)} + \frac{h_3^2 10^{-1.2\mu((\theta_2^2-\theta_{t,u}^q)^2+\epsilon)}}{\left( h_3^2 10^{-1.2\mu((\theta_2^2-\theta_{t,u}^q)^2+\epsilon)} + \left( h_3^2 10^{-1.2\mu((\theta_2^2-\theta_{t,u}^q)^2+\epsilon)} \right) \right)} + \right) \right) +$$

$$\log_2 \left( 1 + \left( \frac{h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)}}{\left( h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)} + \left( h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)} \right) \right)} + \right) \right) +$$

$$\log_2 \left( 1 + \left( \frac{h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)}}{\left( h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)} + \left( h_1^3 10^{-1.2\mu((\theta_1^1-\theta_{t,u}^q)^2+\epsilon)} \right) \right)} + \right) \right) \quad \text{(3.44)}$$

in order to maximize $C$ as a function of tilt angles, we put,

$$\max_{\theta_{t,u}^1, \theta_{t,u}^2, \theta_{t,u}^3} \ C \left( \theta_{t,u}^1, \theta_{t,u}^2, \theta_{t,u}^3 \right) \quad \text{(3.45)}$$
Figure 3.6: Aggregate throughput of the three hot spots in a triplet of sectors as a function of tilt angles of two sectors. Tilt angle of third sector is fixed at 13° for ease of plotting. Though there is a clear global optimum but the function is not concave.

subject to:

$$\theta_{tilt}^1, \theta_{tilt}^2, \theta_{tilt}^3 < \frac{\pi}{2}$$  \hspace{1cm} (3.46)

Figure 3.6 plots $C$ versus $\theta_{tilt}^1$ and $\theta_{tilt}^2$ for fixed value of $\theta_{tilt}^3$. It can be seen that $C$ is not a concave function of $\theta$ but it has a clear global optimum. Since the number of optimization parameters is only three and their range is also limited i.e., $0 < \theta < 90$, so the solution of 3.46 can be easily determined using a nonlinear optimization technique that can tackle a non-concave optimization objective.

Since we have limit constraints on the optimization variables and the objective function is nonlinear, so it is a nonlinear constrained optimization problem. Furthermore, as it can be seen in figure 3.6, that the objective function is not concave (convex), this makes this problem a non-convex constrained optimization problem.
3.4. Hot Spot Traffic Relief by Tilt Optimization through BSOF (TO-BSOF)

This non convexity means that classic lagrangian multipliers method for constrained optimization can not used here because there might be non zero duality gap between the primary i.e. original objective function and dual function i.e. lagrangian function (see Appendix D for details on convex optimization).

Solving a non convex multi variable constrained optimization problem is challenging task in general. Fortunately, here is where the use of BSOF will pay us back, as through the steps of BSOF we have reduced our original problem in (3.10) from a large scale optimization problem (i.e. a optimization problem over large number of variables) to small scale optimization problem as the number of optimization variables in (3.45) are just 3 now.

This transition from large scale to small scale optimization problem allows a variety of optimization techniques to be used for solving this problem. For example, assuming that tilting solution’s accuracy of more than 1° degree is not of practical significance, the problem can be even formulated as integer programming (Appendix D). In this case even the exhaustive search is also feasible as only $90 \times 90 \times 90 = 72900$ evaluations of (3.44) would be required to find the globally optimal solution each time hotspots/CG’s change their location in triplet. A relatively faster option can be the use of combinatorial optimization techniques like Tabu search to reduce the computational time i.e. improve agility. Another option can be the evolutionary heuristics listed in table 2.3 e.g. genetic algorithms or neural networks. However, all of these techniques except the exhaustive search do not guarantee convergence to global solutions.

Exhaustive search although is feasible and stable option because of small solution space, it is not the most time efficient method.

A more sophisticated approach is to use advanced non linear optimization techniques e.g. iterative methods (see Appendix D for details). Noticing the fact

---

7See appendix D for details on various optimization techniques
Chapter 3. Spectral Efficiency Optimization through SO of Antenna Tilting

that our objective function is twice continuously differentiable and the constraint function is continuously differentiable. Sequential Quadratic Programming (SQP) can be used to find the solution here (see appendix for details on SQP and relevant information on optimization theory). Below we explain how the SQP can be used to solve the problem under considerations.

For sake of clarity, we drop the subscript tilt. Instead we use subscript to present the association with sector in the triplet. Then the problem can be written in the standard form as

\[
\min_{\theta} -C(\theta) \tag{3.47}
\]

subject to:

\[
g_j(\theta) < 0 \quad j = 1, 2, 3 \tag{3.48}
\]

where \( \theta = [\theta_1, \theta_2, \theta_3] \) and \( g_j(\theta) = \theta - \frac{\pi}{3} \). If \( \hat{H} \) denotes the approximate of the Hessian matrix \( H \), then we can define quadratic subproblem to be solved at \( r^{th} \) iteration of SQP as follows

\[
\min_{w \in \mathbb{R}^3} \frac{1}{2} w^T \hat{H}(\mathcal{L}(\theta, \lambda)) w + \nabla \mathcal{L}(\theta, \lambda)^T w 
\]

subject to:

\[
w_j + \theta_j - \frac{\pi}{2} < 0 \quad j = 1, 2, 3 \tag{3.50}
\]

where Hessian is given as

\[
H(\mathcal{L}) = \begin{bmatrix}
\frac{\partial^2 \mathcal{L}}{\partial \theta_1^2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_1 \partial \theta_3} \\
\frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2^2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_2 \partial \theta_3} \\
\frac{\partial^2 \mathcal{L}}{\partial \theta_3 \partial \theta_1} & \frac{\partial^2 \mathcal{L}}{\partial \theta_3 \partial \theta_2} & \frac{\partial^2 \mathcal{L}}{\partial \theta_3^2}
\end{bmatrix} \tag{3.51}
\]

Where \( \mathcal{L} \) is lagrangian of above original problem and can be written as

\[
\mathcal{L}(\theta, \lambda) = C(\theta) - \lambda^T g
\]

\[
\mathcal{L}(\theta, \lambda) = C(\theta) - \sum_{j=1}^{3} \lambda_j (\theta_j - \frac{\pi}{2}) \tag{3.53}
\]
3.4. Hot Spot Traffic Relief by Tilt Optimization through BSOF(TO-BSOF)

Below we briefly describe the three main steps taken to solve the above problem through SQP.

1. **Updating the \( \mathbf{H} \):** At each iteration the value of \( \mathbf{H} \) is updated using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) approximation method i.e.

\[
\mathbf{H}_{r+1} = \mathbf{H}_r + \frac{b_r b_r^T}{b_r^T a_r} - \frac{\mathbf{H}_r a_r a_r^T \mathbf{H}_r}{a_r^T \mathbf{H}_r a_r} \quad (3.54)
\]

where

\[
a_r = \theta_{r+1} - \theta_r \quad (3.55)
\]

\[
b_r = \left(\nabla C (\theta)_{(r+1)} - \sum_{j=1}^{3} \lambda_j \nabla g_j (\theta_{r+1})\right) - \left(\nabla C (\theta)_{(r)} - \sum_{j=1}^{3} \lambda_j \nabla g_j (\theta_{r})\right) \quad (3.56)
\]

2. **Solution of Quadratic subproblem:** Once the Hessian is known the problem in (3.49) is a quadratic programming problem that can be solved using standard methods. We use gradient projection method as described in [59].

3. **Line search and Merit function** The solution of the quadratic subproblem in the \( r^{th} \) iteration of SQP algorithm returns the vector \( \mathbf{w}_r \) that provides the locus for next iteration as follows

\[
\theta_{r+1} = \theta_r + \alpha \mathbf{w}_r \quad (3.57)
\]

where \( \alpha \) is set such that sufficient decrease in the merit function is achieved. We use merit function defined in [60] i.e. given as

\[
\psi(\theta) = C(\theta) + \sum_{j=1}^{3} \mu_j \max(0, g_j(\theta_j)) \quad (3.58)
\]

where \( \mu \) is penalty parameter which we set as recommended in [60] i.e.

\[
\mu_{j,(r)} = \mu_{j,(r+1)} = \max_j \left\{ \lambda_j, \frac{\mu_{j,(r)} + \lambda_j}{2} \right\}, \quad j = 1, 2, 3 \quad (3.59)
\]
Through the above steps of SQP, the SO-Function in (3.42) can be solved within each triplet independently to determine the optimal tilt angle to be adapted and maintained by each triplet for given locations of CG's within that triplet.

The execution of these SO-Functions in each triplet in the WCS independently results in achievement of the SO-Goal, that in turn manifests SO-Objective i.e. system wide throughput optimization in presence of hot spot based non uniform traffic distribution. We call this framework TO-BSOF (Tilt Optimization through BSOF). Next subsection concludes this section by presenting a brief recap of the LB-BSOF.

3.4.5 Recap of Tilt Optimization through BSOF (TO-BSOF)

In this section framework called TO-BSOF for SO of system wide antenna tilts to enhance spectral efficiency in presence of traffic hot spots was developed by applying BSOF. Figure 3.7 summarises TO-BSOF developed in this section in simple words. First in subsection 3.4.1 the SO-Objective for this problem was indentified. In this section the problem of spectral efficiency optimization through system wide antenna tilt adaptation in face of non homogenous user geographical distribution was formulated as non linear multi variable constrained optimization problem as equations (3.10)—(3.11). In subsection 3.4.2 the complex SO-Objective was transformed into a constraint free yet complex optimization problem that serve as an SO-Goal given by (3.40). In subsection 3.4.3 this SO-Goal is further disintegrated into SO-Function that are much simpler and local optimization problems in (3.40). Finally, in subsection 3.4.4 a generic solution methodology for solving these SO-Functions is presented.

The execution of these SO-Functions in each triplet in the cellular system independently results in achievement of the SO-Goal, that in turn manifests SO-Objective i.e. system wide throughput optimization in presence of hot spot based non unifo-
3.4. Hot Spot Traffic Relief by Tilt Optimization through BSOF (TO-BSOF)  

Figure 3.7: A visual illustration of steps of TO-BSOF

form traffic distribution. Maintaining the maxima of these SO-Functions locally in each triplet, results in achievement of SO-Goal in (3.40) that in turn manifests the complex SO-Objective in (3.10) approximately and objective in (3.34) exactly. Here it is important to note that achieving the original SO-Objective in (3.10) even approximately is still valuable as there is no overhead of global signalling associated with this solution.

It is interesting to note that how TO-BSOF is built on principle of coopetition, i.e. it is based on cooperation within the sectors of triplet where as competition among triplets of the WCS, as explained in figure 3.5.

TO-BSOF is implementable in distributed and self organising manner and performance very close to optimal can be achieved. The main advantage of TO-BSOF is that it does not have heavy overheads associated. Although the globally optimal performance is not aimed for but this is analogous to the case that birds...
never fly in perfect V-shape, but even maintaining a near V-Shape increases their flight efficiency significantly [15] (significant performance gain achievable through TO-BSOF is evaluated and will be discussed in next section on system level simulation results). Furthermore, as postulated in [4] one of the four main paradigms for designing self organization into system is that for perfect self organization perfect objectives should not be aimed for. So here the self organizing, nature of solution is perfect but at cost of sub-optimal global objective. This sub-optimal solution reaches very close to optimal if the conditions in Proposition (4) are satisfied and investigation of this proximity can be a topic of further future work. In next section the performance of TO-BSOF is evaluated under number of different scenarios through system level simulations.

3.5 System Level Performance Evaluation

In this section we evaluate the performance of TO-BSOF presented in previous sections by implementing it in a system level simulator that models an OFDMA based generic WCS. To avoid unnecessary complexity and make the results more inference able, a circuit switching based simulation model is used. It should be noted that, TO-BSOF mainly plays with physical antenna parameters and its performance is not directly dependent on MAC layer details. Therefore, the results obtained in this section are equally valid to predict usefulness of TO-BSOF for packet switch WCS as well e.g. LTE. Table 3.1 shows the summary of simulation parameters used for performance evaluation of TO-BSOF.

In order to present a comparative analysis of TO-BSOF, we also evaluate the performance of number of other fixed tilting options used in the conventional WCS i.e. Performance is evaluated with all antennas in systems tilted by four different arbitrary tilt angles of $0^\circ, 5^\circ, 20^\circ, 25^\circ$. Results for these system configurations are denoted by legend ‘Tilt=$x$,SO=OFF’. Where $x$ denotes the specific value of
### 3.5. System Level Performance Evaluation

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>System topology</td>
<td>19 BS with 3 sector/cells per BS</td>
</tr>
<tr>
<td>Frequency Reuse</td>
<td>1 (Full frequency reuse)</td>
</tr>
<tr>
<td>BS Transmission Power</td>
<td>39 dBm</td>
</tr>
<tr>
<td>Inter site distance</td>
<td>1200 meters</td>
</tr>
<tr>
<td>BS height</td>
<td>32 meters</td>
</tr>
<tr>
<td>User height</td>
<td>1.5 meters</td>
</tr>
<tr>
<td>User antenna gain</td>
<td>0 dB (Omini directional)</td>
</tr>
<tr>
<td>BS antenna horizontal beamwidth, $B_h$</td>
<td>$70^\circ$</td>
</tr>
<tr>
<td>BS antenna vertical beamwidth, $B_v$</td>
<td>$10^\circ$</td>
</tr>
<tr>
<td>BS antenna vertical Gain Weight, $\lambda_v$</td>
<td>0.5</td>
</tr>
<tr>
<td>BS antenna vertical Gain Weight, $\lambda_h$</td>
<td>0.5</td>
</tr>
<tr>
<td>BS antenna maximum gain, $G_{max}$</td>
<td>18 dB</td>
</tr>
<tr>
<td>BS antenna maximum attenuation, $A_{max}$</td>
<td>20 dB</td>
</tr>
<tr>
<td>Frequency</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Pathloss model</td>
<td>Urban Macro [61]</td>
</tr>
<tr>
<td>Modulation and Coding Schemes (MCS)</td>
<td>LTE compliant [62]</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>5 MHz</td>
</tr>
<tr>
<td>Number of data sub carriers</td>
<td>326</td>
</tr>
<tr>
<td>Shadowing standard deviation</td>
<td>8 dB</td>
</tr>
<tr>
<td>Scheduling</td>
<td>Round Robin</td>
</tr>
<tr>
<td>Total user population</td>
<td>10000 users</td>
</tr>
<tr>
<td>Mean call arrival rate (Poison)</td>
<td>150 calls per second</td>
</tr>
<tr>
<td>Mean call holding time (Exponential)</td>
<td>120 seconds</td>
</tr>
<tr>
<td>Hot Spots Radius (HSR)</td>
<td>10% of cell radius (located randomly)</td>
</tr>
<tr>
<td>Percentage of users concentrated in hot spots</td>
<td>80% and 50%</td>
</tr>
</tbody>
</table>
fixed tilt angle. The results with TO-BSOF in place are denoted with legend \( \text{Tilt=auto,SO=ON} \) that means SO is 'ON' and hence the tilt angles are determined autonomously for each cell by TO-BSOF depending on user geographical distribution in that cell.

Performance is evaluated on the downlink in terms of three different performance aspects i.e. 1) CDF's of bandwidth normalised achievable throughput per users i.e. user spectral efficiency 2) Mean or expected spectral efficiency achieved per user given in units of \( \text{bps/Hz/User} \) and 3) Fairness among user's achievable spectral efficiency's using Jains's index. All these three metrics are evaluated separately for hotspot users only, as well as, for all users in the coverage area. Simulations are performed for two different cases that represent region characterised by \( \rho << 1 \) as indicated in figure 3.2 i.e. traffic distribution with severe hotspots.

**Case 1** is scenario where Traffic Distribution Factor (TDF) is 50% i.e. 50% of the whole users population is concentrated in the hot spots. The radii of these hotspots are 10% of sector radius thus resulting in \( \rho = 0.2 \).

**Case 2** represents more sever hot spots where TDF=80% hot spots size is still 10% of sector radius resulting in even smaller \( \rho = 0.125 \). The location of these hotspots is random as shown in figure 3.8.

### 3.5.1 CDF of User Throughput

CDF's for two types of user throughput are discussed in this section i.e. theoretical and practical throughput. For theoretical throughput, we use the Shannon’s theoretical bound, i.e. we map the SINR of each user to the maximum achievable spectral efficiency (i.e. channel capacity normalised by bandwidth) using Shannon’s capacity equation [63]. Practical throughput is determined using practical...
Figure 3.8: User geographical distribution for case 1. Users are not spread in the outer most tier of sectors, as no complete triplet can be formed there to execute TO-BSOF. Therefore, results based on the user population confined within the inner tiers of cells that can be part of a triplet are reported only.
Figure 3.9: CDF of theoretical throughput for all users and users in the hot spot for Case 1. Solid lines show throughput of all users in the system. Whereas dashed lines show throughput of users in the hotspots only.

MCS's of LTE listed in table 3.2.

3.5.1.1 Theoretical Throughput

In order to evaluate the full potential of TO-BSOF, the impact of TO-BSOF in terms of theoretical throughput is first evaluated as shown in figure 3.9. It can be seen that, without any antenna tilting user throughput is worst as expected because of high interference. A significant gain in throughput achievable with TO-BSOF can be observed in figure 3.9 compared to other fixed tilt configura-
3.5. System Level Performance Evaluation

Another important observation is that for all other fixed tilt configurations throughput for hotspot only and all users are almost same. But TO-BSOF provides a relatively higher throughput to hotspot users. This is because the TO-BSOF optimises the antennas with respect to CG’s in the sectors. In case of sectors with clear hotspots, center of the hotspots is taken as CG as explained above. It is important to mention that this gain is not provided by sacrificing throughput of non hotspot users as it can be seen in figure 3.9 that non hotspot user also get an increased throughput with TO-BSOF compared to all other scheme. Since the gain of TO-BSOF is manifested by increase in spectral efficiency, so the depth of modulation and coding schemes can have significant impact on the actual gain, the TO-BSOF can provide. Therefore, in next subsections we investigate performance of TO-BSOF using practical MCS considered in LTE and shown in table 3.2

3.5.1.2 Throughput with LTE Modulation and Coding Schemes

Figure 3.10 shows the CDF of bandwidth normalised practically achievable user throughput using LTE compliant modulation and Coding Schemes (MCS’s) listed in table 3.2. Despite of the significant decrease in the overall throughput compared to theoretical throughput in figure 3.9, due to limited number MCS’s available in this practical case, the gain of the TO-BSOF compared to all other schemes is fully preserved. It can be seen that, without any antenna tilting user throughput is worst as expected because of high interference. With small tilt of 5° user throughput improves significantly and as the tilt is increased upto 20° throughput remains nearly unchanged except a slight improvement for already privileged users i.e. near cell center users (top right of the CDF curves). This is because tilting down the antenna obviously favors cell center users compared to cell edge users. But for higher tilts the number of disfavoured users increases compared to favoured users and overall throughput starts deteriorating as can be seen by
Figure 3.10: CDF of practical throughput for Case 1. Unlike theoretical continuous modulation that yielded smooth CDFs shown in figure 3.9, the step nature of CDF curves in 3.10 is because of the discreteness of practical MCS's used here.
Figure 3.11: CDF of throughput for all users and users in the hotspot for case 2.
CDF for tilt of 25°. It can be seen that TO-BSOF outperforms all other fixed tilt configurations. The reason for this gain in throughput yielded by BSOF is as follows. As explained earlier there are two types of sectors in the system. Sectors with uniform distribution i.e. $\rho \approx 1$ and sectors with hotspots i.e. with $0 < \rho < 1$. For sectors with $\rho \approx 1$ the value of optimal tilt depends only on system parameters like antenna height, horizontal and vertical beamwidth and pathloss parameters etc. and is determined through (3.18). But for sectors with randomly located hotspots its value will be different for each sector depending on the location of its hot spot as that location will contain most of the users in that sector. To elaborate this fact figure 3.14 (left two graphs) plots average bandwidth normalised throughput i.e. spectral efficiency within triplet as function of the tilt angles of the two sectors in the triplet. Two different set of locations of hotspots in the same triplet are analysed. Totally different values of optimal tilt angles in a given triplet for different locations of hotspots show that how the optimal tilt angles in a triplet are sensitive to the location of hot spots in the sector and need to be determined accordingly to cope with spatio temporal dynamics of hot spots. In contrast to fixed antenna tilting or no tilting at all, implementation of TO-BSOF provides a self organising solution for such randomness both in time and space. Thus by setting optimal tilts adaptively for each sector within triplet depending on its traffic distribution, TO-BSOF yields a significant gain in system throughput. It is to be noted that although gain in throughput is higher for hotspot users (dashed black line in figure 3.10). This is because in TO-BSOF the SO-Objective in (3.10) is particularly designed to improve the throughput of hot spot users. However, with TO-BSOF in place the user throughput in the whole system is also improved with TO-BSOF (solid black line), as ensured by constraint in (3.13).

In figure 3.11 we plot practical throughput CDF's for case 2 of user geographical distribution. It can be seen in figure 3.11 that the gap between the hotspot users throughput and average users throughput shrinks, obviously because most of the
3.5. System Level Performance Evaluation

Table 3.2: Threshold SINR's for LTE modulation and coding schemes [2]

<table>
<thead>
<tr>
<th>MCS Index(l)</th>
<th>Modulation</th>
<th>Coding Rate</th>
<th>SINR</th>
<th>MCE(b/s/Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
<td>N/A</td>
<td>-5.1&gt;</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>QPSK</td>
<td>1/8</td>
<td>-5.1</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>QPSK</td>
<td>1/5</td>
<td>-2.9</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>QPSK</td>
<td>1/4</td>
<td>-1.7</td>
<td>0.5</td>
</tr>
<tr>
<td>4</td>
<td>QPSK</td>
<td>1/3</td>
<td>-1</td>
<td>0.667</td>
</tr>
<tr>
<td>5</td>
<td>QPSK</td>
<td>1/2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>QPSK</td>
<td>2/3</td>
<td>4.3</td>
<td>1.33</td>
</tr>
<tr>
<td>7</td>
<td>QPSK</td>
<td>3/4</td>
<td>5.5</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>QPSK</td>
<td>4/5</td>
<td>6.2</td>
<td>1.6</td>
</tr>
<tr>
<td>9</td>
<td>16QAM</td>
<td>1/2</td>
<td>7.9</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>16QAM</td>
<td>2/3</td>
<td>11.3</td>
<td>2.667</td>
</tr>
<tr>
<td>11</td>
<td>16QAM</td>
<td>3/4</td>
<td>12.2</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>16QAM</td>
<td>4/5</td>
<td>12.8</td>
<td>3.2</td>
</tr>
<tr>
<td>13</td>
<td>64QAM</td>
<td>2/3</td>
<td>15.3</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>64QAM</td>
<td>3/4</td>
<td>17.5</td>
<td>4.5</td>
</tr>
<tr>
<td>15</td>
<td>64QAM</td>
<td>4/5</td>
<td>18.6</td>
<td>4.8</td>
</tr>
</tbody>
</table>

users are now in hotspots. Furthermore, since, TO-BSOF improves spectral efficiency particularly at the hotspots, therefore, if we compare figure 3.10 with 3.11, we can see that the gap between without TO-BSOF CDF and with TO-BSOF CDF is much decisive in case 2 compared to case 1.

3.5.2 Expected Spectral Efficiency

While the CDF's have been useful to have qualitative assessment of gain of TO-BSOF in terms of spectral efficiency, a more quantitative assessment is desirable
to have figurative measure of this gain. In order to assess the net gain in spectral efficiency the TO-BSOF can provide, we evaluate the expected i.e. mean user spectral efficiency in the system. The expected spectral efficiency $\zeta$ that can be given as:

$$\zeta = \sum_{l=0}^{L} \left( MCE_l \times \frac{u_l}{U} \right) \left( \frac{b/s}{Hz} \right)$$  (3.60)

where

$$U = \sum_{l=0}^{L} u_l$$  (3.61)

where $MCE_l$ is the Modulation and Coding Efficiency of $l^{th}$ MCS out of the L total MCS used in the LTE and given in the table 3.2. $l = 0$ represents the case where SINR is so low that even the lowest modulation and coding pair can not be used to serve the user i.e. the user is in outage. $u_l$ denotes number of users that can be served with $l^{th}$ modulation and coding pair $U$ is the total number of users in the system.

The $\zeta$ is although similar to conventional average cell throughput in its semantics but the reason we chose this metric over conventional average cell throughput is that it has two main advantages: First, it takes the soft blocking i.e. outage due to high interference into account by including $l=0$ cases in the calculation of the reflective average cell throughput. Secondly it has in built provision fairness consideration in it as well, as it weighs each possible MCE by its probability of occurrence in the whole system for determine the average spectral efficiency.

Figure 3.12 plots the $\zeta$ for the case 1 and case 2 together. Here, the same trends are observable for all five system configurations as observed from CDF’s i.e. $\zeta$ improves with increasing tilt but then this improvement become saturated after certain tilt at which negative effect on remote users (disadvantaged because of tilting) overrides the positive effect on central users that are advantaged because of tilting. It can be seen that TO-BSOF out performs in both cases compared to all other four configurations with fixed tilting. In comparison to no tilting config-
Figure 3.12: Expected or mean spectral efficiency for all users in the whole area of WCS and users in the hotspots. This figure actually shows the means of the both sets of CDF’s plotted in figures 3.10 and 3.11 for case 1 and case 2 respectively.

Operation BSOF provides a 29.53% gain in $\zeta$ where as compared to best fixed tilting configuration TO-BSOF still provides 11% boost in $\zeta$ of hot spot users. Another important observation is to note that TO-BSOF yields this gain in performance without sacrificing the gain of non hotspot users. This can be observed by comparing the $\zeta$ for all users in the coverage area, for case 1 (i.e. more uniform user distribution compared to case 2), with and without TO-BSOF. It can be observed that in this case also TO-BSOF has a gain of around 29.8% compared to worst case of no tilt and 9% compared to the best case of fixed tilt i.e $tilt = 20^\circ$. 
3.5.3 Fairness

Although fairness is not the primary focus of this work and its not the SO-Objective for which the TO-BSOF is designed, but still its important to evaluate TO-BSOF’s effect on the fairness in the system to assess its practicality. We use Jain’s Fairness Index (JFI) as metric for fairness [64] that is given as:

\[ JFI = \frac{\left( \sum_{n=1}^{N} MCE_n \right)^2}{N \sum_{n=1}^{N} (MCE_n)^2} \]  

Figure 3.13 presents JFI evaluated for both of the cases for all five configurations with and without TO-BSOF. By contrasting with throughput CDF and ζ performance results in figures 3.11-3.12 we can easily see in figure 3.13 that here too exists the time persisting trade off between the fairness and capacity. But we can observe that TO-BSOF decreases fairness only slightly compared to other fixed tilting configurations that are best in terms of throughput and ζ. This decreased JFI is not because of compromise in fairness among hotspot and non hot spot users as can be seen by very closely matching CDF’s of hot spot and non hot spots in figure 3.14. As can be seen in figure 3.14 that within a triplet, optimal tilt angle for throughput optimization fairness optimization are not necessarily same within a triplet for the same hotspot. This means although TO-BSOF does not effect the fairness among the users within a sector, but in order to optimize tilt for throughput TO-BSOF have to slightly compromise on the fairness among users of different sectors within a triplet if those sectors contain hot spots at relatively different locations. Considering the this aspect of global fairness jointly with global throughput optimization is beyond the scope of this study and is one of our potential future works. It is important to highlight that TO-BSOF does not effect fairness in negative sense, i.e. it does not decrease the throughput of some users while increasing that of others rather it increases throughput of some users without decreasing that of others. This can be seen by examining the CDF of throughput in figure 3.9. The negative effect of TO-BSOF is barred through
3.5. System Level Performance Evaluation

In Whole Area; H SR =10; T D F =80%

In Whole Area; H SR =10; T D F =50%

In Hotspots Only, H SR =10; T D F =80%

In Hotspots Only, H SR =10; T D F =50%

Figure 3.13: Fairness for all users and users in the hot spot for hot spot radius of 10% of cell radius. The slight decrease in fairness can be traced back in figure 3.9 where a more transient CDF with TO-BSOF compared to other tilting scheme can be seen. This shows that the decrease in fairness is not because of decrease of throughput of some users, it is because of increase the throughput of some users without effecting the that of others.
Figure 3.14: Average Spectral Efficiency (SE) and Jain’s Fairness Index (JFI) for two different CG locations within the triplet. This result is evaluated for a single stand alone triplet by assuming a single user in each sector of the triplet located at the center of the hotspots or CG’s. Tilt angle of third sector in the triplet is kept constant at 0°. It can be seen that optimal tilt angles for maximum throughput change as location of CG’s changes. Furthermore, optimal tilts for throughput can be noticed to be slightly different than optimal tilts for fairness.
constraint in (3.11) and the slight decrease in throughput observed in left bottom of figure 3.10 and 3.11 is not observable in 3.9. It means that TO-BSOF does not decrease throughput of the users lying in that region rather this loss is because of lack of appropriate MCS such users can use. With a finer resolution of MCS that negligible loss in the throughput can also be recovered improving the overall practical throughput gain as well fairness of TO-BSOF.

In summary it can be seen that, compared other tilting options TO-BSOF does not effect fairness much and still provides a significant gain in overall system throughput while alleviating the hot spot problem by increasing the spectral efficiency of user links for hot spots users particularly and for regular users generally. The most important merit of TO-BSOF is that it provides this gain in performance in pure self organizing manner because its implementable locally within each triplet and still achieves the global objective. This makes TO-BSOF absolutely scalable, less complex and global signalling and coordination free. Furthermore as we will explain in next subsection, TO-BSOF is self sustained and does not require any human intervention for its operation.

3.6 Practical Implementation of TO-BSOF

It is important to highlight that TO-BSOF is practically implementable with the state of the art technologies. The hot spot position information can be easily gathered at the respective BS with existing location estimation technologies like GPS and location identification based on relative received signal delay etc. There is already an X2 interface between the neighbouring BS's in the LTE that can be used to exchange this information among the three adjacent BS's within the triplet. Based on this information, the optimal tilt angle for all the three sectors within the triplet can be determined and their tilts can be adapted accordingly.
Since in state of the art BS tilts can be adjusted electronically, so with the implementation of TO-BSOF BS's can autonomously adapt dynamically maintain their antenna tilts to cope with the spatio temporal dynamics of traffic. Therefore, this algorithm requires no human intervention.

Another advantage of TO-BSOF is that because of its highly localised nature it is very agile as it has no intrinsic delays caused by the required excessive global signalling or complex coordination. Therefore, TO-BSOF can be implemented in an online manner using event based triggering mechanisms. Such triggering mechanisms can log user location information to detect major variation in user geographical distribution and thus can trigger the SO of tilts through TO-BSOF. The execution of TO-BSOF can also be periodic in an off line manner. Such off line execution will not require real time position information of user locations, rather, it can rely on long term measured traffic data profile of the area of interest. In case of periodic execution the time period of re-execution can range from minutes to months and can be set based on the statistics of the long term measured traffic data profile of the area of interest.

3.7 Conclusions

In this chapter an analytical framework for SO of system wide tilts is developed using BSOF. The solution in this framework is designed to be scalable, agile and stable. The performance gain and operational stability of the solution is demonstrated through numerical results as well as extensive system level simulations for multi cell OFDMA based LTE type generic WCS. The results are evaluated for realistic and worst case scenarios with concentrated hot spots and have been shown to yield upto 30% gain in spectral efficiency in WCS in face of non homogenous geographical user distributions. The pragmatic nature of the solution is also established with detailed suggestions for practical implementation. Unlike
previous works on hot spot relief through tilt optimization, this solution does not necessitate handovers, does not require global cooperation, is not prone to oscillations and is fairness considerate and fully practical for OFDMA based cellular network like LTE and LTE-A. The essential novelty of the solution lies in two features of solution. Firstly, it is designed through BSOF and therefore has intrinsic SO marked by scalability, stability and agility, designed into it. Secondly it does not provide hotspot relief through transferring load to other sectors rather it enhances spectral efficiency at the affected regions e.g. hotspot centers to generate more capacity and thus relieves congestion.
Load Balancing through SO of Coverage

4.1 Introduction

In addition to short term pop up hotspots that have been addressed in the last chapter, low user satisfaction and poor resource efficiency also occurs from relatively medium term unbalanced traffic load among different cells. This unbalanced load may result from natural user location variation over day and night, permanent shadowing and various socio economic and demographical factors that may cause relatively permanent hotspots. These factors, altogether make user geographic distribution as well as their cell association non uniform and ever changing. By non uniform cell association we mean different number of users get associated with different cells even if the cell sizes are same.

Although these factors vary at much slower rate\(^1\) and does not behave like pop up hotspots\(^2\), the fact that they make the user distribution ever changing in a hard to predict fashion, makes them difficult to be considered and accounted for during planning deployment phase of WCS. The heterogenous traffic distribution resulting

\(^1\)see medium term dynamics in figure 1.1
\(^2\)see short term dynamics in figure 1.1
from such factors may cause congestion in one cell and increase the hard blocking probability while resulting in under utilization of resources in another cell at the same time. The TO-BSOF can still improve the radio resource efficiency in such scenario, by improving the spectral efficiency through focusing antenna tilts on CG’s of the traffic distribution in a cell in a SO manner. Nevertheless, an additional mechanism is still required to maintain minimum hard blocking through optimal resource utilization by relieving congestion in over loaded cells and avoiding under utilization in less loaded cells i.e. through Load Balancing (LB). This problem becomes more severe in Relay Station (RS) enhanced WCS as LB among the RS and BS adds an extra dimension to the problem due to different propagation characteristics of RS and BS and their largely different cell size. Since RS has to share resources with BS for their back haul access link, as a result adding RS can deteriorate resource efficiency further due to the inherent multiplexing loss caused by RS, if traffic load is not balanced among RS and BS optimally. Since future WCS like LTE-A have RS in its essential deployment plan, so there is dire need for some LB mechanism for relay enhanced WCS. Given the complexity and scale of the problem, it is desirable that this LB mechanism should be self-organising. To this end, in this chapter we present a novel self-organising framework for user satisfaction improvement through load balancing for both LTE and LTE-A type cellular networks. This framework is developed using BSOF and is termed as LB-BSOF.

This chapter is organised as follows: Since much work has been done on LB, therefore, section 4.2 is presents an extensive review of the related work. This section also aims to establish the necessary background and the novelty of our work. In section 4.3 we describe the system model used for problem formulation. In section 4.4 we apply BSOF to the problem of maximising user satisfaction in face of medium term dynamics to develop a SO solution i.e. LB-BSOF. In section 4.5 we discuss the pragmatic implementation aspects of LB-BSOF and
present a simple heuristic algorithm for its practical implementation. In section 4.6 we present the performance evaluation results of our proposed framework obtained through system level simulator. In section 4.7 we conclude this chapter.

4.2 Background and Related Work

Need for LB mechanisms in WCS to mitigate the effects of natural spatio-temporally varying user distribution was realised immediately after the advent of commercialised WCS [65]. A large number of research works since then have embarked on this problem and have proposed variety of very useful LB strategies some of which even reach the required level of scalability, stability and agility and can manifest desired SO [66,67]. However the problem lies in the fact that, most of these LB schemes were specific to the particular generations of WCS evolved so far and only a few are applicable to the emerging WCS e.g. LTE and LTE-A due to the differences in the MAC and physical layer discussed in section 3.2. In following subsection we will discuss the key works by broadly classifying them in four general categories based on their main underlying approach they take towards LB. i.e. 1) Resource Adaptation (RA) based LB, 2) Traffic Shaping (TS) based LB 3) Coverage Adaptation (LB) based LB, and 4) Relay Station (RS) based LB.

4.2.1 Resource Adaptation (RA) based Load Balancing

In this type of LB schemes the main underlying principle is to adapt amount of resources allocated to a cell to match it to the offered traffic load in that cell for optimal LB. More specifically an over loaded cell would borrow channels from the other cells that are less loaded or from a common pool of some free channels. Works like [66–69] presented variations of LB schemes building on this main idea.
Authors in [68] presented and evaluated the performance of channel borrowing algorithms where an overloaded cell will borrow channel from selected cells that are least loaded. Same authors further refined the idea in [68] and presented in [66] an improved version in terms of scalability. This better scalability was achieved by limiting the borrowing process within the hierarchical tier based local structures or clusters. However, this scalability was achieved at the cost of noticeable loss in overall performance due to lack of global optimality in the proposed distributed approach [66].

The scope of RA based LB schemes has been mainly limited to legacy GSM type systems where neighbouring cells use different frequency channels. In case of emerging OFDMA based WCS e.g. LTE and LTE-A, frequency reuse of 1 prevails and such channel borrowing will cause intra-cell interference that is not present otherwise in OFDMA based systems. Albeit, few authors have attempted to extend the channel borrowing concept to make it applicable to CDMA based WCS by introducing the idea of virtual channel borrowing for CDMA based systems [70]. The basic concept of channel borrowing has also been extended for emerging WCS by generalising it in form of bandwidth management strategies [67] [69]. While the work presented in [67] remarkably holds all features of desired SO i.e, scalability, agility and stability due to its cell by cell implementation style, its lack of pragmatism for WCS with universal frequency reuse undermines its applicability to the future WCS. Perhaps that is the reason that RA based LB schemes have not been further investigated in the context of LTE and LTE-A.

\[4.2.2\] Traffic Shaping (TS) based Load Balancing

In this type of schemes the main underlying principle is to shape the traffic offered to a cell either through pre-emptive and strategic admission control or forced handover of the ongoing calls. This Traffic Shaping (TS) is done in or-
order to effectively match the offered traffic load with the available resources for optimal LB. A significant number of works have focused on optimizing such TS strategies to minimize their trade off in terms of hand over overheads and interference [65,71-75]. These work were triggered from the pioneering paper [65] that presented a simple but seminal idea of selecting least loaded cells among candidate neighbour cells for handover with purpose of LB. A comprehensive review of LB balancing algorithms using the traffic shaping approach and other approaches have been presented in [75]. This paper also presents a mathematical general framework that models the underlying principles of variety of LB strategies.

Although call admission control is also a possible way of traffic shaping [72,74,76], but due to randomness of arrival times and user mobility, its use for optimal LB is relatively complex approach. Its complexity lies in the fact that it necessitates extensive cooperation among cells to ensure QoS so that calls rejected by one cell is accepted by others. On the other hand, handover parameters adaptation to trigger forced handovers of ongoing calls to shed off extra load to neighbouring cells, is more straightforward approach [65,71,73,77].

Authors in [77] presented a handover based approach for LB exclusively for LTE. The algorithm requires all cell loads to be known at central control unit and optimal handover parameters are determined based on it. Simulation results show that load can be balanced to a reasonable degree across cells with this approach. Signalling overhead and excessive delays incurred due to centralised nature are a compromise on scalability and agility of the proposed algorithm in [77]. It is important to mention that, in the case of LTE and LTE -A, Handover (HO) based traffic shaping is not as feasible approach for LB as it was for CDMA based UMTS due to the reasons explained in section 3.2. This makes majority of the

In OFDMA based WCS like LTE and LTE-A the luxury of soft and softer handover is
existing LB solutions based on this kind of traffic shaping approach less attractive for LTE and LTE-A.

4.2.3 Coverage Adaptation (CA) based Load Balancing

This type of LB schemes rely on mechanisms to change the effective coverage areas of the cells to match the traffic offered within their areas to the resources available to those cells, either through power adaptation, antenna adaptation or a hybrid of the both of these techniques. This approach has been most extensively studied in literature because of its flexibility and effectiveness. In the following subsection we discuss the key works in this category of LB schemes by further classifying them in three sub categories listed above.

4.2.3.1 LB through Antenna Adaptation (AA)

In this type of LB schemes the basic idea is to reduce the coverage area of the over loaded cells either through tilting down the antenna [78] or by changing its radiation pattern [79–87]. Authors in [79–81] consider a scenario with traffic hotspots that cause congestion in some cells and propose a solution that contracts the antenna patterns of those congested cells around those hot spots, whereas neighbouring cells expand their radiation pattern to fill in any coverage gap. All of these three works presented by same authors, assume negotiation among only neighbouring BS to fill in any coverage gap and thus have basic scalability but at cost of the stability as coverage gap may be left uncovered with this limited local cooperation. In [82] same authors attempt to address the coverage gap issue of the same problem by using a bubble oscillation model where air in the adjacent bubbles fills any gap among adjacent bubbles through oscillation of not available and hard hand over usually involve a change of carrier frequency incurring extra complexity and over heads.
4.2. Background and Related Work

bubbles. Authors use the analogy of air for antenna radiation pattern and the uncovered users as the gap between the bubbles, suggesting that all users can be eventually covered through oscillation of antenna radiation pattern. While the use of model is novel and it is realistic in its physical context, oscillation of antenna radiation pattern in WCS can highly increasing handover frequency. This high handover frequency may be manageable in CDMA based WCS where a user can communicate with multiple BS simultaneously (soft handover). However, in OFDMA based WCS where hard handovers are involved, this solution may be a compromise on agility and stability in general. In [84] same authors present the same solution in context WCDMA instead of CDMA.

Authors in [86] propose and evaluate the performance of cooperative coverage algorithm where coverage is dynamically adapted by antenna pattern adaption based on user location information and cooperation among the BS in the network. Although significant improvement in terms of service probability has been reported, the dependency of the proposed scheme on heavy cooperation and extensive user location tracking makes it less scalable and agile respectively.

Authors in [85] propose a radiation pattern adaptation scheme for adaptive sectorization in WCS. The key feature of this scheme is that unlike the scheme in [86], it does not track individual mobile users rather it relies on statistical information. Furthermore it does not require extensive inter site cooperation as it can be implemented on each BS individually that makes it scalable and more pragmatic for WCS. This scheme makes use of spatial information and mobility pattern of the mobile users to depict user geographical distribution statistically over a relatively longer period. The adaptive sectoring problem is formulated as a shortest path problem, where each path corresponds to a particular sector partition, and the partition is weighted by its outage probability. Simulation results show a significant improvement in outage probability can be achieved with this adaptive sectorization methodology. Since the use of empirical or off line stas-
tical information of user distribution has been made in this scheme, the agility of this scheme may only measure up to requirement of very long term dynamics of the WCS making it less pragmatic for online operation to cope with medium term load variation of loads among cells in real time.

Author in [83] compared the impact of cooperative beam shaping with the cooperative tilt adaptation for LB. The tilt based approach though has relatively less margin for performance improvement but it is more pragmatic as it is implementable with conventional widely commercialised antennas [78] that do not have have beam shaping capabilities. The radiation pattern adaptation approach, on the other hand, has been shown to have huge margin for performance improvement but requires smart antennas that can change their radiation pattern electronically. Another limitation of the radiation pattern adaptation based approach is that, an expansion in beamwidth of neighbouring cell is required to fill up the coverage gaps, when an over loaded cells narrows its beamwidth. This expansion in beamwidth of the antenna of the helping cell (i.e. the cell aiding the over loaded neighbouring cell) will cause an inevitable decrease in the gain of its antenna, hence the QoS of users in that aiding cell may deteriorate proportionally.

4.2.3.2 LB through Power Adaptation (PA)

In this type of LB schemes the coverage areas of individual cells are adapted through control of transmission power of the reference signal carrying cell signature that is used by the users for cell association. It is different from the handover parameter control used for traffic shaping discussed in section 4.2.2. Contrary to traffic shaping through forced handovers, coverage adaption by reference signal power control does not necessarily affect the on the ongoing calls, but it effectively changes the coverage area and thus changes the association of all the users in the coverage area. Our work presented in this chapter exploits this approach. Only few work in literature has ventured on this direction [88–93].
Authors in [88] present a *centralised* scheduling algorithm where users are required to switch BS in every time slot for the joint objective of throughput maximisation and LB simultaneously. Authors in [88] further assessed the heavy overhead of this solution that makes it void of scalability and agility, and therefore proposed a modified version of the algorithm as well. To this end, they suggested to separate LB from throughput maximisation. By doing so, the time scale of the BS switching for LB alone, can be increased to several time slots. This is possible because LB requires consideration of only medium term dynamics\(^4\). Whereas throughput optimization needs counter measures for very short term dynamics\(^5\) as well and therefore may require BS switching in every single slot i.e at very short time scale. The proposed solution reduces signalling overhead significantly, making it relatively more scalable, however central control still remains inevitable in this solution.

Authors in [89] determine such a combination of pilot power levels in the network that guarantees full coverage and maximizes the capacity ratio of the bottleneck cell. While the proposed algorithm in [89] has been shown to yield reasonable performance improvement through realistic planning tools, it is more suitable for off line planning of pilot powers, than an online LB scheme executable during the operational phase of the network. The main reason behind this is that it requires long term traffic statistics in the whole system to determine the bottleneck cell, as well as the load in each cell to determine pilot power vector for all cells. Furthermore, it does not explicitly address the effect of soft handover on the capacity of the network that is important in the context of CDMA for online LB operations. The scope of [89] is also effectively limited to CDMA based system.

Only [90–93] consider an OFDMA based systems and propose algorithms for dynamic association or coverage adaptation and are directly relevant to our work.

\(^4\)see figure 1.1

\(^5\)see figure 1.1
Work in [91] is an extended version of [90] by the same authors and uses coverage adaptation or dynamic association, as termed by authors therein, for joint objective of LB and interference avoidance through fractional frequency reuse. This work shows a significant gain in terms of designed utility as indicator of system wide performance. However, the underlying assumption of network wide feedback and channel estimation, at each MS and BS at each scheduling instant and need for a central control entity, makes this solution short of desired level of scalability and agility required for self organization. Authors in [92] also proposed a similar algorithm for joint problem of cell association and channel assignment with objective of LB, that is again purely centralised and thus lacks scalability and agility.

A LB solution for OFDMA based WCS presented in a recent work in [93] is fully scalable as it is can be implementable in a fully distributed fashion. The basic idea is that each BS periodically broadcasts its average load and MS uses this information along with the signal strength to make the decision for cell association. This is contrary to legacy WCS where association decision is made only on the basis of received signal strength. This distributed algorithm has been shown to achieve the globally optimal performance iteratively but with the two crucial assumptions 1) spatial load are temporally stationary and 2) time scale at which BS broadcasts its load is much larger than time scale of call holding. First assumption has implications in the sense that even in given spatial vicinity loads are bound to vary over time in real WCS environment due to mobility etc. The violation of this assumption will directly compromise on stability of the proposed solution [93]. In order to make the second assumption valid, BS have to keep their broadcast time very large, that may be a compromise on the agility of the solution.
4.2.4 Load Balancing with Hybrid Approach

In addition to the work discussed above, a few works have used multiple approaches of LB simultaneously. Authors in [94] presented analysis for the simultaneous use of TS through both call admission control and handover hysteresis control and CA through both AA and PA for general TDMA/FDMA systems. While simulation results report 3 – 11% network wide gain in performance the proposed method lacks scalability because of requirement of central omnipotent control unit that needs to exchange excessive signalling with all users in network for spatial estimation of traffic. Furthermore, for spatial traffic estimations, the proposed method heavily depends on the use of mobile positioning jointly with cell assignment probability maps, generated by the network planning process. This makes it very less agile to measure up with acute spatio temporal dynamics of WCS. Therefore, the proposed schemes may be used as more of an offline design methodology useful during deployment phase, than an online LB mechanism implementable in the operational phase of the WCS. Authors in [95] proposed a similar centralised LB algorithm for CDMA based system that uses AA and PA together. As highlighted by the authors, it is time consuming and hence not agile enough to be used for real time LB. Rather its use for self healing scenario has been proposed by authors. Furthermore, it is important to mention that this algorithm is also centralised and hence lacks scalability.

4.2.5 Load Balancing through Relay Stations

In addition to the approaches discussed above, RS has also been extensively investigated as possible LB tool [76,88,96-107]. This is due to fact that RS can help achieving LB through at least three different means. 1) Through CA by improving coverage and signal quality at point of interest e.g hotspot [102]; 2) Through RA by local or opportunistic reuse of spectrum. [104] 3) Through
TS by relaying or routing traffic from over loaded cells to less loaded cells. The last of the three approaches has been most extensively investigated in literature [96-102,105]. However, most of these works assume CDMA based WCS assuming ad hoc RS operating on an out of band spectrum e.g. ISM band. As highlighted by authors in [100] the realistic performance of such ad hoc systems in terms of dynamic load balancing and load sharing is heavily dependent upon the number of channels available to these ad hoc RS. It is concluded in [100] that for dynamic load balancing the number of such channels required is much more than the number of channels required for load sharing (i.e., for bringing the call blocking probability of a hotspot to 2%). Such out of band RS are not included in scope of our work.

Only authors in [105] and [106] consider RS with inband spectrum and of non adhoc nature for LB in a fully architecture based OFDMA WCS. The main idea in [105] is that all users establish their association with the BS or RS dynamically to maximise a utility function designed to reflect system wide performance. It is further proposed that RS also establish their association with the BS’s dynamically to maximise the same utility. Stability of the system is ensured by confining the reassociation process to one RS and (Mobile Station) MS per time slot to avoid ping pong effect. Substantial improvement in user throughput is reported with the proposed scheme. Although most of the information exchange required is local among neighbouring cells, still a central control unit is required to receive, process and feed back the dynamically changing system wide utility, to and from all MS and RS in the network. This may have an adverse effect on agility of solution in practical system because of the delays incurring from large amount of data processing and its relaying to and from a central unit.

Authors in [106] exploit the idea of dynamics clustering through cooperation among neighbour cells thereby avoiding the need for a central control unit to keep the solution scalable. They introduce a dynamic clustering approach, where
an overloaded cell forms cluster by selecting out of the 6 immediate neighbouring cells those with least load. The traffic to be transferred from the over loaded cell to the other cells in the cluster is determined and transferred for LB. The performance of this algorithm is evaluated for hypothetical scenario where only one central cell is over loaded and more realistic evaluation is indicted as intended future work. It is anticipated that in the more realistic scenario, where multiple over loaded cells may coexist, this dynamics clustering approach might need to be improved. This improvement is required to eliminate stability issues that may arise when the same cells are neighbours to more than one overloaded cell and may result in ping pong effect in the clustering process.

4.2.6 Summary

The need for LB is not newly realised, therefore, a large number of brilliant works have already embarked on this problem in WCS. However it can be inferred from the discussion above, that most of these works have their specific scope limited to legacy WCS. Furthermore, a few works that propose LB solutions that are applicable to OFDMA based WCS are not purposefully designed to be self organising and lack one feature of SO or other. i.e. They compromise on scalability, stability or agility of the solution to achieve the desired objectives as discussed above. To the best of our knowledge, a self organising solution for LB applicable to generic WCS as well as relay enhanced WCS including OFDMA based emerging WCS like LTE and LTE-A is still missing. To this end, in next sections of this chapter we develop a novel LB framework (LB-BSOF) designed with built in SO by applying BSOF to the problem of user satisfaction maximisation in WCS. Next section lays out the main assumption and describes system model over which the LB-BSOF is developed.
4.3 System Model and Assumptions

For analytical treatment of the problem of LB, we consider a generic WCS with \( N \) cells. Each sector is considered as one cell so the set notation introduced in section 3.3 representing system as a set of sectors \( \mathcal{N} \) remains valid such that \(|\mathcal{N}| = N\). These cells can be projected by omni or directional antenna of the macro BS or a RS. For mathematical traceability we assume circuit switching model where even the data services are provided by allocating permanent resources throughout duration of the call by some sort of tunneling mechanism. Since the main objective of this investigation is establish a LB mechanism, we assume a worst case scenario of a lossy system with no queuing in place. i.e. all calls, or data requests that do not find a free channel in their respective cell are considered blocked.

The total traffic in the system is \( T_t \) such that \( T_t = \sum_{n=1}^{N} T_n \) where subscript \( n \) denotes association with \( n^{th} \) cell, and subscript \( t \) denotes total. Total number of available radio resource channels in the system are \( M_t \) such that \( M_t = \sum_{n=1}^{N} M_n \).

The blocking in the \( n^{th} \) cell can be given using Erlang B formula [108]

\[
B_n(M_n, T_n) = \frac{\frac{T_n^{M_n}}{M_n!}}{\sum_{m=0}^{M_n} \frac{T_n^m}{m!}}
\]

(4.1)

The average probability or more precisely the expected blocking in the whole system can be given as

\[
\overline{B}(M_N, T_N) = \frac{1}{T_t} \sum_{n=1}^{N} \left( \frac{\frac{T_n^{M_n}}{M_n!}}{\sum_{m=0}^{M_n} \frac{T_n^m}{m!}} \times T_n \right)
\]

(4.2)

where \( T_N = [T_1, T_2, T_3, \ldots T_N] \) and \( M_N = [M_1, M_2, M_3, \ldots M_N] \) are vectors denoting the traffic and radio resources associated with \( N \) cells. For sake of simplicity of expression we will express the dependency of \( \overline{B} \) on these vectors only where it is necessary, however it is understood that \( \overline{B} \) is function of these vectors.
4.4 Biomimetic SO Framework for Load Balancing (LB-BSOF)

In this section we follow the steps of BSOF explained in section 2.7.3 to design a SO solution for the problem of users' low satisfaction rising from the unbalanced load in WCS.

4.4.1 Problem Formulation: Identifying SO-Objective

Our problem can be described as follows: for given total traffic and radio resources in the system, system should self organise such that the user satisfaction is optimal for that traffic and resources available. User are unsatisfied due to either hard or soft blocking. By hard blocking we mean blocking due to unavailability of free channels, whereas soft blocking means although free channels are available but interference on those channels is too high to achieve the lowest required QoS and hence the attempted call is rejected. TO-BSOF in chapter 3 focused on maximising spectral efficiency by reducing the interference and thus can help in decreasing soft blocking as well. Here we focus specifically on hard blocking. Hence, our SO-Objective here is to minimise the system wide average hard blocking for given total traffic and radio resources in the system, i.e.

$$\min_{T^*_N, M^*_N} \hat{B}(M_N, T_N) ; \quad n = 1, 2, ..., N$$  \hspace{1cm} (4.3)

subject to:

$$T_t = \sum_{n=1}^{N} T_n$$  \hspace{1cm} (4.4)

and

$$M_t = \sum_{n=1}^{N} M_n$$  \hspace{1cm} (4.5)
This SO-Objective is not only a system wide holistic objective, but it is also large scale optimization problem that is seemingly unscalable, as it requires achieving the right traffic density and amount of resources in each of the \( N \) cell. First we need to reduce the complexity of this problem, i.e. we need to determine a SO-Goal for this objective, just like SO-Objective of flight efficiency maximization is simplified into SO-Goal of maintaining a V-shape (see section 2.7).

### 4.4.2 Transforming SO-Objective into SO-Goal

With the aim to determine the optimal traffic and resource distribution among all the cells in WCS for minimal system wide average blocking, we start with a simpler scenario, where each cell has same fixed amount of resources i.e. \( M = \tilde{M} = \frac{\tilde{M}}{N} \), and aim to determine the right radio resource distribution among the cells in the system. In this case the average blocking will be given as:

\[
\bar{B} = \frac{1}{T} \sum_{n=1}^{N} \left( \frac{\frac{T_n}{M_i}}{\frac{M_i}{M}} \times T_n \right)
\]  

(4.6)

To proceed further with our analysis, we put forth following set of propositions that will help us transform SO-Objective into a simpler SO-Goal.

**Proposition 6.** In a cellular system with given total traffic load and same number of radio channels in each cell, the average blocking will be minimum if each cell has same amount of traffic load. Mathematically, for minimum average blocking in such system:

\[
\Rightarrow T_1 = T_2 = \ldots = T_N = \frac{T}{N} = \bar{T}
\]

(4.7)

**Proof.** This proposition can be proved by solving for the minima of the average blocking function in (4.6). A detailed proof is given in appendix B.  

\(^6\text{see Appendix D for relevant information on optimization theory}\)
Proposition (6) provides a simple condition for minimizing the blocking in a WCS with given total traffic and radio resources. Equation (4.7) can be used as SO-Goal for minimum blocking in cellular system that have *same number of radio resources in each cell*. However, this assumption is only true for legacy WCS in their early deployment stages or most probably in rural areas as resource allocation per cell in that phase is usually symmetric among all cells. In emerging WCS cells might have different radio resources available in them due to their different size and user densities. Furthermore, in future WCS like, LTE-A, the advent of RS and Femto cells means that the assumption of having same radio resources in each cell will not remain practical any more. To obtain more pragmatic SO-Goal for such practical systems, we present following proposition.

**Proposition 7.** In a cellular system of $N$ cells with given total traffic $T_i$ and radio channels $M_i$, the average blocking will be minimum if

$$\frac{T_i}{T_j} = \frac{M_i + 1}{M_j + 1}; \forall i \neq j, \text{ and } i, j = 1, 2, 3 \ldots N$$

(4.8)

**Proof.** Proof of this proposition is similar to proof of proposition 6 and is therefore skipped for brevity. □

Equation (4.8) is our potential SO-Goal for more generic WCS with heterogeneous cell sizes and radio resources. In other words in order to minimise average blocking in a system with $N$ cells, $N \times N$ conditions like the one in (4.8) need to be maintained across whole system dynamically such that $T_i = \sum_{n=1}^{N} T_n$ and $M_i = \sum_{n=1}^{N} M_n$. To increase scalability and agility of this solution i.e. to get a simpler SO-Goal we present and prove following prepositions.

**Proposition 8.** In a cellular system of $N$ cells with given traffic offered in each cell and total radio resources $M_n$, the average blocking is minimum if the radio resources allocated to each of the $N$ cell are such that

$$M_n = \frac{a \times M_i - \sum_{i \in N \setminus n} (T_i - T_n)}{N \times a}, \forall n \in N$$

(4.9)
where $T_i$ and $T_n$ denote traffic in $i^{th}$ and $n^{th}$ cell respectively.

Proof. From proposition 7 we know that in a WCS of $N$ cells with given total traffic $T_t$ and radio channels $M_t$, the average blocking is minimum if

$$T_n = M_n + 1, \forall n \in N$$

(4.10)

we can write above set of equations in standard form of linear equation

$$T_n = aM_n + b, \forall n \in N$$

(4.11)

where $a = b = 1$. In order to obtain minimum average blocking, $N$ such equations need to be maintained in WCS, one for each cell such that $T_t = \sum_{n=1}^{N} T_n$ and $M_t = \sum_{n=1}^{N} M_n$. In order to solve this system of linear equations to determine the optimal radio resources to be allocated or traffic to be offered to each cell, we use basic elimination method and by subtracting the equation for one cell from the other we proceed as follows

$$T_i - T_j = a(M_i - M_j) \quad ; \quad i, j \in N$$

$$M_i = \frac{(T_i - T_j)}{a} + M_j \quad ; \quad i, j \in N$$

(4.12)

As

$$M_1 + M_2 + M_3 + ... + M_N = M_t$$

(4.13)

Using (4.12) in (4.13) to solve for $M_1$

$$M_1 + \left(\frac{(T_2 - T_1)}{a} + M_1\right) + \left(\frac{(T_3 - T_1)}{a} + M_1\right) + ... + \left(\frac{(T_N - T_1)}{a} + M_1\right) = M_t$$

(4.14)

$$aN \times M_1 + (T_2 - T_1) + (T_3 - T_1) + ... + (T_N - T_1) = aM_t$$

(4.15)

$$aN \times M_1 + \sum_{n \in N \setminus 1} (T_i - T_1) = aM_t$$

(4.16)

$$M_1 = \frac{a \times M_t - \sum_{n \in N \setminus 1} (T_i - T_1)}{N \times a}$$
similarly solving for $M_2$, $M_3$ and so on we can obtain a general formula to calculate optimal resources for minimum blocking i.e.

$$M_n = \frac{a \times M_e - \sum_{i \in N/n} (T_i - T_n)}{N \times a}, \forall \ n \in N$$

From proposition 8 we also derive following important proposition:

**Proposition 9.** For given total resources $M_e$ and given traffic distribution for $N$ cells, radio channels can be optimally allocated to all cells for minimum average blocking only if

$$\sum_{i \in N/n} (T_i - T_n) \leq M_e \ orall \ n \in N \quad (4.17)$$

**Proof.** The proof follows directly from proposition 8. Since radio resources allocated to cell cannot be negative, so the numerator in (4.9) should be always be positive hence the proposition. □

In other words proposition 9 states that with given $M_e$, we can not obtain optimal load balancing for unboundedly different traffic distributions among the cells. i.e. For larger traffic variations among cells larger total number of radio resources in WCS are required to obtain optimal minimum blocking.

Given the condition in proposition (4.17) is met, proposition 8 can be applied and minimum blocking can be achieved by LB through dynamic resource allocation among cells by satisfying proposition 8. However, dynamics resource allocation is not always best approach because of its complexity and need for central control. Rather, distributing the actual traffic load among cells through various procedures, like handover, power control, tilting or relaying is another way to reach the same optimality as discussed in section 4.2. Here we present a proposition that can help us to implement these alternative approaches other than dynamic resource allocation towards optimal LB.
Proposition 10. In a cellular system of $N$ cells with given radio resources allocated to each cell and total traffic $T_t$, the average blocking is minimum if traffic in each cell is such that

$$T_n = \frac{T_t - a \sum_{i \in \mathcal{N}} (M_i - M_n)}{N}, \quad \forall \ n \in \mathcal{N} \tag{4.18}$$

Proof. From proposition 7 we know that in cellular system of $N$ cells with given total traffic $T_t$ and radio channels $M_i$, the average blocking is minimum if:

$$T_n = M_n + 1, \quad \forall n \in \mathcal{N} \tag{4.19}$$

we can write above set of equations as:

$$T_n = aM_n + b, \quad \forall n \in \mathcal{N} \tag{4.20}$$

In order to solve this system of linear equations, to determine the optimal radio resources to be allocated or traffic to be offered to each cell, we use basic elimination method. By subtracting the equation for one cell from the other we proceed as follows:

$$T_i - T_j = a(M_i - M_j), \quad i, j \in \mathcal{N}$$

$$T_i = a(M_i - M_j) + T_j, \quad i, j \in \mathcal{N} \tag{4.21}$$

As

$$T_1 + T_2 + T_3 + \ldots + T_N = T_t \tag{4.22}$$

Using (4.21) in (4.22) to solve for $T_1$

$$T_1 + (a(M_2 - M_1) + T_1) + (a(M_3 - M_1) + T_1) + \ldots + (a(M_N - M_1) + M_1) = T_t \tag{4.23}$$

$$\frac{N \times T_1}{a} + (M_2 - M_1) + (M_3 - M_1) + \ldots + (M_N - M_1) = \frac{T_t}{a} \tag{4.24}$$

$$\frac{N \times T_1}{a} + \sum_{i \in \mathcal{N}/1} (M_i - M_1) = \frac{T_t}{a} \tag{4.25}$$

$$T_1 = \frac{T_t - a \sum_{i \in \mathcal{N}/1} (M_i - M_1)}{N}$$
simply solving for $M_2, M_3$ and so on we can obtain a general formula to calculate optimal traffic in each cell for minimum system wide average blocking i.e.

$$T_n = \frac{T_t - a \sum_{\forall n \in N} (M_i - M_n)}{N}, \quad \forall \quad n \in N$$

From proposition 10 we can derive following important inference.

**Proposition 11.** For given total traffic $T_t$ optimal blocking can be achieved if the difference among the radio resources allocated to different cells is such that

$$a \sum_{\forall n \in N} (M_i - M_n) \leq T_t, \quad \forall \quad n \in N$$

where $M_i$ denotes traffic in the $i^{th}$ cell.

**Proof.** The proof follows directly form proposition 10. Since traffic in a cell cannot be negative, so the numerator in (4.18) should be always be positive hence proposition.

Proposition (11) is a dual of proposition (9) and has same implications as proposition (11). More specifically, proposition (11) states that with given $T_t$ we can not obtain optimal load balancing for unboundedly different resource allocation among the cells.

In other words, proposition (9) highlights the limit of *dynamic resource allocation* based LB. Whereas, proposition (11) puts forth the limit of *traffic shaping* based load balancing. Both of these constraints will be considered when designing SO-Functions in next sections. Here we can summarise the findings of this section as follows:

Equation (4.8) was obtained as solution to the problem of system wide average hard blocking minimization as described in (4.3) for any generic WCS. This solution was not worthy of being an SO-Goal because of its high complexity so...
Chapter 4. Load Balancing through SO of Coverage

it is further simplified to obtain suitable SO-Goal that has potential for better scalability and agility enabled by low computational complexity. This SO-Goal has two forms given in (4.9) and (4.18) (the use of the both SO-Goals will be explained in next section 4.4.5) and can be achieved through many possible ways as discussed in detail in section 4.2.

However, this SO-Goal, instead of its flexibility of the ways in which it can be achieved and very low implementation complexity, is still a solution that requires global or central coordination and control. So in order to bring full scalability and agility we follow the next step of BSOF and decompose this global SO-Goal into suitable local SO-Functions.

4.4.3 Decomposing SO-Goal into SO-Functions

In order to decompose the global SO-Goal into local SO-Functions we introduce a notion of super cell (see figure 4.1). Super cell is a virtual cell that consists of $S$ underlying cells. Using the concept of super cell, the global SO-Goals presented in (4.9) and (4.18) can be decomposed into local SO-Functions. i.e. each super cell can aim to achieve the SO-Goals (4.9) and (4.18) locally. This can be done by assuming that each super cell itself is the whole system. We present following proposition to further clarify the advantage of the concept of super cell.

**Proposition 12.** For given total traffic and resources in a group of $S$ cells, the blocking probability in the super cell of those $S$ cells is lower than average blocking probability in the group of $S$ cells, even if the load among those cells is optimally balanced for minimal blocking. i.e.

$$\bar{B}_{s,\min}(M_s, T_s) > B(M_s, T_s) \quad \forall \quad 1 < S \leq N$$

(4.27)

where subscript $s$ denotes super cell such that $M_s = \sum_{n=1}^{S} M_n$ and $T_s = \sum_{n=1}^{S} T_n$ and $\bar{B}_{s,\min}$ is minimum average blocking probability in a group of $S$ cells that can
be archived with optimal LB for given radio resources and traffic in that group of cells. \( M_s \) and \( T_s \) are vectors of radio resources and traffic in each of the cells in the super cell.

**Proof.** First we prove the above proposition for a simpler scenario where same number of resources are dedicated to each cell in the group of \( S \) cells. In that case through proposition 6 we know average blocking in group of \( S \) cells will be minimum if \( T_1 = T_2 = T_3 = T = \frac{T_s}{S} \). Therefore, for perfectly balanced load, the minimum average blocking in this case is same as blocking in any of the \( S \) cells and can be given as

\[
B_{s, \text{min}}(M, T) = \frac{T^M}{M!} \sum_{m=0}^{M} \left( \frac{T^m}{m!} \right)
\]

For sake of simplicity of expression we drop the subscript \( \text{min} \) onward.

The blocking in the super cell on the other hand would be:

\[
B(SM, ST) = \frac{g^{SM} M!}{\sum_{m=0}^{SM} \left( \frac{T^m}{m!} \right)}
\]

Where \( SM = S \times M \), and \( ST = S \times T \), in order to prove \( B(SM, ST) < B_s(M, T) \) we prove that \( \frac{1}{B(SM, ST)} > \frac{1}{B_s(M, T)} \) consider

\[
\frac{1}{B_s(M, T)} = \frac{M! \sum_{m=0}^{M} \left( \frac{T^m}{m!} \right)}{T^M}
\]

\[
\frac{1}{B_s(M, T)} = M! \sum_{m=0}^{M} \left( \frac{T^m - M}{m!} \right)
\]

\[
\frac{1}{B_s(M, T)} = \sum_{m=0}^{M} \left( \frac{M! T^m - M}{m!} \right)
\]

Let \( p = M - m \); then

\[
\frac{1}{B_s(M, T)} = \sum_{m=0}^{M} \left( \frac{M(M-1)(M-2)\ldots(M-p+1)m!}{m! T^{M-m}} \right)
\]
\[
\frac{1}{B_s(M,s)} = \sum_{p=0}^{M} \left( \frac{M(M-1)...(M-p+1)}{T^p} \right)
\]

Similarly,

\[
\frac{1}{B_s(M_s,T_s)} = \sum_{p=0}^{S_M} \left( \frac{S_M(S_M-1)...(S_M-p+1)}{S^p} \right)
\]

(4.30)

Since all terms in (4.30) and (4.31) are positive hence proved that \(\frac{1}{B_s(M_s,T_s)} > \frac{1}{B_s(M,T)} \Rightarrow B(M_s,T_s) < B(M,T)\). Similarly for the case where cells in the group have different resources dedicated to them, following the same steps as above it can be shown that, \(B(M_s,T_s) < B(M_n,T_n)\). \(\square\)

Since proposition 12 proves that for a given amount of total traffic and radio resources a super cell yields even higher user satisfaction by having lower blocking compared a group of cells having same total traffic and resources, with fixed dedicated resources to each cell. To achieve the advantage of super cell, brought forth in proposition 12 all the cells within same super cell need to use the whole spectrum available to all cells in the super cell. i.e. A frequency reuse of \(1\) within a super cell can yield even lower blocking then optimal LB. In addition to this advantage of super cell, we propose following use of supper cell to decompose the global SO-Goals into local SO-Functions : we propose to divide the whole WCS in virtual super cells and meet the load balancing conditions (4.9) and (4.18) within in each super independently rather than in the whole system as whole. That is, the SO-Goals (4.9) and (4.18) should be achieved within each super

\(^7\)All future WCS are also aiming for frequency reuse of 1

cell independently making them effectively *SO-Functions*. This localisation of (4.9) and (4.18) will be further explained in next section 4.4.5. Here it important to highlight following main motivations behind the proposed super cell concept:

The main advantages of the proposed concept of super cell is that it will bring scalability and agility into the solutions in (4.9) and (4.18). These two characteristics of SO are achieved by reducing the complexity and spatial scope of the solutions in (4.9) and (4.18). Less signalling and processing will be required to implement the solution locally and independently for small number of cells in the super cell compared to that required for implementing it on system wide scale. The second advantage of this concept is that, as proved in proposition 12 each super cell will have better user satisfaction than that could be achieved through perfect load balancing among the cells in the super cell. In next section we investigate the actual WCS actuators and functions that can be executed to preform these SO-Functions.

### 4.4.4 Design Considerations for SO-Functions and Practical Implementation

An important step in designing a SO solution is identification of appropriate actuators to execute SO-Functions under the principle of coopetition to achieve the designated SO-Goal. For this we need to identify the specific capabilities of individual entities (these entities are super cells in our case) of the system that are executable independently within the local scope of these individual entities. Although there are number of actuators that can be performed to achieve LB as discussed in detail in section 4.2, only those actuators which bear specific characteristics (see section 2.7.2.2) to qualify as SO-Functions are suitable candidates to achieve SO. It is clear that the cells within the coverage of same site i.e. cells projected from same BS and its associated RS's can observe each other's
state with minimal signaling cost. This is because each sector or cell projected from same site is physically co-located. Even though RS are not co-located with their donor BS but they are connected to its donor BS through an over the air link. Therefore, we proposed to define \textit{super cell as group of all cells formed by sectors projected from the same BS and RS associated to that BS}. In addition to advantages of co-existance, another reason to choose this entity as super cell is that, unlike some prior works it does not require dynamic coordination for cluster making and thus promises perfect scalability and agility. By avoiding the need for dynamic clustering the vulnerability to oscillations is also avoided thus ensuring stability. Since the design of actual actuators for executions of SO-Functions in this case is dependent on the Deployment Architecture (DA) of WCS, therefore first we describe an exemplary DA below:

\subsection*{4.4.4.1 Exemplary Deployment Architecture for Illustration of LB-BSOF}

A heterogenous WCS i.e. containing both Macro cells projected by BS and micro/pico/femto cells projected by small RS as shown in figure 4.1 is considered in this study to demonstrate the concepts. Each BS has certain number(6 in this case) of sectors/cells represented by red colored area. Each cell has certain number of RS strategically positioned either to cover a dead zone area or fixed traffic hot spot or generally at the edge of two adjacent sectors where the coverage by BS is minimum. All cells and RS associated to same BS form a super cell as shown in figure 4.1. Each RS has one omni-directional antenna to provide \textit{coverage links} to MSs. The coverage areas of RS are shown by green colored area in figure 4.1. In this particular instance of WCS considered, there are three RS per BS and all are systematically located on edge of cell between sectors. The BS to which RS are attached is called its \textit{donor BS}. Each RS is installed with one highly directional antenna with switchable beam direction to establish a back haul link called

Figure 4.1: Exemplary deployment architecture for illustration of LB-BSOF. Small green circles show RS based micro/pico/femto cells. Whereas each elliptical shape shows a cell projected by the BS access link with its donor BS. An OFDM based air interface similar to LTE is considered. The radio resources between access link and coverage link are shared using time division. Access link part of frame is further divided in time, among the RS attached to the same BS.

4.4.5 Using Two SO-Goals to Achieve Same SO-Objective

As discussed above, two different SO-Goals have been obtained as solution of the SO-Objective in (4.3), through proposition 8 and 10. In fact both of these solutions complement each other by being useful in different scenarios as we explain below.
From proposition 8 and its subsequent propositions, it can be anticipated that with limited total radio resources, optimal minimum blocking cannot be achieved by mere dynamic resource allocation among cells. This approach may fail in cases where large difference in traffic among the cells occur as highlighted by Proposition 9. In such cases the solution is given by proposition 10 that provides method to optimally balance traffic among cells for given resource allocation. So, proposition 8 and 10 provide two alternative pragmatic ways to achieve same SO-Objective through two different SO-Goals. Both of these SO-Goals have their own advantages and limitations. For example, shedding of extra load from the over loaded cell i.e. traffic shaping (SO-Goal obtained by proposition 10) is easy to implement in a distributed way, compared to dynamic resource allocation among cells (SO-Goal obtained by proposition 8). However, for access link traffic shaping is not feasible, as users are already accepted on the coverage link and must be catered for on the access link to avoid wastage of coverage link resources. Therefore, in order to achieve the SO-Objective in (4.3) for access link we propose to use the SO-Goal in (4.9) i.e. dynamic resource allocation based LB, and for coverage link we propose to use the SO-Goal (4.18) i.e dynamic traffic shaping based LB. This means the SO-Functions to achieve these SO-Goals should include both kind of actuators i.e. traffic shaping based as well as dynamic resource allocation based. In other words, there should be SO-Functions to adjust resource allocation among access links (according to proposition 8) and there should be SO-Functions to transfer traffic to other cells (according to proposition 10). By exploiting the specific deployment architecture of relay enhanced WCS like LTE-A e.g the one explained in figure 4.1, we propose the following SO-Functions to achieve desired SO-Goal given in (4.8).

4.4.5.1 Coverage Adaptation at BS (BCA)

BS in each super cell maintains its coverage such that traffic in each of its cells satisfies (4.18). When applied at super cell level, (4.18) can be written as:

\[ T_n = \frac{T_s - \alpha \sum_{i \in N_s} (M_i - M_n)}{N_s}, \quad \forall \ n \in N_s \]  

(4.32)

Where subscript \( s \) denotes super cell and thus \( N_s \) is set of cells in the super cell such that \( |N_s| = N_s \). This coverage adaptation can be performed by any of the means discussed in section 4.2.3 though reference signal power control based coverage adaptation is relatively practically advantageous as discussed in that section.

4.4.5.2 Coverage Adaptation at RS (RCA)

BS in each super cell commands its RSs to maintain their coverage such that the traffic in each of the coverage areas of that satisfies (4.32).

4.4.5.3 Access Link Adaptation (ALA)

The condition in (4.17) when applied at the super cell level for the access links, can be written as:

\[ \sum_{i \in N_s^A} \frac{T_i^A - T_n^A}{a} \leq M_s^A, \quad \forall \ n \in N_s^A \]  

(4.33)

Where superscript \( A \) denotes access link and thus \( N_s^A \) is a set of access links in the super cell that is equal to number of RS in a super cell; \( M_s^A \) is total number of radio resources in all access links in the super cell. If condition (4.33) is true BS in each super cell maintains optimal resource allocation among the access links of its three RS according to (4.9) that at the super cell level can be written as:

\[ M_n^A = \frac{a \times M_s^A - \sum_{i \in N_s^A} (T_i^A - T_n^A)}{N_s^A \times a}, \quad \forall \ n \in N_s^A \]  

(4.34)
In other words BS increases the amount of radio resources allocated to the access link of a RSs when its overloaded, by reducing the radio resources allocated to access link of other RS in same super cell. The quantities of this increment or reduction in resources allocated to the \( n^{th} \) access link in super cell are determined through (4.34). The principle of coopetition requires that ALA should happen only when there is a RS in the same super cell with under loaded access link, otherwise, this SO-Function will not be executed. i.e. RS will cooperate in a super cell but only when they do not have to sacrifice their own users.

4.4.5.4 Beam Switching of RS (BSR)

As mentioned above, the SO-Function ALA can not be helpful if condition in (4.33) is not true. Or in more simple words when access link radio resources in super cell run shorter than certain threshold required for optimal LB for given traffic on the access link in the super cell, ALA can not help.

Similarly the SO-Functions BCA and RCS can not be applied when (4.26) is violated. i.e. when at super cell level following condition is not true:

\[
a \sum_{v \in N/n} (M_i - M_v) \leq T \quad \forall \quad n \in N
\]  

(4.35)

Or in simple words when total traffic in super cell is too low to perform optimal traffic shaping within the super cell.

For such scenario we propose a SO-Function in which a RS in a super cell changes its donor BS and gets associated to another BS in the neighbouring cell. This SO-Function can cope with both of the above situations as follows:

In first case, i.e. for over loaded scenario the over loaded super cell makes one of its RS switch the donor BS station to one of the neighbouring super cells. This effectively increases available resources to cover the traffic in the same geographic
area of the overloaded super cell, i.e. increases $M^A$ in (4.33) for that overloaded super cell to make (4.33) true.

In second case, i.e. scenario with too low traffic $T_\gamma$ to satisfy condition (4.35), super cell requests one of its neighbour super cell to switch donor BS of one or more of its RS's to itself. Now the traffic in the coverage area of the newly obtained RS that geographically lies in the neighbouring super cell will be served by this under loaded super cell. Thus the effective $T_\gamma$ in that super cell increases and condition (4.35) becomes true.

Physically this SO-Function means Beam Switching at RS (BSR) which is easily possible because directional antenna on RS used to establish access link with the donor BS can have beam switching capability. The proposed DA (figure 4.1) of WCS makes this functionality even easier to achieve with minimal cost because there are only a small number of (four in this instance of WCS) predetermined directions for each RS in which it might have to switch its beam. Therefore, for this DA, the SO-Function BSR can be implemented by less expensive fixed beam switching antennas with only four predefined radiation patterns, instead of the fully adaptive smart antennas that are relatively expensive.

### 4.4.6 Recap of LB-BSOF

Having identified the SO-Objective, SO-Goals and SO-Functions the over all LB-BSOF can be recap in figure 4.2. The complex optimization problem i.e. SO-Objective of blocking minimization over amount of resources and traffic in a multi cell environment is first solved analytically. This solution is further simplified semi analytically to develop low complexity SO-Goals. These SO-Goals are further decomposed into local SO-Functions by introducing concept of supper cell and proving its advantage. Four potential SO-Functions BCA, RCA, ALA and RBS are proposed. Practical actuators for execution of those SO-Function are
Chapter 4. Load Balancing through SO of Coverage

Operation of SO System

SO-Goal in turn manifests SO-Objective(s)

Global SO-Objective transformed into a much simpler, though still global but locally achieve code SO-Goal

On coverage links traffic should be shaped according to:

\[ T_i = \frac{M_i + 1}{M_j + 1}; \forall i \neq j, \text{ and } i, j = 1, 2, 3...N \]

On the access links resources should be allocated according to:

\[ M_a = \frac{u \times M_i - \sum_{i=1}^{N} (T_i - T_{a})}{N \times a} \]

SO-Goal: in all cells in system following conditions are maintained

\[ \frac{T_i}{T_j} = \frac{M_i + 1}{M_j + 1}; \forall i \neq j, \text{ and } i, j = 1, 2, 3...N \]

SO-GOal (s) decomposed into basic elementary functionalities that each super cell can execute them independently or semi independently

SO-Objective: \( \min_{T_x, M_x} B(M_N; T_N) \); \( n = 1, 2,..., N \)

Subject to:

\[ T_i = \sum_{n=1}^{N} T_{n} \]

\[ M_i = \sum_{n=1}^{N} M_{n} \]

Figure 4.2: Recap of LB-BSOF
4.5. Operational Use Cases of LB-BSOF

In this section we present the potential use cases of LB-BSOF that explain its operation by dividing it into three hierarchical levels.

4.5.1 SO at Micro level

Micro level SO will use appropriate SO-Functions to achieve SO-Goal at small spatial scope i.e. load balancing within a super cell in this case. There can be two cases of imbalanced traffic distribution within a super cell,

1. BS in the super cell is overloaded in one or more of its cells.

2. One or more RS are overloaded on coverage link.

In first case, BCA will be invoked to achieve optimal minimum blocking in a super cell as explained in section 4.4.5.1. In second case, SO-Function RCA will be used to resolve the situation as explained in section 4.4.5.2.

4.5.2 SO at Macro level

Macro level SO will use appropriate SO-Functions to achieve SO-Goal at relatively larger spatial scope i.e. load balancing across the super cells at system wide scale. It is required in addition to micro level SO explained in section 4.5.1, because there can be scenarios where the traffic in the coverage area of a super cell grows to
an extent that it is larger than the LB potential of the super cell as determined by (4.17). That condition applied at super cell level instead of system level, becomes:

$$\sum_{m \in N_s/n} (T_m - T_n) \leq M_s \quad \forall \ n \in N_s$$  

(4.36)

If the traffic distribution is such that the condition in (4.36) is violated, Micro level SO explained in last section 4.5.1 will not be able to handle it. In such a scenario, the SO-Functions ALA and BSR will be invoked. That is, one or more of the RSs will change their serving BS to shift their load to a neighbour super cell with lowest load, by switching the direction of their access link beam. By doing so, the RS can not only transfer the traffic load from their own coverage area to under loaded neighbouring super cell, but also can alleviate congestion on originally serving super cell by taking no radio resources from it. Thus more radio resources will be available for the remaining RSs and BS cells in that super cell.

The time scale at which these SO-Functions ALA and BSA are invoked for Macro level SO, is anticipated to be much larger compared to that of SO-Functions used for Micro level SO. Therefore, rate of inter super cell signalling required to execute ALA and BSA can be kept low. Furthermore, since RSs are located at edges of the super cells, so they are as close to neighbouring super cells as they are to their donor BS itself. A good channel quality means even less bandwidth will be required for signalling between the neighbouring BS and the RS. Furthermore it also means that no deterioration in the quality of channel of the access link is anticipated with SO-Function BSR. It is to be noted that, here the principle of coopetition would require that each super cell should give priority to its own RS, and only after satisfying its own local demand it will accept traffic load from the RSs of the neighbouring super cells.
4.5.3 SO at Global Level

LB-BSOF can play a threefold role for LB at very large spatial and temporal scale as explained below.

4.5.3.1 Load Balancing Across Operators

There can be a scenario when operator X’s network is over loaded to an extent that both micro and macro level SO cannot resolve the issue completely for the reasons determined in previous sections. i.e. when (4.17) is violated which can be written for operator Y as follows:

\[
\sum_{m \in N^X} \left( T_t^X - T_n^X \right) \leq M_t^X \quad \forall \quad n \in \mathcal{N}^Y
\]  

(4.37)

where superscript X denotes association with operator X. In this case SO-Functions BCA, RSA, ALA and BSR can be invoked iteratively in such a manner that they can transfer the extra load in operator X’s network to appropriate BS or RS adjacent to another operator Y’s network. The RS in the cell adjacent Operator X’s network can then transfer the load to operator Y’s network using SO-Function BSR thus achieving inter operator load balancing. Again, of course the lending operator will lend its radio resources only if it has some free resources after serving its own users i.e. relationship among operators has to be based on coopetition to yield an overall progressive outcome as explained in section 2.7.2.3. Here coopetition plays its role in favour of both of the operators; the borrowing operator is helped to maintain its service even in overloaded conditions, while the lending operator can generate revenue from its resources which were otherwise going to be wasted.
4.5.3.2 Load Balancing Across Infrastructures

LB-BSOF can also be used for load balancing across infrastructures. An example of across-infrastructure load balancing can be transfer of traffic from the terrestrial WCS to an HAP (High Altitude platform) [109,110] or satellite based backbone link. Similar to the scenario discussed in previous section 4.5.3.1, both micro and macro level SO jointly can transfer some load in given over loaded infrastructure to a pre-defined node in that infrastructure which can again call SO-Function BSR to transfer load to an aerial platform based link either belonging to HAP or a satellite. This will not require any extra hardware or power requirements on users end making it seamless to them.

4.5.3.3 Self Healing

Proposed LB-BSOF framework has potential for self healing features as well. For example, in case a RS or BS cell fails, SO-Functions BCA and RCA can be called to in neighbouring RS and cells to fill the coverage gap. At the same time SO-Function ALA can be called to allow the rest of the RSs to occupy the spectrum allocated to that RS, thus minimising the effect of node failure. In case an entire BS fails, SO-Function RCA is called on all the RSs in that area. This will increase the coverage area of RSs in the affected area to at least partially fill the coverage gap caused by BS failure. Then SO-Function BSR will be called to link the RS associated to failing BS to a suitable neighbour BS.

While the use cases of LB-BSOF explained in previous three subsections highlight its extreme potential, demonstrating all these use cases through simulation is not viable for scope of this thesis. We will confine our further discussion, evaluation and demonstration of LB-BSOF to the micro level SO only in the remaining sections of this chapter.
4.5.4 A Heuristic Algorithm for Execution of LB-BSOF

In order to illustrate the practical implementation of LB-BSOF a heuristic algorithm is developed and illustrated in figure 4.3. It can be seen that local state of information at the level of super cell are required at most for operation of this algorithm, minimizing the signaling overhead. The self-organising and autonomous nature of the LB-BSOF is also clear here as no central or external control is required [4]. The purpose of this detailed algorithm using all the four SO-Functions identified above, is to illustrate the full scale practical implementation of LB-BSOF. For the demonstration of concepts in next section we use simpler version of this algorithm using only two SO-Functions i.e. RCA and BCA out of the four because of the limitations of the simulation platform.

4.6 Simulation Results

In this section performance of LB-BSOF is evaluated for a number of different scenarios and and DA’s thorough system level simulations.

4.6.1 Some key Assumptions

Because of the very large scale of simulator model required to demonstrate use cases presented in sections 4.5.2 and 4.5.3, the simulation results in this section demonstrate SO at Micro level only, as explained in section 4.5.1. This means the SO-Functions namely BCA and RCA are implemented only that aim for one out of the two SO-Goals i.e. one given in (4.32).

The algorithm essentially remains the same as shown in figure 4.3 with difference that LB on coverage link is only performed i.e. only one SO-Goal shown in (4.18) as indicated by left branch of LB-BSOF in figures 4.3 and 4.2 is considered. this
Figure 4.3: A heuristic algorithm for execution of LB-BSOF for continuous optimization of user satisfaction.

means LB through traffic shaping is performed only. Since no SO-Function to achieve dynamic resource allocation based alternative SO-Goal in (4.9) is used therefore results for scenario for which condition (4.35) is true are presented only. As expected, for scenarios were (4.35) is not true, other two SO-Functions ALA and BSR are required to achieve SO-Goal in (4.9) and only BCA and RCA based LB-BSOF implementation, can not and has been observed to not yield any reduction in blocking and hence those results are omitted for sake of brevity.  

We assume a scenario where all cells have the same amount of radio resources then (4.32)

\[ M^a = \frac{a \times M^a - \sum_{k=1}^{n} (T_k - T^a)}{M^a} \]

\[ M^b = \frac{a \times M^b - \sum_{k=1}^{n} (T_k - T^b)}{M^b} \]

\[ T^a = \frac{a \times \sum_{k=1}^{n} (M^a_k - M_k)}{N_a} \]

\[ T^b = \frac{a \times \sum_{k=1}^{n} (M^b_k - M_k)}{N_b} \]

\[ T = \frac{\sum_{k=1}^{n} (M^a_k - M_k) + \sum_{k=1}^{n} (M^b_k - M_k)}{2} \]
becomes $T_n = \frac{\gamma_n}{K_n}$, $\forall n \in N$. We assume that each user produces same amount of traffic. Thus (4.32) becomes $K_n = \frac{K_s}{K_n}$ where $K_n$ and $K_s$ represents number of users within $n^{th}$ cell and a super cell respectively.

Both BCA and RCA are implemented through power control of reference signal that is measured by users to decide cell associations and thus $K_n = \frac{K_s}{K_n}$ is maintained in each super cell.

Performance is evaluated in terms of Hard Blocking (HB) i.e. percentage of rejected calls due to unavailability of free data channels. Impact of proposed algorithm is also evaluated on Soft Blocking (SB) as well i.e. percentage of rejected calls due to high interference on the available channels.

The performance of LB-BSOF is compared with scenario with no LB in place. To establish a fair benchmark we also compare performance of LB-BSOF with that of Ideal Central Control (ICC) based LB. ICC shows the minimal possible HB for given total radio resources and traffic in the system by perfect system wide LB according to (4.18) unlike LB-BSOF that uses (4.32) for LB within each super cell in the system independently. ICC is based on the hypothetical assumption that each sector in the WCS knows the load of all other sectors in the WCS and will adjust its coverage based on that global knowledge and intelligence. Thus ICC is neither scalable, nor agile hence lacks the two important features of SO. The purpose of evaluating ICC is to determine the absolute minimum to which blocking can be reduced for given radio resources per cell and total traffic in the system. Therefore, the HB observed with ICC will be some times referred to as absolute minimum HB and provides benchmark to compare the performance of super cell based distributed and self organising LB i.e. LB-BSOF.
Parameter | Value | Parameter | Value
---|---|---|---
System Parameters | Access link Parameters
Average call holding time (exponential) | 120s | Max Tx Power | 30dB
No. Sectors per site | 6,3 | Max Antenna Gain | 33 dB
Call arrival rate | adjusted to observe blocking | Vertical Beam width | 4 deg
Frequency | 2GHz | Horizontal beam width | 3 deg
Pathloss Model | 3GPP Urban | Antenna Radiation Model (with 3D extension) | 3GPP
Total User Population | 20000 | Max Antenna Attenuation | 30 dB
Inter site distance | 1200m | Resource sharing factor with BS | 50%
Bandwidth | 5MHz | BS Parameters | RS Parameters
User Distribution | Uniform, Non Uniform | Height | 32m | 10m
No. of BS | 19 | No. of RS per Site | 0.3
Max Tx Power | 39dB | Max Tx Power | 24dB
Antenna Radiation Model (with 3D extension) | 3GPP | Antenna Radiation Model (with 3D extension) | Omni directional
Vertical Beam width | 10deg | Vertical Beam width | 10deg
Horizontal beam width | 70,35 deg | Horizontal beam width | 360 deg
Max Antenna Gain | 19dB | Max Antenna Gain | 9dB
Max Attenuation | 20dB | Max Attenuation | 0 dB
BS antenna tilt | 20deg | RS antenna Tilt | 10 deg

Figure 4.4: Key system level simulation parameters for evaluations of LB-BSOF

4.6.2 Simulation Model

Our system level simulator models a multi cell WCS with DA similar to that discussed in section 4.4.4.1 and shown in figure 4.1. The performance of LB-BSOF can be anticipated to be dependent on the size of super cells. Therefore, two different sizes of super cell are considered with 3 and 6 cells per super cell. As highlighted in [3] and also discussed in the beginning of this chapter that LB in heterogenous networks becomes more difficult because of the different cell sizes and propagation factors of macro and Femto or RS based cells. To this end we
4.6 Simulation Results

Present extensive simulation-based evaluation of LB-BSOF for conventional as well as heterogeneous WCS. The key system-level simulation parameters are listed in the table in figure 4.4.

4.6.3 LB in Macro Cell based WCS

4.6.3.1 LB-BSOF with Super cell of 3 cells

We start with the conventional 3 sector deployments scenario and assume a super cell consisting of 3 sectors projected from the same site.

Figure 4.5 shows the average HB logged in WCS for a scenario of uniform geographical distribution shown in figure 4.6. Logging period is ensured to be long enough such that system reaches a steady state i.e. blocking percentage becomes almost stable. Therefore, in the following discussion when we will refer to blocking, either HB or SB, it would mean blocking in steady state.

It can be seen in figure 4.5 that LB-BSOF reduces HB significantly compared to no LB and brings it reasonably close to HB with ICC (i.e. absolute minimum). What is more important is to notice that even in case of uniform user distribution, LB-BSOF can yield this significant performance improvement. This is because, instead of perfect uniform geographical distribution and same transmission power, different cells can have different number of users associated to them because of shadowing. This can be seen in figure 4.7 that shows users associated per sector based on maximum reference signal received power, for perfectly uniform user distribution shown in figure 4.6.

Figure 4.8 indicates observed HB as percentage of the ICC HB. The gain of LB-BSOF is more figuratively observable in figure 4.8. In this scenario of perfect uniform distribution and only with super cell of three sectors, LB-BSOF yields over 150% reduction in HB by reducing it from 190% above absolute minimum to
Figure 4.5: Hard blocking with super cell consisting of three macro cells in uniform user distribution scenario. LB-BSOF performance is compared against the Ideal Central Control (ICC) based LB.

Just 35% over ICC, as can be seen in figure 4.8. It should be noted that although the LB-BSOF balances user association among super cells *independently* (perfect global balance among all cells in the system is neither aimed for nor achieved through LB-BSOF) still significant improvement in performance is achieved with LB-BSOF. Most importantly, since the cooperation within super cell i.e. sectors associated to same site would cost negligible overheads, this significant improvement in performance is achieved without compromising on agility and scalability i.e. in a purely self organising manner.
4.6. Simulation Results

The effect of LB-BSOF on interference is shown in figure 4.9. Although LB-BSOF has less adverse effect in terms of increasing interference and thus SB compared to ICC, but compared to no LB in place, it increases the interference significantly. This was expected because an independent variation in transmission powers of certain sectors of some super cells in order to adapt their coverage may increase interference in absence of any interference mitigation and tight frequency reuse in place.

Commercially, full frequency reuse is hardly used with three sectors per site as it will cause intolerable interference as observed in figure 4.9. Therefore, a deployment architecture with 6 sector per site is widely used in WCS. In this case a super cell can reasonably be assumed to be made of 6 sectors associated to same
Figure 4.7: User association per sector for uniform user distribution with three cells per super cell. Even with same transmission power and perfectly uniform user distribution, user association per sector is largely different, mainly because of shadowing.
4.6. Simulation Results

4.6.3.2 LB-BSOF with Super Cell of 6 Cells

Figure 4.10 shows the HB and as well as SB for the same uniform user distribution based scenario as discussed in previous section but now with 6 sector per super cell. Before we discuss the performance gain of LB-BSOF in this case, it is important to explain the two key observations that can be made by comparing the results for six sector case in figure 4.10 with results for case with 3 sectors. (figure 4.5 and 4.9)

First, the SB in 6 sector case has almost reduced to zero. This large improvement compared to 50% SB in 3 sector scenario is because of the following two main factors. First, with deployment topology used in case of 6 sectors each sector has only one strongest interferer sector. Whereas, in case of three sectors, each sector has two strongest interferes. Second, in case of 6 sectors inference from other
Interfering cells is also avoided due to better antenna directivity and resulting interference isolation. This can be seen by comparing sector layout in figure 4.6 and 4.11. This shows that if reasonable interference avoidance scheme is in place, LB-BSOF can substantially reduce HIB without having noticeable impact on interference i.e. SB and thus can provide significant net gain in grade of service. This observation can be made consistently in all results for six sector based scenario subsequently presented in this chapter.

Second observation on comparison of 3 and 6 sector scenarios is that, in 3 sector scenario when 50% of users were being blocked because of high interference, it
4.6. Simulation Results

Figure 4.10: Hard blocking and soft blocking with super cell consisting of six for macro cells based scenario with uniform user distribution.

means effectively only 50% of users were loading the system. Even then, the Hard Blocking (HB) was above 5%. Whereas, in case of 6 sectors, with same user distribution and traffic profile, no users is being blocked due to interference, still the HB is just 2%. This is due to the fact that, with 6 sectors per site instead of 3, with full frequency reuse, number of available channels have almost doubled the capacity of the system.

Another factor that effects system performance in terms of blocking is the Adaptive Modulation and Coding (ACM). With ACM in place, use of higher order MCS’s is possible when interference level in system is low (i.e. 6 sector case).
Figure 4.11: Non uniform user distribution used for evaluation of LB-BSOF performance. 50% of the total 20000 users are concentrated in randomly located pockets that are 114 in count, in the whole coverage area. Rest 50% of users are spread uniformly in the whole coverage area.

This in turn may result in much less system loading for same traffic demand and users distribution compared to system where interference level is high (i.e. 3 sector case). By loading here we mean percentage of engaged channels.

Due to these multiple factors that affect system performance in terms of blocking, we have used ICC instead of an analytical attempt to estimate blocking. ICC not only provides an alternative centralised algorithm for implementation of LB but it also provides a benchmark for comparison of minimum blocking obtainable with
LB-BSOF with the absolute minimum blocking obtainable with its centralised counter part i.e. ICC.

Although a perfectly uniform distribution is neither real, nor the main use case of LB-BSOF, still it can be seen in figure 4.10 as well as 4.5 that even in this hypothetical scenario of uniform user distribution, LB-BSOF reduces HB noticeably.

As discussed earlier, in general, the user geographical distribution in real world is not perfectly uniform. Usually there are some concentrated pockets of users i.e. medium term or long term hot spots at some locations in the coverage area while user density can be relatively low at other places. To test LB-BSOF in such realistic scenario, we consider a non uniform traffic distribution scenario as shown in figure 4.11.

Figure 4.12 shows both SB and HB, for this scenario of non uniform user distribution, referred to as realistic user distribution for its better resemblance to real world user population compared to uniform user distribution that would be referred to as ideal. By comparing figures 4.12 and 4.10 it is worth noticing that for same traffic requirement and total number of users in the coverage area, the average HB with realistic user distribution is much higher i.e. 5% compared to the just 2% observed in case of uniform user distribution. This shows the substantial impact user geographical distribution can have on system performance.

Furthermore, it should be noticed that HB with ICC stays same i.e. 1% in both cases of ideal and realistic user distribution. This is because ICC performs perfect LB among all cells in the system. By using hypothetical global control ICC adapts the cell sizes to take exactly same number of users in each cell. Therefore with ICC the effect of user distribution is fully undone. Thus the HB with ICC is dependent on total number of users in system, their traffic demands, and the amount of resources available per cells only, as explained above. Since these parameters stay same in both scenarios hence the blocking for these parameters, observed with ICC is same.
Figure 4.12: Hard blocking and soft blocking with super cell consisting of six macro cells in non uniform user distribution scenario.

The key inference from figure 4.12 is that LB-BSOF reduces HB very close to absolute minimum HB. The HB with LB-BSOF in terms of percentage of ICC HB, is shown for both non uniform and uniform user distribution respectively, in figure 4.13. LB-BSOF reduces HB form 280% above absolute minimum to just 7% above absolute minimum, in the realistic user distribution scenario. In the ideal scenario of perfectly uniform user distribution, a noticeable performance improvement is obtained though not as close to absolute minimum as in case of realistic user distribution. This gap in performance gain in two scenarios can be explained with help of figure 4.14 that compares user associations with and
4.6. Simulation Results

Figure 4.13: HB as percentage of minimum blocking achievable with ICC, with and without LB, both for uniform and non uniform user distribution with super cell consisting of six macro cells.

A realistic user distribution results in more diverse number of users per sector, compared to ideal user distribution, providing LB-BSOF more margin to adapt and thus converge to a local average user per super cell that is closer to the global average.

The impact of user distribution is more clearly investigatable in figure 4.15 that plots average loading in the cells and compares average loading with LB-BSOF and ICC in both uniform and non uniform user distribution scenarios. The CDF curves indicate the average loading percentage in different cells. The load per cell is averaged over the whole simulation period from start of the simulation to the point where the average blocking becomes stable. The CDF of loading with ICC in both scenarios, shows the normal loading range with perfectly balanced users association across cells. It is important to observe here that even with perfectly
Chapter 4. Load Balancing through SO of Coverage

Figure 4.14: User association with ICC, LB-BSOF and without LB, both for uniform and non uniform user distribution with super cell consisting of six macro cells. Notice the fact that each adjacent 6 cells make a super cells thus have almost even user association among them.

balanced user association i.e. with ICC, the CDF is not a straight vertical line, i.e. loading is not same across cells. This is because of the randomness that arises from arrival process, holding times and the ACM. However, in stead of presence of all these random factors, CDF with ICC curve can be used to roughly determine a region of natural loading with a given system configuration, with given number of users and traffic demands types in the system. This region is identified as the central region in figure 4.15. We name the zones out side this natural loading zone as artificially under loaded and artificially over loaded zone. The term artificial
4.6. Simulation Results

Figure 4.15: CDF of average load observed in different cells, with ICC, LB-BSOF and without LB, both for uniform and non-uniform user distribution with super cell consisting of six macro cells.

reflects the fact that, this overloading or underloading is because of imbalance of traffic among cells and not because of shortage or excess of user population on system wide scale. It can be seen that with ICC i.e. perfectly balanced users association, the average loading per sector should stay roughly between 40-57% in the system under consideration. But this is not the case, even with the ideal user distribution. This is because as explained above, even ideal user distribution does not guarantee a balanced user association in real WCS because of shadowing etc.
With the realistic user distribution the slope of the CDF is even lower, that means loading among sectors varies largely. With this realistic user distribution, as high as 37% of the cells can be artificially under loaded. This pushes around 33% of the cells into artificially over loaded zone, causing artificial congestion. Finally it can be seen in figure 4.15 that LB-BSOF confines the loading back to natural zone irrespective of the geographical distribution and therefore yields a significant reduction in HB as shown in figure 4.12 and 4.13.

It is important to mention that this reduction in HB achieved by LB-BSOF is at almost no cost in terms of either interference or signalling overheads. However, it is interesting to note at the bottom of figure 4.12 that, not LB-BSOF, but ICC causes slight increase in SB. This is because in order to achieve perfectly balanced global user association ICC has to play with the power in an unconstrained manner. Where as in LB-BSOF does not try to achieve global balancing, rather aims for local balancing at super cell level only and therefore does not have to change power very largely. Furthermore, LB-BSOF also has threshold and minimum and maximum powers. Therefore LB-BSOF has less impact on the interference compared to ICC.

To further investigate this aspect, the individual transmission power levels of all cells with ICC and with LB-BSOF are shown against the fixed transmission power levels with out any LB in place, in figure 4.16 along with the average power levels per cell in figure 4.17. It can be seen that ICC not only varies power levels across sectors more vigorously but it also raises the average transmission power level as well. Both of these facts results in higher interference. The advantage of LB-BSOF over ICC in terms of low interference can be seen in both 6 sector as well as 3 sector based scenarios in figure 4.12 and 4.9 respectively.

In addition to this advantage in terms of improvement in grade of service, LB-BSOF also may have positive gain in terms of the energy efficiency as is indicated in figure 4.17 by the reduced average transmission power per sector. Further
4.6. Simulation Results

Figure 4.16: Transmission power of association control reference signal with ICC, LB-BSOF and without LB, for non-uniform user distribution with super cell consisting of six macro cells.

Investigation of the energy saving aspects and the modification of LB-BSOF for objective of energy efficiency can be scope of the future work.

4.6.4 LB in Heterogenous WCS

In addition to the conventional macro cell based WCS, future WCS e.g. LTE-A are supposed to be heterogenous in the sense that they would contain Micro/Pico/Femto cells and relay stations as well. All of different cells types other than RS based cell will be referred to as Femto cells without loss of applicability of results presented to other cell types.
In case of Femto cells different antenna type and low radiation power is used to keep the cell size very small to limit the interference. Difference between the Femto and RS cells lies mainly in the fact that Femto cells have their own backhaul either wired or wireless, whereas RS rely on a donor base station to provide usually wireless backhaul access. Therefore, in case of Femto cells load balancing, we need to consider the coverage link only both for Femto and macro cells, but in case of RS the access link also need to be considered by the LB balancing mechanism. In the following two subsection we apply and evaluate LB-BSOF and ICC performance in heterogeneous WCS with Femto cells and RS respectively.
4.6. Simulation Results

4.6.4.1 LB with Femto cells

Figure 4.18 plots HB for heterogeneous network containing Femto cells in addition to macro cells. The layout of the network is shown in the figure 4.1 with small circles representing the Femto cells. It can be seen in figure 4.18 that without any load balancing, HB on macro cells is very high i.e. 7% whereas HB on Femto cells is very low i.e. 0.1%. This is mainly because of large difference in the size of Femto and macro cells. Very low transmission power and the low antenna gain of Femto cells means that very small number of users get associated with them. If same amount of radio resources are available in Femto cells, this results in large
difference in the loading as shown figure 4.19. It can be seen in figure 4.18 that

![Graph](image)

**Figure 4.19: Loading in the heterogenous WCS with Femto Cells**

LB-BSOF not only reduces the blocking in macro to very close to the absolute minimum, it also reduces the blocking in Femto cells to zero. This is because it not only performs inter macro cell LB, but also performs macro to Femto cell LB to some extent to achieve better LB. This fact is observable through the slightly increased average loading of Femto cells with the LB-BSOF as shown in figure 4.19.

**4.6.4.2 LB with RS**

Figure 4.20 shows the HB in a heterogenous network containing macro cells and RS cells. HB is logged for coverage links of both macro cells and RS cells as well
4.6. Simulation Results

Figure 4.20: HB in heterogeneous WCS with RS. Blocking on coverage as well as access link is logged

as for the RS-BS access link. It can be seen in figure 4.20 that with the given traffic and user distribution, without any LB, the blocking on BS is as high as 18% whereas blocking on the RS is only 0.2% and the blocking on the access link is almost 0%. The low blocking on RS is because of the very small coverage area of RS as explained in previous section for Femto cells. This very small number of user associations results in very low load on the access link, hence the access link also does not face any congestion in this case, as all the congestion is being faced by Macro cell users. With the LB-BSOF in place, congestion on BS is significantly reduced i.e. from 18% to just 4.6% which is very close to that with ICC 4.5% i.e. absolute minimum blocking possible with the given amount traffic and radio resources. However, it is interesting to note that the blocking on the access link has slightly risen with LB. This is because the access link has only $\frac{1}{3}$
of the resources available compared to those available on the coverage link of RS as explained in the section 4.3.

As mentioned in section 4.6.1 since SO-Functions ALA and BSR, are not modeled, so no LB on access link is performed, however, these results for heterogeneous network with RS, are interesting as they highlight the need for LB not only on access link but also at macro and global level as discussed in sections 4.5.2 and 4.5.3.

Nevertheless, it should be noted even without SO-Functions ALA and BSR the LB-BSOF provides a significant improvement in grade of service in relay enhanced WCS, by reducing the overall blocking as shown in figure 4.21.

Figure 4.21 plots the total blocking aggregated on all the three links, i.e. MS-BS, MS-RS and RS-BS in the heterogeneous network for the same scenario discussed above. It can be seen that, instead of increase in the blocking on access link,
4.7 Conclusions

In this chapter, we built on extensive survey of the state of the art load balancing schemes and developed a novel generic analytical framework for system-wide load balancing using BSOF and named it as LB-BSOF. First, globally optimal solution of the problem (SO-Objective) of minimising blocking, over resources and traffic per cell in system, was determined as SO-Goal in terms of number of global optimality conditions. A simpler solution was developed to reduce the complexity of the global solution. In addition to the assistance in decomposition of SO-Goal into local SO-Functions, this proposed simplification also helped us to implement an Ideal Central Control (ICC) based LB algorithm to benchmark our LB-BSOF based distributed SO solution developed in this chapter. A novel concept of supercell was proposed and used to develop the SO solution implementable thorough SO-Functions at local level. A number of practical actuators and functionalities that can be executed as SO-Functions to achieve the designed SO-Goals were identified and their use was elaborated with pragmatic use case illustrations. Finally, a simplified version of the proposed framework i.e. LB-BSOF has been investigated through extensive system level simulations for conventional as well as heterogenous WCS. The performance is compared against the optimal benchmark and with no LB, cases.

The numerical results demonstrate that even with limited functionality implemented due to constraints of simulator, LB-BSOF can achieve as high as 270% reduction in blocking for realistic non-uniform user distribution scenario in conventional WCS compared to no LB. In case of heterogenous WCS, where LB-BSOF with all four SO-Functions is anticipated to be more advantageous, the
limited implemented version of LB-BSOF with two SO-Functions, shows a performance gain of up to 100%.

The key advantage of proposed framework is its built in SO as it is designed using BSOF. It is fully scalable because of distributed implementation and has very low complexity. Its agility is limited for short scale dynamics because of the handovers it may trigger, but it is perfectly agile for medium to large time scale dynamics.

It is also autonomous as it does not require any central controller or external human intervention. Its stability has been marked by comparing its performance in number of scenarios with ideal central control based solution.
Multi Objective Optimization through SO of Frequency Reuse & Deployment Architecture

5.1 Introduction

In the previous two chapters, we addressed two different problems of concentrated hot spots and uneven load and user distribution and developed SO solutions for maximisation of spectral efficiency and minimisation of hard blocking using BSOF, respectively. The SO frameworks presented there focused on single optimization objectives and are mainly aimed for small to medium time scale dynamics. In this chapter we present a novel SO framework that deals with relatively long term dynamics of WCS (figure 1.1). Furthermore, the SO-framework presented in this chapter focuses on jointly optimizing the multiple long term optimization objectives in cellular system namely spectrum efficiency, service area fairness and energy efficiency.

This chapter has two major contributions. First, we develop a novel Performance
Chapter 5. Multi Objective Optimization through SO of Frequency Reuse & Deployment Architecture

Characterisation Framework (PCF) consisting of three novel metrics to characterise the three major long term performance aspects of WCS in holistic manner. Through PCF we also evaluate and investigate the long term performance of variety of WCS with different possible frequency reuse and deployment architectures.

Second, using PCF we develop a framework for Self Organizing Frequency reuse and Deployment (SOFD) by applying the principles of BSOF. SOFD can maintain the desired operational objectives in terms of the three aforementioned performance aspects in face of long term dynamics of WCS. We define our SO-Objective as multi objective optimization problem. Given the complexity of problem, here we adapt a semi analytical approach to dissolve the SO-Objective into SO-Goal and SO-Functions and developed a pragmatic solution. We also demonstrate our concept through extensive simulations and numerical results.

This chapter is organised as follows, in section 5.2 we present the necessary background and context and identify the WCS's main long term design factor i.e. Frequency reuse and Deployment architecture (FD) factors that affect WCS performance over long term in terms of major performance criteria like capacity, QoS and Operational Cost. In this section we also present a brief overview of the literature that is most pertinent to the contribution in this chapter to establish the novelty of our work. In section 5.3, we present PCF i.e. a novel performance characterization framework to characterise the performance of various FD's for WCS in terms of the three major long term performance aspects identified in section 5.2. In section 5.4 we evaluate numerical values of the developed metrics for different possible FD's for future WCS through extensive numerical simulations. By comparing these performance results we investigate and developed insight into the dependency of performance of a WCS on major long term WCS design factors i.e. FD factors like number of RS's per site, number of sectors per site and frequency reuse factors. In section 5.5 we use the insight developed in pervious sections and present a SO solution SOFD to meet the multiple perfor-
mance objectives in face of long term WCS dynamics. SO-Objective is defined as multiobjective optimization problem. It is transformed to SO-Goal by defining a bimodal utility function to represent all three objectives together. Pragmatic SO-Functions are then proposed to achieve these goals and numerical results are presented to demonstrate the whole concept of SOFD. Finally section 5.7 concludes the contributions in this chapter.

5.2 Context and Background

5.2.1 Need for SO for Long Term Dynamics

Attraction triggered random hot spots and non uniform traffic distribution due to mobility in general causes small to medium time scale variations (figure 1.1) in load offered to cellular system and have been addressed extensively in previous chapters. In addition to these geographical user distribution based dynamics, users' behaviour i.e. data rate demands and expectations of QoS also changes with the specific hours of the day, days in the week, week in a months and seasons of years etc. For example, traffic volumes can be expected to be much higher during the day in business hours compared to the early morning hours. This can be seen in figure 5.1 showing the statistics of traffic volume logged over several days and then averaged over hours of day. In this figure certain peak hours can be clearly identified when traffic is much higher than the overall average. Such statistics show a certain pattern of peaks and troughs corresponding to the specific hours of day. Similarly, different days of week can have different traffic patterns specific to them. For example week days can be more busy than the weekends. Similarly seasons and festival also have an effect on the amount and type of traffic users generate.

Another aspect of such dynamics that remain persistent over relatively long term
basis, is the change of expectation of users in terms of cost, capacity and quality of service. For example, during business days users might have more sensitivity to QoS than on weekends. On the contrary, the sensitivity to cost might be high on last days of month or weekends then during business hours or days. Similarly festive seasons also might have an effect on the cost, QoS and capacity expectation of users.

In addition to user behaviour, operator’s objectives and considerations are also bound to vary with such socio economically triggered dynamics. For example, cost of energy during night and weekend is much lower than during the working hours and days respectively. Similarly, in winter operational cost and energy cost may go higher than in summer.

Most of these variations in user traffic demand, expectation, operators consideration and objectives remain persistent over long term and can be predicted through the statistical analysis of previous data or through forecasts based on the surveys. Instead of this semi predictability, they are still difficult to deal with the conventional fixed and rigid FD’s based WCS. In case of fixed FD, designed with respect to the average traffic requirements, shortage of capacity in peak times and seasons and wastage of energy and radio resources in off peak times or seasons, is inevitable. Therefore, adaptive and flexible system design is the only solution for optimal radio resource and energy efficiency and QoS service. However, in order to be pragmatic and cost effective for large systems with huge complexity, this adaptation solution should require minimal human intervention i.e. should be autonomous and at same time agile, scalable and stable. So a self organising mechanisms that can cope with such large time scale dynamics by adapting the WCS is inevitable and is the focus of this chapter.
5.2. Context and Background

5.2.2 Major System Optimization Objectives on Long Term

Optimization process of wireless networks can have a number of different target objectives like maximization of capacity, coverage, fairness, spectral efficiency, throughput and QoS or minimization of cost, energy consumption and outage etc. However over long term, all these objectives boil down to three main categories of performance measures.

1. Capacity Oriented Performance Measures: These include cellular capacity, spectral efficiency, throughput, and goodput.

2. QoS Oriented Performance Measures: Fairness and outage are well known examples of QoS measures.

3. Operational Cost Oriented Performance Measures: Operational cost of cellular systems further have two major factors:

   (a) Labour Cost: Cost of labour required for operation, optimization and maintenance of sites and the switching networks

   (b) Energy Consumption: Energy consumed to keep the cellular system running is major factor of operational cost.
As discussed in the chapter 1, if the solutions in WCS that deal with optimization of aforementioned objectives are self-organizing, a large fraction of operation cost (i.e., the labour cost) can be reduced. Therefore, in all solutions featuring SO, the objective of operational cost minimization actually reduces largely to minimization of energy consumption or in other words maximizing of energy efficiency. Particular energy efficiency is becoming increasingly important not only to decrease the operational cost of the WCS (hence increase the profit of operators), but also to provide greener and echo-friendly WCS. Since a huge amount of energy is wasted merely due to the continuous running of continually required equipment or mismanagement in general, SO is a promising approach to save energy in face of the long term dynamics of the WCS.

To enable a SO solution to achieve the objectives of energy efficiency, capacity and fairness, in this chapter, we develop PCF consisting of three metrics to characterize each of these three types of performance aspects of various FD's for WCS. PCF is further used in SOFD for joint optimization of these three aspects in face of long term dynamics of WCS.

5.2.3 Major Factors Effecting Long Term Performance of WCS

WCS has myriad of FD factors that can affect the performance of WCS in terms of the three types of aforementioned performance aspects, i.e., spectral efficiency, energy efficiency and long term fairness of the service in the coverage area. These factors include inter-site distance, number of sectors per site, antenna type and antenna gain, antenna height, number of relay stations per site, type of relay stations (i.e., amplify and forward and decode and forward), transmission power, modulation and coding schemes, frequency reuse factor etc. In this chapter we focus on four major parameters that affect the WCS performance in terms of
capacity, QoS and energy efficiency, most pronouncedly. These are number of sectors per site (denoted by S onward), number of RS per site (denoted by R onward), Frequency reuse factor (denoted by F onward) and Modulation and Coding Efficiency (denoted by MCE onward). Below we briefly discuss the trade offs these four parameters offer among the various performance measures in next generation WCS.

Unlike legacy WCS, next generation WCSs are based on OFDM/OFDMA MAC and therefore allow to dynamically adapt individual user links according to time and frequency varying channel conditions. This enables the use of modulation and coding schemes with higher MCE for users with better link quality, thus exploiting the multi user and channel diversity to improve the overall spectrum efficiency. Over long time scale, where the effect of short term channel variation is averaged out, a higher average MCE in WCS can be achieved by designing a FD that improves the overall SIR geographic distribution in the whole coverage area of the system. In a fully loaded WCS that does not resort to any feedback and cooperation based interference mitigation techniques, the larger the F the better will be the average available SINR in the coverage area. This brings in a tradeoff between increase in spectral efficiency achievable by increasing spectrum reuse and increase in spectral efficiency by resorting to higher MCE through link adaptation. Another degree of freedom is added to this tradeoff through sectorization since sectorization can potentially improve SINR in the coverage area by reducing the effective number of interfering cells. However, at the same time it incurs loss in terms of spectrum reuse efficiency as well as trunking efficiency (proposition 12 in section 4.4.3) if the available spectrum is divided among the sectors to avoid inter sector interference.

As mentioned above, in addition to the spectral efficiency or capacity, another desirable performance goal is homogeneity of the level of service that can be provided to the users. Over long term, this variation of service level that is
dependent on very short term channel conditions, is averaged out in temporal sense. However the homogeneity of service level in spatial sense in the whole coverage area i.e. fairness over the service area (simply called fairness onward) still remains an important performance criteria. Specifically, consideration of the service profile of cell edge users is crucial in WCS design as they are most vulnerable to receive lowest SINR due to their large distance from desired cell and their close proximity to interfering cells. This goal is one of the top priorities of 3GPP [111]. To achieve this goal, addition of Relay Stations (RS) has been considered in WCS FD e.g. LTE-A and 802.16m. RSs have been shown to yield a significant improvement in SINR distribution in the low coverage areas e.g. cell edge or heavily shadowed zones [112]. Although RS also offer potential for reduction in cost [113] but a down side of RSs is that they need extra radio resources to multiplex either in time or frequency with their donor BS in order to avoid mutual interference [113]. This introduces another tradeoff between the gain in spectral efficiency that RSs can provide by boosting SINR and the loss in spectral efficiency that the RSs cause due to multiplexing with BS. Most of the studies on RS enhanced WCS report the advantage of relays assuming centralized resource allocation scenario. Thus, the heavy amount of signaling required to implement an interference mitigation technique is neglected in these analysis [114]- [115]. Therefore, in this chapter we consider a system without any feedback or cooperation based interference mitigation techniques i.e. a distributed OFDM/OFDMA cellular system where interference and hence performance is determined mainly by FD factors i.e. $F$, $S$ and $R$ and thus $MCE$.

### 5.2.4 Relevant Work

Most of research on SO has focused so far on extension of the already available solutions in MAC and physical layer to bring in short to medium term adaptability [3]. Relatively less attention has been paid to the need and potential of long
5.2. Context and Background

In addition to conventional F e.g. F=1, F=3, many advanced Frequency reuse (F) schemes have also been proposed to achieve a tradeoff between the spectrum efficiency achievable by spectrum reuse factor and spectral efficiency achievable by using higher coding and modulation schemes, adaptively. These advanced F schemes can be classified in three main categories; 1) fully isolated fractional F [116], 2) partially isolated fractional F [117], 3) dynamic fractional [118].

In fully isolated fractional F, cell is divided into two geographical parts. Central part uses F=1 and edge part uses higher F e.g. 3 for three sector case. This scheme improves the cell edge performance but sacrifices significant throughput at the same time due to F=3 [116]. In partially isolated F schemes, all cells use all subcarriers but outer parts of the cells use a group of carriers with low power. This same subcarrier can then be used in adjacent cell with high power. This scheme yields better throughput than fully isolated fractional reuse because of resorting to F=1 but its performance degrades rapidly as the system load increases [117].

Dynamic fractional F does not divide the cell area on geographical basis into cell edge or cell center, neither does it split subcarriers. Rather it establishes virtual groups of carriers to be used by virtual groups of users. These virtual groups of subcarriers and corresponding users are determined dynamically for each frame by estimating the channel condition for each user on each subcarrier in each BS. Although this scheme has been shown to have relatively better average throughput compared to the other two schemes, throughput at cell edge is worst in this case [118]. Furthermore, the need for global cooperation based on heavy signaling and huge computational power required to implement this scheme renders it effectively impractical because of its low scalability and agility. In summary,
each of these frequency reuse schemes proposed so far is optimal for a specific scenario and meets high performance criterion for some metrics while sacrificing performance in other metrics.

In order to address the issues of complexity and signaling overhead in RRM for next generation WCS, SO Deployment (SOFD) framework presented in this chapter combines the simplicity of conventional frequency reuse with the adaptive FD potential of future wireless networks. Unlike previous works on frequency reuse that deal with spectral efficiency or capacity optimization on short to medium time scale SOFD deals with long time scale dynamics. It optimizes not only spectral efficiency but also considers energy efficiency and long term fairness of service in the coverage area as well. Furthermore, unlike previous works that focus on frequency reuse, SOFD also incorporates effect of number of sectors $S$ and relays per site $R$ on these multiple performance objectives.

To enable SOFD, in the next section we first design suitable metrics to quantify the performance of WCS over long time scale in the three aforementioned major performance aspects.

### 5.3 A Performance Characterisation Framework (PCF)

In order to define the SO-Objective that consists of multiple optimization objectives in this case, there is need to have a unified framework for quantifying these multiple objectives as function of same set of parameters under consideration. To this end, in this section we propose a novel performance characterisation framework that consist of three metrics one representing each category of the three main performance measures i.e. the capacity, QoS and energy based performance measures. This framework is further used to simplify the SO-Objective into SO-
5.3. A Performance Characterisation Framework (PCF)

Goal and SO-Function in next sections of this chapter.

In subsection 5.2.3 first we discussed how the long term performance of WCS is dependent on four major FD factors: 1) spectrum reuse factor $F$, 2) number of sectors per site $S$, 3) number of relay station per site $R$ and 4) modulation and coding efficiency $MCE$ achievable through link adaptation. In following subsection we extend that discussion in quantitative sense to develop a performance characterization framework.

5.3.1 System Model

Consider down link scenario of WCS where $\mathcal{N} = \{1, 2, 3...N\}$ is a set of BSs in the coverage area, $\mathcal{S} = \{1, 2, 3...S\}$ is set of sectors per BS and $\mathcal{R} = \{1, 2, 3...R\}$ is a set of RSs per BS. $\mathcal{K} = \{1, 2, 3...K\}$ is a set of users in the coverage area of the system, out of which $|\mathcal{K}_b|$ are in the coverage area of the BS’s and $|\mathcal{K}_r|$ are in the coverage area of RS’s, such that $|\mathcal{K}_b| + |\mathcal{K}_r| = |\mathcal{K}|$. $\mathcal{M} = \{1, 2, 3...M\}$ is a set of sub carriers allocated to each BS that further shares it with its child RS either in time or frequency with a sharing factor $\rho^b < 1$ such that $\rho^r = 1 - \rho^b$. Since BS and RS multiplex in frequency or time, hence they do not interfere with each other. Received Signal Level (RSL) in dBm from sector $s$ of the $n^{th}$ BS on $m^{th}$ subcarrier for $k^{th}$ user at a given location in the coverage area can be given as

$$RSI_{k,m}^{n,s} = P_{m}^{n,s} + G_{k}^{n,s}(\theta_{k}^{n,s}, \phi_{k}^{n,s}) + L_{k,m}^{n} (d_{k}^{n}, f) + \alpha_{k,m}^{n,s} \quad (5.1)$$

where $P_{m}^{n,s}$ is the transmission power on $m^{th}$ sub-carrier from the sector $s$ of $n^{th}$ BS. $G_{k}^{n,s}$ is the antenna gain of sector $s$ of $n^{th}$ BS towards user $k$. It is a function of the elevation angle $\theta_{k}^{n,s}$ and azimuth angle $\phi_{k}^{n,s}$ between location $p$ of $k^{th}$ user and bore site of respective antenna. $L_{k,m}^{n}$ is the pathloss as a function of distance $d_{k}^{n}$ between user $k$ and BS $n$ and the frequency of operation $f$. $\alpha_{k,m}^{n,s}$ is the log normal shadowing faced by the $s^{th}$ user, while receiving signal form $s^{th}$ sector of
Chapter 5. Multi Objective Optimization through SO of Frequency Reuse & Deployment Architecture

$n^{th}$ BS. Similarly, the received signal level from the $r^{th}$ RS of $n^{th}$ BS for user $k$ on $m^{th}$ carrier can be written as.

$$RSL_{k,m} = P_r + G_r (\phi_k, s) + L_{k,m} (d_k, f) + \alpha_{k,m}$$  \hspace{1cm} (5.2)

SINR for the $k^{th}$ user associated to a BS on $m^{th}$ subcarrier will be

$$\gamma_{k,m}^{n,s} = \frac{RSL_{k,m}^{n,s}}{\sigma_k^2 + I_{k,m}}$$  \hspace{1cm} (5.3)

$$I_{k,m}^{n,s} = \sum_{n \in N} \sum_{s \in S} RSL_{k,m}^{n,s} u(n, s, m)$$  \hspace{1cm} (5.4)

$$u(n, s, m) = \begin{cases} 
1 & m = m_k, n \neq n_k, s \neq s_k \\
0 & \text{otherwise}
\end{cases}$$  \hspace{1cm} (5.5)

$\sigma_k^2$ is thermal noise floor of $k^{th}$ users receiver and $n_k$ and $s_k$ respectively denote that particular BS and the sector to which user $k$ is being served on subcarrier $m_k$.

5.3.2 Relationship between the WCS Performance and Frequency Reuse and Deployment (FD)

The MCE achievable on a given link is dependent on the SINR available on that link. Theoretically, the maximum achievable MCE on a link can be determined by the Shannon bound i.e.

$$MCE_{k,m} = \log_2(1 + \gamma_{k,m})$$  \hspace{1cm} (5.6)

However, in practice MCE is a discrete function of SINR at the receiver and depends on the set of modulation and coding schemes being used WCS i.e.

$$MCE_{k,m} = f[\gamma_{k,m}]$$  \hspace{1cm} (5.7)
where \([\cdot]\) represents discrete function and \(MCE_{k,m}\) is modulation and coding efficiency of the link for the \(k^{th}\) user on the \(m^{th}\) subcarrier. Thus the total throughput of users attached to BSs can be given by:

\[
C^b_p = B \rho^b \times \sum_{\forall k \in K_v} \sum_{\forall m \in M_k} MCE_{k,m} 
\]

(5.8)

where \(\rho^b\) is the sharing factor with which resources are shared between BS and RS such that \(\rho^s = 1 - \rho^b\) where \(\rho^s\) denotes sharing factor for RS. \(M_k\) is a set of subcarriers allocated to user \(k\), and \(B\) is the sub-carrier bandwidth. By substituting (5.6)-(5.3) in (5.8), the maximum theoretically achievable aggregate throughput of users attached to BS can be determined by

\[
C^b_p = B \rho^b \sum_{\forall k \in K_v} \sum_{\forall m \in M_k} \log_2 \left( 1 + \frac{RSL_{k,m}^{n_s,m_s}}{\sigma_k^2 + \sum_{\forall m' \in M_k} \sum_{\forall n \in S} RSL_{k,m}^{n_s,m_s} u(n, s, m) } \right) 
\]

(5.9)

However, in cellular networks where link adaptation is in operation, the actual achievable aggregate throughput of all users attached to BS's can be represented by substituting (5.7) in (5.8)

\[
C^b_t = B \rho^b \sum_{\forall k \in K_v} \sum_{\forall m \in M_k} f[\gamma_{k,m}] 
\]

(5.10)

Similarly if the user \(k\) is attached to a RS instead of BS the SINR perceived can be given as

\[
\gamma_{k,m} = \frac{RSL_{k,m}^{r_h}}{\sigma_k^2 + \sum_{\forall r \in R \setminus r_k} RSL_{k,m}^{r_s,m_s} u(r,m) } 
\]

(5.11)

Then the aggregate theoretical and practical throughput of all users attached to RS in the coverage can be given as:

\[
C^s_t = B (1 - \rho^b) \sum_{\forall k \in K_v} \sum_{\forall m \in M_k} \log_2 \left( 1 + \frac{RSL_{k,m}^{n_s,m_s}}{\sigma_{k,m}^2 + \sum_{\forall r \in R \setminus r_k} RSL_{k,m}^{r_s,m_s} u(r,m) } \right) 
\]

(5.12)

where

\[
u(r,m) = \begin{cases} 
1 & m = m_k, r \neq r_k \\
0 & \text{otherwise}
\end{cases}
\]

(5.13)
\[ C_t = B \left( 1 - \rho^b \right) \sum_{\forall k \in \mathcal{K}} \sum_{\forall m \in \mathcal{M}_k} f \left[ \gamma_{k,m} \right] \]  

(5.14)

The total achievable throughput in the coverage area can be written using (5.9) and (5.12)

\[ C_t = B \rho^b \sum_{\forall k \in \mathcal{K}} \sum_{\forall m \in \mathcal{M}_k} \log_2 \left( 1 + \frac{RSL_{k,m}^{n_k,m_n}}{\sigma_k^2 + \sum_{\forall n \in \mathcal{N}_m} \sum_{\forall s \in \mathcal{S}} RSL_{k,m}^{n_k,m_n} u(n, s, m) + B \left( 1 - \rho^b \right) \sum_{\forall k \in \mathcal{K}} \sum_{\forall m \in \mathcal{M}_k} \log_2 \left( 1 + \frac{RSL_{k,m}^{n_k,m_n}}{\sigma_{k,m}^2 + \sum_{\forall r \in \mathcal{R} \setminus \mathcal{R}_k} RSL_{k,m}^{n_k,m_n} u(m) \right) \right) \]  

(5.15)

The unity functions in (5.5) and (5.13) capture the effect of resources reutilization. With our assumption of fully loaded system with no dynamic coordination for interference avoidance through cooperated or coordinated scheduling, (5.5) and (5.13) represent the effect of frequency reuse factor \( F \). It can be thus seen in (5.3) and (5.11) that SINR perceived by user is mainly dependent on the \( F \), \( S \) and \( R \). Furthermore (5.15) shows that the system throughput and hence the spectral efficiency is also dependent on resource sharing factor between BS and RS as well as actual mapping of SINR to MCE i.e. \( f[.] \). This mapping is determined by the set of modulation and coding schemes used in the system. In next subsection we build on these dependencies in order to design three holistic metrics to reflect capacity, fairness, and energy consumption oriented performance measures.

### 5.3.3 Proposed Performance Metrics

#### 5.3.3.1 Effective Spectral Efficiency (ESE)

The conventional definition of spectral efficiency is

\[ \text{Spectral Efficiency} = \frac{C_t}{BW} \text{ (bps/Hz)} \]  

(5.16)
where $BW = B \times |\mathcal{M}|$. While this metric is widely used to estimate spectral efficiency, it is not suitable in the context of our problem. This is because throughput is strongly dependent on very short term dynamics like fast fading and temporary shadowing and also on scheduling schemes used. Furthermore, it is also dependent on the medium term dynamics like the number of users and their geographical distribution and therefore heavily over shadows the effect FD factors such as $F$, $R$, and $S$ have on the long term performance of WCS. It is very difficult if not possible to determine optimal $F$, $R$, $S$ for long time scale objectives using this conventional throughput based measure for capacity oriented performance. Another difficulty with the throughput based evaluation of spectral efficiency in the scope of this problem is that a large number of full scale system level simulation with different FD’s and user distributions are required to estimate throughput as analytical assessment of throughput is not a viable option either.

A second option for measuring capacity oriented performance is conventional area spectral efficiency. However for evaluating this metric too throughput calculation is prerequisite rendering it unsuitable to represent the explicit effect of FD factors of system long term performance.

Our basic aim here is to quantify the long term performance of WCS by incorporating its dependencies on long term design factors i.e FD factors e.g. $F$, $R$, and $S$ rather than short term design parameters e.g scheduling and sub carrier power allocation etc. Therefore conventional throughput based metrics that depend on short term dynamics are not suitable to our purpose and a metric to quantify the capacity oriented performance independent of very short to medium term dynamics is essential. To this end, in this subsection we present novel metric to represent long term capacity oriented performance of WCS that can be used to directly characterize the spectral efficiency of various FD’s while explicitly accounting for $S$, $R$, $F$ and MCE. This metric has semantics similar to the area spectral efficiency but it does not require throughput estimation for its
calculation, rather it can be calculated with semi analytical approach through much reduced analytical and simulation complexity\(^1\). We call this metric *Effective Spectral Efficiency* (ESE) and represent it by \(\Upsilon\). Below we explain rational behind ESE.

Since the sub carrier bandwidth in given WCS is fixed so the throughput on single sub-carrier in a given link and hence the total throughput of the system depends on MCE on each link. The MCE in turn depends on SINR available on that link. Thus, from (5.15) it can be seen that with total bandwidth fixed, the theoretical and actual throughput hence the spectral efficiency of a WCS depends on the SINR's geographical distribution in the coverage area that in turn depends mainly on \(S, R\) and \(F\). Therefore, in interference limited scenario, \(\sigma << I\), with full frequency reuse among all sectors in the system, under full load conditions, the SINR available on sub-carrier \(m\) to user \(k\) is mainly dependent on the location \(p\) of the user within cell and can be written as.

\[
\gamma_p = \frac{\text{RSL}_p^{n,p}}{\sum_{n \in \mathcal{N}} \sum_{s \in \mathcal{S}} \text{RSL}_p^{n,s}}
\]  

(5.17)

Where \(s_p\) is sector in which point \(p\) lies. If the Power \(P\), antenna tilts \(\theta\) and frequency of operation \(f\) as shown in equation (5.2) and (5.1) are assumed to be constant, then the long time scale average RSL and hence the SINR at point \(p\) will be a function of its distance \(d\) and angle \(\phi\) from the its serving BS or RS and can be simply written as

\[
\gamma_p = f (\phi, d)
\]  

(5.18)

Now the average modulation and coding efficiency in cell can be given as

\[
\text{MCE}_{\text{cell}} = \frac{1}{A_{\text{cell}}} \int \int \log_2 (1 + \gamma_p(\phi, d)) \, d\phi \, dd
\]  

(5.19)

Where \(A_{\text{cell}}\) is the total coverage area of cell. In order to evaluate the system wide spectral efficiency in more practical manner, let's consider \(\mathcal{P} = \{1, 2, 3...P\}\)

\(^1\)A detailed comparison of effective spectral efficiency and area spectral efficiency is given in appendix C
5.3. A Performance Characterisation Framework (PCF)

is set of all points in the coverage area.

\[ \text{MCE}_{\text{area}} = \frac{1}{|P|} \sum_{p=1}^{P} \log_2 (1 + \gamma_p) \] (5.20)

In order to have an actual area measure \( P \to \infty \), but for sake of practicality and implementation in the simulations we assume that the total coverage area is divided into finite set of B virtual bins i.e. \( A = \{a_1, a_2, a_3, \ldots a_B\} \) each with area \( a \text{ m}^2 \) within which SINR remains constant i.e. \( B \times a = |A| a = A \), therefore \( P \) in (5.20) can be assumed to be a finite set of points representing center of each bin. Now the (5.20) can be written as

\[ \text{MCE}_{\text{area}} = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{\log_2 (1 + \gamma_b)} \] (5.21)

Now let \( L = \{0, 1, 2, 3, \ldots L\} \) is set of modulation and coding schemes available to be used in and \( \text{MCE}_i \) denotes the respective modulation and efficiency of \( i^{th} \) scheme. Where \( i = 0 \) means modulation and coding scheme with zero spectral efficiency i.e. no link and \( L \) is modulation and coding scheme with highest spectral efficiency. Now the pdf of MCE can be estimated as

\[ f(MCE_i) = \frac{B_i}{B} \] (5.22)

where \( B \) is total number of bins in the coverage area of the system and 

\[ B_i = \sum_{\forall p \in P} U_i(\gamma_p) \] (5.23)

and \( U_i(\gamma_p) \) is defined as follows.

For \( i \in L \setminus \{0, L\} \) \( L : U(\gamma_p) = \begin{cases} 1 & T_l < \gamma_p < T_{l+1} \\ 0 & \text{otherwise} \end{cases} \)

For \( l = L \) : \( U(\gamma_p) = \begin{cases} 1 & T_{l-1} < \gamma_p \\ 0 & \text{otherwise} \end{cases} \)
Chapter 5. Multi Objective Optimization through SO of Frequency Reuse & Deployment Architecture

And for \( t = 0 \): \( U(\gamma_p) = \begin{cases} 
1 & \gamma_p < T_1 \\
0 & \text{otherwise}
\end{cases} \)

\( T_1 \) is the threshold SINR required to use \( t^{\text{th}} \) modulation and coding scheme from set \( C \). \( T_2 \) is the threshold of minimum \( \gamma \) below which link cannot be maintained with pre-decided performance criterion and all such points in coverage area constitute the outage area.

Similarly CDF of MCE can be given as

\[
F(MCE_i) = \frac{\sum_{l=0}^{L} B_l}{B} (5.24)
\]

While (5.22) and (5.24) give PDF and CDF of MCE achievable with given PD, a numeric metric is also required to quantify this MCE. We define this metric to quantify the spectral efficiency achievable through MCE for a given SINR geographical distribution, as follows.

\[
T_{MCE} = \sum_{i=0}^{L} \left( MCE_i \times \frac{|P_l|}{|P|} \right) (5.25)
\]

where \(|P_l|\) is the cardinality of set of all points in which \( \gamma_p \) is such that \( t^{\text{th}} \) modulation and coding scheme can be used. Alternatively a more easily assessable version of (5.25) can be written as follows

\[
T_{MCE} = \sum_{i=0}^{L} \left( MCE_i \times \frac{B_l}{B} \right) (5.26)
\]

Note that

\[
\sum_{l=0}^{L} B_l = B (5.27)
\]

Hence (5.27) in conjunction with (5.22) and (5.24) implies that \( T_{MCE} \) in (5.26) is actually expected value of MCE i.e.

\[
T_{MCE} = E(MCE) = \sum_{i \in C} MCE_i \times f(MCE_i) (5.28)
\]
using $T_{MCE}$ we define the new metric that reflects the capacity oriented performance of FD namely effective spectrum efficiency as follows:

$$T = \frac{T_{MCE}}{T_{MF}} \times T_{SRF} \left( \frac{\text{bps/Hz}}{\text{site}} \right)$$  \hspace{1cm} (5.29)

where subscripts $SRF$ denote Spectrum Reuse Factor and $MF$ denote multiplexing factor. $T_{SRF}$ thus denotes number of times spectrum is reused within a cell (set of all sectors attached with same BS). Thus it depends on the number of sectors per cell $S$ and frequency reuse $F$. For example, if in system with 6 cells per site, if total spectrum is divided in two parts (i.e. $F=2$) and is used in each alternative sectors of the same site then $T_{SRF} = 6/2 = 3$. If FD has RS as well and $p^*$ is the factor with which spectrum is shared between BS and RS associated to it then $T_{MF} = 1/p^*$. If FD does not have RS $T_{MF} = 1$.

In (5.29) $T_{MCE}$ reflects the expected MCE and thus reflects spectral efficiency achieved through the use of various modulation and coding schemes in given geographical SINR distribution resulting from particular FD’s. A combination of $R, F, S$ i.e. $SRF$ denotes the spectral efficiency achieved through spectrum reuse and $MF$ denotes the multiplexing loss due to the use of RS. Thus, the ESE (5.29) represents the effective spectral efficiency while directly reflecting the effect of key FD factors and their respective tradeoffs highlighted in the previous sections. The main advantage of this metric is its ease of calculation as calculation of throughput through dynamic simulation is not required. Rather only the SINR geographical distribution for various FD need to be determined. This distribution then can be mapped to MCE’s using theoretical Shannon bound or using practical SINR thresholds of the standards under consideration.

Another key advantage of ESE is that it has the potential to reflect geographical areas of high importance with weighting factors to pronounce their importance and thus reflect them in the holistic ESE measure proportionally. For this $B$ in (5.26) will not be representing simply number of bins but it will represent...
sum of weights associated with each bin. The advantages of ESE evaluated in this way are explained in Appendix C and their exploitation is a scope for future work. In this chapter however will assume each bin of the coverage area has equal importance thus $B$ is simply number of total bins in the coverage area.

The process of calculating ESE is further explained in section 5.4 while presenting numerical results. In next subsection we extend this analysis to derive a suitable metric for fairness of service area that represents the effect of $F$, $R$, $S$ and MCE in the way similar to that explained in this section.

### 5.3.3.2 Service Area Fairness (SAF)

As discussed earlier, the notion of fairness in the context of long term dynamics is significantly different from the conventional notion of fairness that is considered when designing for very short time scale adaptive mechanisms e.g scheduling or power allocation to subcarriers etc. In case of long term dynamics under consideration in this chapter, all such short term dynamics can be neglected as they are averaged out. So, in scope of this chapter, by fairness we mean long term fairness among all users in the coverage area. More precisely its fairness in space than in time. We build on above derivations and define a metric for such long term fairness that reflects the effect of MCE, $S$, $R$ and $F$ on fairness among the type of service users are capable to receive dependent on their location in the service area and name it Service Area Fairness (SAF) given as:

$$\Phi = SAF = 1 \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left( MCE_b - \sum_{l=0}^{L} \left( MCE_l \times \frac{B_l}{B} \right) \right)^2}$$

(5.30)

where $MCE_b$ is the modulation and coding efficiency for the $b^{th}$ bin. SAF characterizes fairness among the users in the coverage area of a system by measuring how much the potential data rates of individual users within the service area deviate from the long term average data rate in the the service area, given same
amount of radio resources are allocated to all users. This deviation depends on
the SINR geographical distribution as well as mapping of that SINR to actual
data rate achievable by a user. Advantage of this metric of fairness is that it
exclusively captures the actual effect of link adaptation which is a key factor
in determining effective data rate a user can be served with. Furthermore, this
fairness metric treats justly all the users in the coverage area. This is because it
gives the cell edge users judiciously higher importance because as area is square
function of radius, thus more area lies farther from the cell center. In case of
uniform user distribution this means more users will lie farther from the cell cen­
ter and thus should have naturally larger influence in determining SAF. SAF is
maximum i.e.∞ when all users in the whole coverage area can establish links
with same spectral efficiency i.e. they can be served with same data rate, given
that radio resources are divided equally among them. More spatially uneven is
the spectral efficiency available to the all users in the coverage area, the smaller
would be SAF. In case, a finite bounds based more refined estimation of SAF is
required, Jain’s fairness index can be used to estimate this fairness of the service
profile as follows.

\[
JSAF = \frac{\left(\sum_{b=1}^{B} MCE_b\right)^2}{N \sum_{b=1}^{B} (MCE_b)^2}
\] (5.31)

The advantage of JSAF is that it has finite range starting from 0 (worst case)
to 1 (best case), and it is maximum when all users receive the same allocation.
JSAF is \( \frac{c}{B} \) when \( c \) out of \( B \) bins have same potential spectral efficiency and the
other \( B - c \) bins are in outage i.e. SINR is so low that even lowest available
modulation and coding pair cannot be used in that bin of the coverage area.

Although , SAF does not provide finite bounds on degree of fairness like JSAF, it
is still useful for the scope of our problem where quantitative comparison among
the service area fairness of various FD will serve the purpose and can help us to
select relatively the best among them. Of particular interest is the fact that SAF
weighs the cell edge users proportionally more, and gives deviation from the ESE,
makes it more suitable compared to JSAF in our context. Therefore, in the rest of this chapter SAF is used only to characterise long term spatial fairness.

5.3.3.3 Energy Consumption (EC)

We propose $\Omega$ as a negative measure of energy efficiency (i.e. Energy Consumption (EC) instead of saving) given as

$$\Omega = \frac{P}{T} \left( \frac{H_s}{\text{site}} \right)$$

(5.32)

where $P$ is power consumption per site which incorporates both fixed, as well as, variable power consumption per site, on downlink in a cellular system. Fixed power consumption is the power that is consumed in keeping the circuitry of BS sectors or RS alive no matter if there is traffic or not, until that sector or RS is completely switched off. Variable power consumption is power required for transmission on air interface and varies with the traffic load. Thus, power consumption on a site can be written as

$$P = \sum_{s=1}^{S} \left\{ P_{f,s} + P_{v,s} \left( G (\Gamma, D'), P_t, \omega_s \right) \right\} + \sum_{r=1}^{R} \left\{ P_{f,r} + P_{v,r} \left( G (\Gamma, D'), P_t, \omega_r \right) \right\}$$

(5.33)

where subscripts $f, v$ and $t$ denote fixed, variable, and transmission powers respectively. Post scripts $s$, and $r$ denote sector and relay respectively. For sake of simplicity we do not consider any stray losses e.g. feeder loss, connectors loss as they are negligible for the purpose of this analysis. Variable power consumption further depends on the transmission power $P_t$, traffic loading factor $\omega$ and antenna gain $G$. Antenna gain is further a function of efficiency of antenna $\Gamma$, and directivity $D$. The directivity of antenna has an important role in determining its gain and hence the transmission power required to provide certain coverage level. It can be written as

$$D = 4\pi / \left( \frac{\int_{0}^{2\pi} \int_{0}^{\pi} \beta (\theta, \phi) \sin \theta d\theta d\phi}{\beta (\theta, \phi) \text{max}} \right)$$

(5.34)
5.3. A Performance Characterisation Framework (PCF)

Where $\beta$ is function representing radiation pattern of antenna as function of spherical co-ordinate angles $\theta$ and $\phi$. For practical purposes the denominator of (5.34) can be approximated by product of half power beam widths $\psi_h$ and $\psi_v$ in horizontal and vertical plane. So (5.34) can be approximated as

$$D = \frac{4\pi}{\psi_h \psi_v}$$

(5.35)

In cellular systems the desired vertical beam width of antenna is around $\pi/18 \approx 10^\circ$ and horizontal beam width depends on the number of sectors per site e.g. for three sectors and six sectors, beam width of around $70^\circ$ and $35^\circ$ are usually used respectively. If we define $\alpha$ as factor determining the overlap between the adjacent sectors, we can write horizontal beam width as a function of $S$ as $\phi_h = \alpha \pi / S$.

Then (5.35) can be written as

$$D \approx \frac{72S}{\alpha \pi}$$

(5.36)

Typical value can be assumed to be $\alpha = 1.1$. To achieve a desired EIRP (Effective Isotropic Radiated Power) in the coverage area, less transmission power $P_t$ will be required for antennas with higher gains as

$$EIRP = \Gamma DP_t$$

(5.37)

If $P_d$ is the power required to achieve desired $EIRP_d$ with an omnidirectional antenna

$$P_d = \frac{EIRP_d}{\Gamma \times D}$$

(5.38)

Then the variable circuit power per sector for desired $EIRP_d$ can be written in dB as

$$P_v^s = 10 \log_{10} P_d^s - 10 \log_{10} \left( \frac{4\pi^2 S}{\alpha \psi^2} \right) + 10 \log_{10} \omega^s$$

(5.39)

Similarly, the variable circuit power on a RS can be written as

$$P_v^r = 10 \log_{10} P_d^r - 10 \log_{10} \left( \frac{4\pi^2 \omega}{\psi^2} \right) + 10 \log_{10} \omega^r$$

(5.40)
Putting (5.39)-(5.40) to in (5.33) and

\[ P = \left( \sum_{s=1}^{S} \left\{ P_{s}^f + \alpha \left( \frac{\omega \psi P_{d}^f}{41^a S} \right) \right\} \right) + \sum_{r=1}^{R} \left\{ P_{r}^f + \frac{\omega \psi P_{r}^d}{41^r} \right\} \]

(5.41)

\[ \Omega = \frac{\Upsilon_{MF} \times \left( \sum_{s=1}^{S} \left\{ P_{s}^f + \alpha \left( \frac{\omega \psi P_{s}^f}{41^a S} \right) \right\} \right) + \sum_{r=1}^{R} \left\{ P_{r}^f + \frac{\omega \psi P_{r}^d}{41^r} \right\}}{\sum_{l=0}^{L} (MCE_l \times \frac{B_l}{R}) \times \Upsilon_{SRF}} \]

(5.42)

Equation 5.42 provides metric to quantify the long term energy efficiency of the FD as function of number of sectors per site S, number of relay station per site R and and frequency reuse F.

Having developed metrics to quantify long time scale performance of WCS in all the three aspects of interest, in next section the numeric values for the metrics derived in this section are evaluated for set of possible FD's for WCS.

### 5.4 Long Term Performance of WCS's with different FD's

In this section, first we will discuss the results of ESE, SAF and EC to highlight the gains and respective tradeoffs in performances of different FDs offer (in both conventional WCS and Relay enhanced-WCS(R-WCS) separately). This is followed by a discussion on comparison of the performance of WCS and R-WCS in general.

#### 5.4.1 System Model for Performance Evaluation

Since, there are many potential candidate FDs for next generation WCS with different F, S and R, therefore, in order to evaluate and compare ESE, SAF and EC and the tradeoff between the them in various FD, total of 26 FDs with a
5.4. Long Term Performance of WCS’s with different FD’s

<table>
<thead>
<tr>
<th>FD parameters</th>
<th>Feasible S, F, R combinations usually used.</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>1, 2, 3, 4, 6</td>
</tr>
<tr>
<td>F</td>
<td>1, 2, 3, 4, 5, 6</td>
</tr>
<tr>
<td>R</td>
<td>0, 1</td>
</tr>
</tbody>
</table>

Figure 5.2: List of potential Frequency reuse and Deployment architectures (FD’s) for future WCS that are investigated in this chapter.

A wide range of F, S and R as listed in table in figure 5.2 are modeled through static system level simulations. The major system design parameters used in simulations of various FDs are given in table in figure 5.3. Two tiers of cells are modeled for each FD to consider realistic amount of interference in multi cellular scenario. Other real features, like shadowing and appropriate pathloss models for BS and RS considering both LOS and NLOS conditions similar to [119] are used in order to model a realistic WCS and R-WCS propagation environment. In R-WCS, RS are optimally located at half of inter site distance where the SINR.
5.4.2 ESE of Various FD’s for WCS

ESE is evaluated through two different methods. 1) Pragmatic: Based on the SINR thresholds for a set of modulation and coding schemes described in LTE standard used in [120], 2) Theoretical: i.e. based on (5.21).

5.4.2.1 ESE for Conventional FD’s

Figure 5.4 shows the ESE evaluated through simulations of multi cellular scenarios for 12 different FDs of WCS. The tradeoff among S, F and MCE can be seen playing its role in the overall ESE of different FD. For ease of discussion while...
probing into the underlying trends and tradeoffs we focus on FD 9-12, all with S=6. It can be seen that for FD=9 where full frequency reuse (F=1) is used, ESE is lowest and gap from single link Shannon bound is largest. This is due to high inter-sector interference which results in very low $T_{MCE}$ and hence low $T$. In FD=10 and 11 when F increases to 2 and 3, although $T_{SRF}$ decreases from 6 to 6/2 and 6/3 respectively, still the ESE increases. This is because the increase in $T_{MCE}$ due to decreased interference is more than the loss in $T_{SRF}$. Hence as a net result ESE is larger in FD=10, 11 compared to FD=9. But in FD=12, where F further rises to 6, the loss in ESE due to low $T_{SRF}$ (i.e 6/6=1) is much larger than the gain in $T_{MCE}$ through lower interference. This causes a lower ESE in FD=12 as a net result. On the other hand, the gap between practically achieved and theoretical ESE, monotonically decreases as F increases $T_{SRF}$ decreases in FDs 9-12, mainly because higher average SINR is yielded with larger F due to the decreased interference and at higher SINR theoretical spectral efficiency becomes saturated allowing practically achievable MCE to catch up.

5.4.2.2 ESE for Relay Enhanced FD's

Figure 5.5 shows theoretical and practical ESE evaluated for various FDs for R-WCS. By comparing the ESEs of R-WCS with those for WCS it can be easily seen that RSs bring a huge improvement in ESE. This improvement is due to two reasons.

First the gap between the practically achievable and theoretical ESE is reduced significantly in R-WCS compared to WCS. This is because of the fact that RS boost SINR distribution more effectively than higher frequency reuse can. This argument can be justified by comparing the SINR distribution of WCS and R-WCS in figure 5.6 and 5.7 respectively.

The relatively much better SINR distribution in R-WCS is mainly because of
much smaller height and lower transmission power of RS. This makes the interference caused by RS much lesser than that caused by the interfering sectors of BS. Secondly, in addition to better SINR distribution and hence higher $T_{MCE}$, there is another positive contribution of RS towards higher ESE that explained as follows: Lets assume 3 RS are working in a cell, the spectrum is divided into two parts for sharing between BS and RS thus reducing the $T_{SRF}$ by half only compared to scenario with three sectors as $T_{SRF}$ will reduce by factor of 3 in this case. These two reasons make RS more advantageous method to boost ESE because they can boost SINR and thus $T_{MCE}$ more effectively while causing relatively lesser decrease in ESE through $T_{MF}$ compared to F or S based method of improving SINR. This fact can be further confirmed by comparing the ESE for $FD=23$ to 26 in figure 5.5. As the F increases, figure 5.7 shows that SINR improves and thus the $T_{MCE}$ improves boosting the ESE. But the net ESE decreases
5.4. Long Term Performance of WCS's with different FD's

Figure 5.6: CDF of SINR distribution in the coverage area for different number of sectors per site and frequency reuse factors.

Figure 5.7: CDF of SINR distribution in the coverage area for different number of sectors and relay stations per site and frequency reuse factors.
because the $T_{SRF}$ decreases more than $T_{MCE}$ can increase through increase in F. Finally, it can be seen highest ESE is yielded by FD=23. This is so because it not only resorts to F=1 to achieve high $T_{SRF}$ but also avails better SINR distribution (see figure 5.7) than counterpart FD=9 due to the advantages of RS explained above.

5.4.3 SAF of Various FD’s for WCS

5.4.3.1 SAF for Conventional FD

Figure 5.8 shows the values of SAF evaluated for all the 12 DAs of WCS using (5.30). In general it can be noted that in WCS, SAF increases with increase in number of sectors but it decreases with increase in F (or in other words decrease in $T_{SRF}$). This is because increasing the number of sectors in general
5.4. Long Term Performance of WCS’s with different FD’s

decrease the cell edge interference thus makes SINRs geographical distribution more uniform in a cell. On the other hand a low $T_{SRF}$ has same effect but in different way. A low $T_{SRF}$ means the interfering cells are farther, thus making SINR distribution less dependent on distance from the cell center hence more uniform geographically.

5.4.2.2 SAF for Relay Enhanced FD’s

Figure 5.9: SAF for different number of sectors and relay stations per site and frequency reuse factors.

Figure 5.9 shows the SAF for all 14 FDs of R-WCS. It can be seen that although the trends with respect to $S$ and $F$ are the same as for WCS but in general SAF in R-WCS is significantly lower than that in WCS. The reason behind this is the drastic change in distribution of SINR brought by RS as can be seen in figure 5.7, the span of cdf of SINR in the R-WCS is much larger than that of
WCSs. This is because, although RSs improve the SINR, this improvement is not in the whole coverage area. Rather they provide an up shift in SINR in their own small coverage area only, leaving the rest of the coverage area served by sectors of BS unaffected. This increases the deviation of SINR values from the mean and hence the SAF decreases.

5.4.4 Energy Consumption of Various FD’s

Figures 5.10, 5.11 and 5.12 plot fixed, variable and total power consumptions respectively. \( \omega^s = \omega^r = 1 \) is assumed because we are considering full load scenario. Antenna efficiency of commercial antennas is used. i.e. \( \Gamma^r = \Gamma^s = 60\% \). \( P_f^s = 15W \) with \( P_f^r = 0.5P_f^s \) is used due to reasons explained in [113]. It is important to note that variable power consumption does not increase with number of sectors. This is because the additional gain due to higher directivity of sectorized antennas cancels out the additional power required to transmit on sectors. Figure 5.12 shows that power consumption per site increases more rapidly with the increase in number of RS (i.e. R) than in number of sectors per site (i.e S). This is mainly because each RS has an omnidirectional antenna, so there is no compensating factor as in case of sectors as explained in subsection 5.3.3.3.

5.4.5 Comparison of Performance of Conventional FD’s with Relay Enhanced FD’s

Results in figure 5.4-5.9 show that R-FD has potential for higher ESE but they have naturally low SAF. Whereas conventional FD although offer lesser ESE but have much higher SAF. So there is tradeoff between the ESE and SAF which can be exploited by adding RS. Furthermore, higher ESE of R-FD in general shows that with RS in place at the cell edges larger \( T_{MCE} \) without significant decrease in \( T_{SRF} \) can be obtained.
5.4. Long Term Performance of WCS's with different FD's

Figure 5.10: Fixed power consumption per site

Figure 5.11: Variable power consumption per site.
5.5 A Novel Self Organising Frequency Reuse and Deployment Framework (SOFD)

In this section, we present a novel self organising frequency reuse and deployment framework (SOFD) designed on principles of BSOF to cope with large time scale dynamics of WCS. We exploit the long term performance characterisation framework (PCF) developed in last sections and follow the steps of BSOF to achieve a SO solution.

5.5.1 Problem Formulation: Identifying SO-Objective

Having established the metrics to characterise the long term performance of WCS for the objectives of interest, identifying the SO-Objective is straightforward i.e. maintain optimal ESE, EC and SAF. As discussed above there are myriad of factors that determine the performance of cellular system in terms of these objec-
5.5. A Novel Self Organising Frequency Reuse and Deployment Framework (SOFD)

tives, but as explained in previous sections, over long time scale these factor can
be boiled down to three major factors 1) spectrum reuse factor $F$, 2) number of
sectors per site $S$, 3) number of relay station per site $R$. Hence the SO-Objective
can be written as

$$\max_{F,S,R} \{ \Upsilon(F,S,R), \Phi(F,S,R), \Omega(F,S,R) \}$$  \hspace{1cm} (5.43)

For sake of simplicity of expression, we will avoid expressing the variable $F$, $S$ and
$R$, however from the previous section it is understood that $\Upsilon, \Phi$ and $\Omega$ denote
ESE, SAF and EC respectively and are functions of $F$, $S$ and $R$ in this context.

It should be noted that these objectives are mutually contradicting due to inter
dependence of their controlling parameters as explained in previous sections. We
know that no FD is optimal simultaneously for ESE, SAF and EC. This makes
such problem non convex hence difficult if not impossible to solve with purely
analytical approaches. In next sections we show how to simplify this problem to
design SO-Goal and SO-Functions

5.5.2 Transforming SO-Objective into SO-Goal

We propose to transform the complex SO-objective in (5.43) into a simpler SO-
Goal through multi objective optimization as used in [121] by representing the
three objectives simultaneously as a single utility function. Since the mutual
priority of these objectives and their target values are strongly dependent on the
operator's policy [3] so we design the utility function $v$ to incorporate the operator
policy and preferences as well i.e.

$$v = \begin{cases} 
  v_g(\Upsilon, \Phi, \Omega) & \text{Case 1: General Optimization} \\
  v_t(\Upsilon, \Phi, \Omega) & \text{Case 2: Targeted Optimization} 
\end{cases}$$  \hspace{1cm} (5.44)

where the subscripts $g$ and $t$ denote the general and targeted cases respectively.
These cases and the respective utility functions are explained in the section below.
Chapter 5. Multi Objective Optimisation through SO of Frequency Reuse & Deployment Architecture

5.5.2.1 Case 1: General Optimisation

This case represents a scenarios where the operator of WCS does not have any specific target values for the performance aspects:

In this case the SO-Goal will be

$$\max_{\mathcal{F}, \mathcal{S}, \mathcal{R}} \nu_2 (T, \Phi, \Omega) = \max_{\mathcal{F}, \mathcal{S}, \mathcal{R}} (\lambda_1 T + \lambda_2 \Phi - \lambda_3 \Omega)$$

(5.45)

This utility function is flexible to adjust mutual priority of these objectives, below we present some exemplary rules to manifest the adaptation of mutual priority among the objectives.

Rules to adapt Utility

1. If the system does not have any priority among objectives in (5.45) set

$$\lambda_1 = \lambda_2 = \lambda_3 = 1/3$$

(5.46)

2. If system wants to maximize some objectives, while neglecting others, In (5.45) set

$$\lambda_i = \begin{cases} 1 & \text{if } i = d, j = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases}$$

(5.47)

where $d$ is the index representing desired objective.

3. If system has specific priority of each objective, system can represents its priority by the weights such that

$$\lambda_1 + \lambda_2 + \lambda_3 = 1$$

(5.48)

5.5.2.2 Case 2: Targeted Optimization

This case represents the scenario where the operator has specific target values for each performance aspect.
5.5. A Novel Self Organising Frequency Reuse and Deployment Framework (SOFD)  

In this case the SO-Goal can be written as

$$\min_{F,S,R} u_t (T, \Phi, \Omega) = \min_{F,S,R} \left[ \sqrt{\lambda_1 (T - T_i)^2 + \lambda_2 (\Phi - \Phi_i)^2 + \lambda_3 (\Omega - \Omega_i)^2} \right]$$  \hspace{1cm} (5.49)

Rules for Utility Adaptation

In this case the rules for utility adaptation would be

1. If system wants to achieve desired targets in each metric with same priority, substitute (5.46) in (5.49)

2. If system has desired target value in one objective, but has no priority in others substitute (5.47) in (5.49)

3. If system has specific values of each metric as target but has different priority of each target to be met, substitute (5.48) in (5.49)

As discussed in previous sections, very intricate mutual dependency of the optimization objectives makes the analytical solution of the problem in (5.43) very difficult. Having designed a more tangible SO-Goals in (5.45) and (5.49) we can solve the rest of the problem semi-analytically i.e. since there are only finite combinations of $F$, $S$ and $R$ that are technically viable as shown in table 5.2, so the search space of optimization problem is finite and small. By evaluating all these FD's performance in terms of long term performance metrics $T$, $\Phi$ and $\Omega$ as function of $F$, $S$ and $R$ in section 5.4 the required solution space has already been established. This very small solution space can be searched easily to manifest the SO-Goals in (5.45) and (5.49).

Having developed the SO-Goal, in next section we discuss how this SO-Goal can be achieved through some practical SO-Functions and hence SOFD can be implemented practically.
5.5.3 Identifying SO-Functions

From practical implementation point of view SOFD, the required SO-Functions that each BS should be capable to execute without human intervention, in order to achieve desired SO-Goals and hence SO-Objective, are

1. Adaptation for BS antenna patterns to change the number of sectors projected

2. Adaptation of frequency reuse scheme

3. Capabilities to switch on or off remote RS

4. Capability to switch on or off the circuitry associated to each sector and RS, within BS.

Emerging WCS feature highly intelligent BSs and each site can have smart antennas and remotely controllable RS. Therefore, changing the radiation pattern of particular antenna or switching on or off a whole sector or RS should not be an issue. Given the fact that SOFD has to be executed usually on large time scale, makes execution of these SO-Functions even more pragmatic.

5.5.4 Practical Implementation

The main idea of proposed SOFD framework is that, in order to cope with long time scale dynamics in conjunction with operator's policy, WCS can switch to a optimal FD scheme in SO manner based on the adaptive utility functions (5.45) or (5.49). These utility functions are designed to reflect operators policy. It is important to highlight here that adaptation of these utilities can be according to predefined empirical rules as listed in
section 5.5.2 or it can be acquired through the learning techniques listed in figure 2.3.

The implementation of SOFD can be both, distributed or centralised. A distributed implementation would require additional capabilities at each BS e.g learning and cognitive decision making with locally gathered intelligence. Such distributed implementation may have limitation in terms of stability as different number of sectors and frequency reuses in neighbouring cells may have negative impact on system wide smooth operation of SOFD. Therefore we recommend centralised implementation to ensure stability. It should be noted that time scale of SOFD execution is so large that even with centralised implementation, required level of agility can be easily achieved. Scalability is also retained because of the large time scale and negligible amount signalling required between the central control and each BS in the system. Signalling is negligible, because central control only need to transmit to each BS an identifier for new FD whenever a switching of FDs is required in WCS. Since total number of FD is at most in the order of few tens (26 in table 5.2 ) so this translates to a few bits of signalling over large period of time. For example, for the set of FD's considered and evaluated here, only \( 2^5 = 32 > 26 \) 5 bits need to be transmitted by the central controller to each BS whenever a change in FD is required. This is almost negligible given the time scale of operation of SOFD, making SOFD scalable as well as agile.

Finally, in next section we present some numerical results to illustrate the operation of SOFD.
Chapter 5. Multi Objective Optimization through SO of Frequency Reuse

5.5.5 Numerical Results

Figure 5.13 and 5.14 show the normalized values ESE, SAF, and EC worked out through 5.29, 5.30 and 5.42. For the ease of plotting on the same scale, values of each metric are normalized by their maximum for both cellular and relay enhanced cellular network respectively. These two graphs form the solution space for the problem i.e. SO-Objective in (5.43). Before we explain the use of this solution space for SOFD it is important to highlight some interesting tradeoffs we can observe among the performance metrics or optimization objectives in figure 5.13 as well as figure 5.14. Figure 5.13 shows that FD=1 is optimal w.r.t. energy efficiency, but has suboptimal spectral efficiency and worst fairness. Compared to FD=1, in FD=2 spectral efficiency and fairness both improve but at the expense of more energy
5.5. A Novel Self Organising Frequency Reuse and Deployment Framework (SOFD)

![Figure 5.14: ESE, SAF and EC normalized by their respective maximum value in relay enhanced WCS](image)

Figure 5.14: ESE, SAF and EC normalized by their respective maximum value in relay enhanced WCS

consumption. FD=3 provides some gain over FD=2 in terms of spectral efficiency as well as energy efficiency but at a heavy expense of fairness and so on. In figure 5.13 it can be seen that relays bring in an additional factor in this tradeoff.

In summary it can be seen that no single FD strategy is optimal in all the three performance metrics simultaneously. In other words no single FD can meet all objectives together. Rather each FD is optimal in a particular sense. This is where SOFD provides a useful solution.

As explained in section 5.5.2, depending on the current system requirements, SOFD will select an appropriate utility i.e. either (5.45) or (5.49). Then it will set the weighting parameters to reflect the operator's priorities among the desired objectives. If the FD that optimizes the utility, in the
Figure 5.15: Value of general optimization utility function $v_g$ for different priorit-

Figure 5.16: Value of general optimization utility function $v_g$ for conventional FD's with different priorities in terms of ESE, SAF and EC.
5.5. A Novel Self Organising Frequency Reuse and Deployment Framework (SOFD)

Figure 5.17: Value of targeted optimization utility function $v_x$ for convectional FDs with different target values of ESE, SAF and EC.

worked out solution space, is not the systems current FD mode, the system will switch to the optimal FD through the SO-Functions explained in section 5.5.3. On next trigger of poor performance e.g. a seasonal event, chorionic high blocking, blocking or poor fairness, or power shortage alarm, system will repeat same process to go to the new FD mode which is optimal to achieve target objectives under systems newly changed state.

Figure 5.15 and 5.16 plot utility $v_y$ for four sets of different objective priorities. To avoid repetition of ideas, we will explain operation of SOFD framework using results of relay enhanced WCS only (i.e. figure 5.16). If FD schemes without relay are also included, the process will essentially remain the same only the search space will become larger.

With equal priority of all three objectives, we can see FD=23 in figure 5.16 is optimal choice. When spectral efficiency has highest priority i.e. 80% and fairness and energy efficiency has lower and equal priorities of 10% each, FD=23 is the optimal state again. On the other hand, when fairness has
Figure 5.18: Value of targeted optimization utility function $v_t$ for relay enhanced convectional FDs with different target values of ESE, SAF and EC.

highest importance i.e. 80%, and the spectrum and energy efficiency have lower and equal priorities of 10%, FD=17 becomes optimal state. When energy efficiency is most important target with 80% importance factor, and fairness and spectral efficiency are lower priorities with importance of just 10%, SOFD will switch WCS the system to FD=23.

Figure 5.17 and 5.18 plot $v_t$ for three different set of target values of the three objectives, each having same priority i.e $\lambda_1 = \lambda_2 = \lambda_3 = 1/3$. First case (blue) represents the scenario when system wants spectral efficiency and fairness both be closes to their optimal values 100% but have some flexibility in energy efficiency. In this scenario SOFD frame work will switch to FD=23. In second case (red), EC is needed to be closest to optimal, followed by spectral efficiency followed by fairness. Now the FD=24 is the optimal solution. In the last case (green), when fairness need to be closest

$^2$Although discussion is valid to conventional FD's as well, but we focus on relay enhanced FD's i.e. figure 5.18
5.6 Multi-Objective Optimization Vs Constrained Optimization

In this chapter we used multi-objective optimization to tackle more than one optimization objectives at same time. The problem could also be formulated as constrained optimization problem by taking one objective as optimization function and the rest of the objectives as constraints. There are two main reasons multi-objective optimization approach was used instead of constrained optimization

- Formulating all the objectives as optimization functions allows more control over the optimality in terms of all the objectives compared to scenario when only one objective is optimised whereas other can lie any where in the feasible region allowed by the constraints. This extra control is handy to reflect the operators policy more precisely in system optimization.

- Absence of constraints allowed us to neglect non convexity of our problem without worrying for the duality gap (see Appendix D). This is because we did not have to formulate any dual problem using lagrangian multipliers to incorporate the constraints. Instead we were able to tackle the problem directly with combinatorial optimization.

Further discussion on the relevant aspects of optimization theory can be found in Appendix D
5.7 Conclusions

Compensating the effects of the long term dynamics is one of the major use case of SO. Remarkably less attention has been given to this use case, compared to the attention that solutions for very short to medium term dynamics have received.

The optimization of long term performance objectives like spectral efficiency, service area fairness and energy consumption, through deployment architecture factors is also a widely unexplored area. This chapter has attempted to fill this gap and presented two novel frameworks to address long term dynamics of SO through SO of Frequency reuse and Deployment (FD) architecture.

First a novel performance characterisation framework (PCF) is developed to characterise long term performance of WCS as function of FD design parameters. PCF consists of a set of three novel metrics to characterise performance of WCS in terms of spectral efficiency, service area fairness and energy consumption. Performance of a variety of potential FD's for future WCS have been evaluated using these metrics through extensive numerical simulations. Results show that no single FD is optimal simultaneously in terms of all the WCS performance aspects. Rather different FD's provide a different level of trade off among these performance objectives showing their close coupling.

This observation is used as a key rational to propose and design novel self Organising Frequency reuse and Deployment (SOFD) framework.

SOFD builds on PCF and performance evaluation results obtained through semi analytical evaluations, and exploits the principle of BSOF to achieve SO solution for adaptation of FD's.

SOFD can enable WCS operators to achieve and maintain multiple long
term objectives through autonomous switching to FD that is optimal in face of changing dynamics and requirements of WCS operation.

The main advantageous features of SOFD are its low complexity of operation, effectively negligible inter site signaling and potential to meet multiple designated objectives over long term efficiently.

The key advantage of SOFD are huge improvement potential for WCS performance by avoiding the wastage of radio and energy resources. With current fixed FD based WCS, such wastage is inevitable due to continual occurrence of socio-economical and seasonal changes in user demography. Finally, SO-Functions proposed for SOFD can also help to save a lot of operational cost by avoiding the need for manual updates of FD's.
Chapter 5. Multi Objective Optimization through SO of Frequency Reuse & Deployment Architecture

206
Chapter 6

Conclusions and Future Work

In this chapter we finally conclude this thesis by providing a conclusive summary and elaborative directions for the future work.

This chapter is organised as follows. In section 6.1 we present a conclusive summary of the thesis. Section 6.2 summarises the key achievements of this thesis. Identifying the implications and directions for further improvements in the works presented in this thesis, in section 6.3 we provide specific research directions that can lead to more promising results by building on our work. In section 6.4, we highlight some general directions for future work that are indirectly motivated by the work presented in this thesis. Finally in section 6.5 we discuss some interesting non deterministic problem solving approaches, that can complement deterministic approach taken in this thesis to design SO.

6.1 Conclusions

The main objective of this thesis has been to develop self organizing performance enhancement algorithms for future WCS to cope with short to long
time scale dynamics.

In chapter 1 we presented the key motivation behind this study and clarified the scope of this thesis by classifying the dynamics of cellular system. Given all the ambiguity associated with the term self organization, we devoted chapter 2 to discuss out these ambiguities and laid down a solid foundation for the rest work presented in this thesis. BSOF and characteristics of SO, were the key contribution of this chapter that paved way for rest of the chapters.

In chapter 3, unlike previous works that addressed hotspot problem through load balancing, a novel approach of optimizing spectral efficiency at the hotspots is used. This approach has advantage that it does not necessitate handovers that are required in load balancing based hotspot relief. Being developed on principles of BSOF, the developed solution features SO and its gain is demonstrated through system level simulation and compared against number of available bench marks. Over 30% gain in spectral efficiency was observed with no cost in terms of signalling.

In chapter 4, the problem of congestion arising from the medium time scale dynamics was addressed and SO solution namely LB-BSOF was developed. The problem was formulated as global blocking minimization problem and was solved analytically with help of a set of mathematical propositions. A novel concept of super cell is introduced and proved to make the solution distributed and localised hence scalable and agile. Practical implementation and use cases of LB-BSOF were also discussed in detail. Being designed on the principle of BSOF, LB-BSOF bore all features of SO. Finally the performance LB-BSOF was evaluated through extensive system level simulations and was compared against, no load balancing scenario and an ideal central control based load balancing algorithm that is neither scalable not agile. Performance was evaluated for conventional macro cell based WCS
with 3 and 6 sector scenarios, as well as heterogeneous WCS with Femto cells and RS. A reduction in blocking by 280% compared to no load balancing scenario was observed. In most cases, LB-BSOF was shown to have same reduction in hard blocking as ICC could provide with an additional advantage that LB-BSOF has relatively lesser negative impact on interference compared to ICC.

Chapter 5, focused on the long time scale dynamics that are most neglected in literature. Two main contributions were presented in chapter 5. First in order to quantify long term performance of WCS a novel performance characterisation framework consisting of three novel metrics was developed. These metrics were derived to capture the performance of WCS by incorporating its long term design features of WCS i.e. frequency reuse, number of sectors, number of relay stations per site and the set of modulation and coding schemes. A novel self organising frequency reuse and deployment framework, namely SOFD, building on the long time scale performance metrics. SOFD is built on the principle of BSOF to ensure scalability, agility and stability and promises a significant performance and profit improvement.

### 6.2 Summary of Key Achievements

The major novel contributions and achievements accomplished in this thesis can be summarised as follows:

(a) Identification of basic characteristics of SO systems and their underlining principles leading to these characteristics.

(b) A novel biomimetic Self Organization Framework (BSOF) for enabling and designing SO in a generic engineering systems.

(c) A novel framework for SO of system wide antanna Tilts (TO-BSOF) for optimization of spectral efficiency and throughput with focus on
Chapter 6. Conclusions and Future Work

short term dynamics

(d) A novel framework for SO for optimal Load Balancing (LB-BSOF) to minimize blocking and maximise user satisfaction with focus on medium term dynamics for WCS.

(e) A novel Performance Characterisation Framework for WCS (PCF)

(f) A novel framework for SO of Frequency reuse and Deployment (SOFD) of WCS for optimization of multiple objectives of spectral efficiency, service area fairness and energy efficiency with the focus on long term dynamics.

6.3 Specific Future work Directions

The work presented in this thesis has not only introduced some novel SO solutions but it has also opened a number of avenues for the future work. Below, we discuss the potential future works that can directly build on the work presented in this thesis.

6.3.1 Future work on TO-BSOF

- TO-BSOF presented in chapter 3 focused on optimization of single objective of spectral efficiency. A subtle tradeoff between spectral efficiency and fairness was observed in the performance evaluation results of TO-BSOF. It would be of merit to extend TO-BSOF for joint optimization of spectral efficiency and fairness using the multiobjective optimization paradigm as exploited in chapter 5.

- Investigation of impact of TO-BSOF on energy consumption is also an interesting extension of TO-BSOF.
6.3. Specific Future work Directions

Although the analysis that lead to TO-BSOF would remain unchanged in case of mobile hot spots, but including the mobility in the performance evaluation scenario to observe the agility of proposed solution in action would be an interesting aspect as well.

- TO-BSOF focused on dynamic and distributed optimization of one parameter of antenna i.e. tilt only. Extending the similar triplet based distributed framework to include azimuth into the optimization objective is also challenging task.

- TO-BSOF was developed and evaluated in context of macro cellular scenario but it is also applicable to scenario, where, hot spots are not directly served by BS rather they are served by RS. In this case TO-BSOF can be used to optimize the spectral efficiency at the access link of the RS assuming location of RS as the center of gravity. Such extension may prove handy to sort out the access link congestion problem in relay enhanced cellular networks as highlighted in section 4.6.4.2.

- Extension of TO-BSOF to scenario with 6 sector based WCS is also straightforward direction for future work.

6.3.2 Future work on LB-BSOF

- In chapter 4, the analysis that lead to LB-BSOF, assumed circuit switching with constant bandwidth channel allocation. Although optimality conditions obtained were generic and yielded performance very close to the optimal even with variable bandwidth allocation, it would be challenging yet an interesting task to extend this analysis to packet switching based traffic with variable bandwidth allocation to find the exact optimality conditions in such scenario. Comparing the performance based on that analysis, with that based on analysis in chapter 4 will be also interesting.
A potential of LB-BSOF to reduce the power consumption was observed in performance evaluation section (figure 4.17). A further work to investigate and improve the energy efficiency aspect of LB-BSOF would also be a fruitful direction.

Scope of LB-BSOF presented in chapter 5 is limited to medium to long term dynamics as it does not include any mechanism to deal with handovers. The fact that LB-BSOF is distributed and have very low implementation complexity makes it resourceful for short term dynamics as well, if proper mechanism are incorporated in it, to deal with resulting high frequency of handovers.

Evaluation of LB-BSOF was done with two selected SO-Functions only. Development of simulator or emulator that can model all four SO-Functions to evaluate full potential of LB-BSOF identified, can be challenging yet fruitful future work.

In chapter 4 we presented elaborative use cases of LB-BSOF. Only the use case of LB-BSOF for SO at micro level was demonstrated. Rest of the cases are discussed but not demonstrated because of lack of evaluation tools and high complexity of modelling. This leaves significant room for future work on LB-BSOF to extend it to rest of the use case discussed there.

6.3.3 Future work on SOFD

In chapter 5, the Performance Characterisation Framework (PCF) that was used to enable SOFD has huge potential for extension by making it online. i.e. in this case we demonstrated this framework through system level simulations, but in real WCS it can be implemented in online manner, where ESE, SAF and PC can be calculated based on
user feedbacks and power consumption logged at each BS. Furthermore, a similar approach, as used to develop ESE, SAF and PC, can also be used to develop other performance metrics to capture variety of aspects of WCS performance and hence enable SO.

- In chapter 5, numerical results for PCF were determined assuming that each bin in the coverage area has same importance. However it was highlighted there that PCF has potential to reflect the more realistic scenarios where different parts of the coverage area have different importance e.g. parks have low importance in determining WCS than the service level in town center. It would be interesting to extend SOFD for such more realistic scenarios.

- SOFD was evaluated assuming a centralised implementation where BS in the system switch their state to same FD simultaneously. While this approach was chosen for its ensured stability, it would be interesting to investigate SOFD with distributed implementation relying on stochastic optimization or doctive learning approaches (explained in section 6.5.4).

- SOFD was developed and evaluated assuming integer frequency reuse that is most commonly used so far. An interesting direction for future work would be to extend this framework to consider fractional frequency reuse instead of fixed frequency reuse. This would increase the complexity of solution, but can be equally beneficial in terms of performance gain as fractional frequency reuse provides more flexibility in trade off between spectral efficiency and spectrum reuse efficiency compared to integer frequency reuse.

- With SOFD extended for fractional frequency reuse, the solution space will not remain small, and the semi analytical approach used in chapter 5 might not be viable. In such case non deterministic or learning
approaches might be exploited and are discussed in next section

6.4 Some General Research Issues in SO

In this section we discuss some general open research issues in context of SO that are motivated by implications of the work presented in this thesis.

6.4.1 Parallel Operation of Multiple Time Scale SO

In future WCS, multiple SO algorithms might be required to achieve different objectives. These algorithms would be operating simultaneously but at different time scales as explained in figure 1.1. There can be scenarios, when there are some common parameters of control among these different time scale algorithms. There would be dire need to investigate their mutual dependency and effect of potential overlapping among these algorithms that might cause instability.

6.4.2 Parallel Operation of Multiple Spatial Scope SO

In addition to having different scope in time, various SO algorithms might have different scope in space as well and there is a need to investigate the stability issues arising from their concurrent operation and designing appropriate coordination mechanisms or interfaces to ensure a holistic smooth operation of SO in future WCS. e.g. in given WCS, one SO algorithm, say for energy optimisation, may need to operate over large number of nodes cooperatively and another SO algorithm in the same network, say for capacity optimization, might be operating locally. Since these two SO algorithms might have to play with same parameters of each node, say transmission power, proper coordination among them is required to avoid instability.
6.4.3 Challenges in Evaluation of Long Time Scale SO

While SO requires a significant upgrade over the classic adaptive algorithms, tools to validate and evaluate SO algorithms also require an upgrade. Classical system level simulator are usually designed to evaluate short term adaptive algorithms, e.g. scheduling, Transmission Control Protocol (TCP), power control etc. The main problems in research on medium to long time scale self organisation emanates from difficulty of modelling and simulating the environmental variation, technical anomalies over large time scale. In order to assess the performance of SO functionalities like, self configuration, multi objective self optimization, or self healing, different kind of simulation and evaluation methodologies are required. The most eminent feature of such simulation tools would be their capability to capture much larger picture of the system both in time and space. Secondly such simulators also need to have potential to model variety of potential SO-Functions, like beam switching and FD adaptation.

6.4.4 Enabling Self Organization

With a lot of activity on SO solutions, attention also needs to be channelled towards enabling SON algorithms/solutions. Enabling SO is a key as we seek to implement SO in future networks through a gradual evolution from reduced human intervention to minimal human intervention and finally one day zero human intervention in deploying optimising and managing future cellular networks. Key enablers for SO, would be seamless and scalable algorithms for autonomous coverage and service estimation, autonomous cell boundary estimation, autonomous energy consumptions and QoS estimation. All these estimation algorithms will essentially provide the necessary data to trigger appropriate SO eliminating the need for human intervention
making future wireless systems truly SO. Determining the right Key Performance Indicators (KPIs) for SO systems is also an open and grey area so far and would be one of the first step toward enabling SO.

6.5 Alternative Design Techniques towards SO

Having identified the specific and general research issues for future work in this section we discuss some interesting alternative techniques that can be used to design SO solutions.

6.5.1 Use of Non Deterministic Techniques towards Design of SO

In chapter 2 we presented the generic biomimetic framework for design of SO i.e. BSOF shown in figure 2.6 (section 2.7). It was highlighted that, in natural systems usually both nature and nurture go hand in hand to implement BSOF. But in engineering systems, any of problem solving techniques listed in figure 2.3 can be used to follow the steps of BSOF. In this thesis throughout we have used deterministic techniques. But there may be problems where deterministic techniques are not possible to use particularly when the architecture of the system is not clear and system model optimization parameters are abstract e.g for self configuration and self healing phases. In such scenarios, it would be very interesting to venture on the stochastic approach and learning based problem solving approach to transform from one step of BSOF to other. Particularly, when designing solutions for self configuration and self healing, learning based approaches become more viable because of their flexibility to evolve and mimic intelligence. The learning
algorithms can be further classified broadly as supervised and unsupervised learning algorithms.

6.5.2 Supervised Learning

Supervised learning involves learning as a result of the training received from a teacher. There is usually a desired target system response and the trained network gives the input - output mapping by minimising a defined cost function [122]. Albeit, when such systems are faced with new set of inputs they could choose wrong actions which would lead to poor performance.

6.5.3 Unsupervised Learning

Unsupervised learning sometimes referred to as self organised learning [122] involves learning process without a teacher (training sample). The system optimises its parameters based on interactions with its inputs. This is based on principal component analysis (PCA) or through clustering of similar input patterns. A major problem with unsupervised learning is uncertainty about the convergence time and hence agility and stability of the solution.

In addition to the classic supervised and unsupervised learning there are relatively new emerging paradigms that are hybrid of the the two approaches towards learning. Below we briefly explain these promising learning paradigms that can be exploited for design of SO.

6.5.4 Docitive Learning

Docitive Learning (DL) is a new emerging paradigm recently addressed in [37–39] and promises an improvement over classical supervised learning
in terms of complexity and convergence time. The improvement comes from the fact that instead of following a pre-assigned fixed supervisor, the learning node can dynamically choose its supervisor among set of available nodes based on the correlation in its own operational environment and environment of other accessible nodes. Thus the learning node can choose node(s) with a high degree of correlation as role model rather than classic supervisor, as is the case of conventional supervised learning. Thus the learning already acquired by the nodes with high degree of correlation with learning node, can be used as it is, saving a lot of time and avoiding higher complexity. Although much promising, DL is yet largely an unexplored realm and has huge potential for improvements.

6.5.5 Reinforcement Learning

RL involves attaching a reward and penalty scheme for each action that helps a learning agent to characterise its own performance. The learning agent tries to minimise the penalty received in each iteration thus improves/learns to take actions that minimises a cost function. A number of authors have found RL suitable for developing self optimisation algorithms [123], [124] and [55].
Appendix A

Proof of Center of Gravity Proposition

In order to prove proposition 1 we show that

$$\bar{T}^n = \eta_{\text{max}} = \frac{\sum_{\forall k \in \mathbb{K}^n} \log_2(1 + \bar{T}^n_k)}{\log_2(1 + \gamma_0^n)}$$  
(A.1)

if

$$\int_x \int_y \left( (\bar{T}^n_{x,y} - \bar{T}^n_{\text{tilt}}) \gamma^n_{x,y} \right) dx dy = 0$$  
(A.2)

where $\bar{T}^n_k$ is post optimization SIR in $n^{th}$ sector at $k^{th}$ point given by (3.14). Rest of the symbols have same definition as given in section 3.3. The sum of SIR for all the $|\mathbb{K}^n|$ users ($\forall k \in \mathbb{K}^n$) in $n^{th}$ sector will be given as:

$$\gamma^n = \sum_{k=1}^{|\mathbb{K}^n|} \left( \frac{d^n_{k}^{m} - \beta 10 \left( \lambda_c \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_0} \right)^2 + \lambda_h \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_h} \right)^2 \right) \right)} {\sum_{\forall n \in \mathbb{N} \setminus \mathbb{N}} \left( d^n_{k}^{m} - \beta 10 \left( \lambda_c \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_0} \right)^2 + \lambda_h \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_h} \right)^2 \right) \right) \right)$$  
(A.3)

In order to maximise this sum SIR over the tilt angle of the $n^{th}$ sector:

$$\frac{\partial \gamma^n}{\partial \theta_{\text{tilt}}} = \frac{\partial}{\partial \theta_{\text{tilt}}} \left( \sum_{k=1}^{|\mathbb{K}^n|} \left( \frac{d^n_{k}^{m} - \beta 10 \left( \lambda_c \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_0} \right)^2 + \lambda_h \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_h} \right)^2 \right) \right)} {\sum_{\forall n \in \mathbb{N} \setminus \mathbb{N}} \left( d^n_{k}^{m} - \beta 10 \left( \lambda_c \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_0} \right)^2 + \lambda_h \left( \frac{\sigma^n_{k} - \sigma_k^{\text{tilt}}} {B_h} \right)^2 \right) \right) \right) \right) = 0$$  
(A.4)
Appendix A. Proof of Center of Gravity Proposition

\[
\frac{\partial \gamma^n}{\partial \theta^n_{\text{tilt}}} = \sum_{k=1}^{\mathcal{N}} \left( \frac{1}{\sum_{n \in \mathcal{N} \setminus \mathcal{N}} \left( d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right) \right) - \frac{1}{\sum_{n \in \mathcal{N} \setminus \mathcal{N}} \left( d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right) \right)} \right) = 0
\]  
(A.5)

\[
\frac{\partial \gamma^n}{\partial \theta^n_{\text{tilt}}} = \frac{\partial}{\partial \theta^n_{\text{tilt}}} \left( d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right) \right)
\]  
(A.6)

\[
= d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right) \ln 10 \frac{\partial}{\partial \theta^n_{\text{tilt}}} \left( -1.2 \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right) \right)
\]  
(A.7)

\[
= C \left( \theta_k^n - \theta_n^{m} \right) d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right)
\]  
(A.8)

\[
C = \frac{2.4 \ln 10 \lambda_n}{B^2}
\]  
(A.9)

Putting \( v \) back in (A.5)

\[
\frac{\partial \gamma^n}{\partial \theta^n_{\text{tilt}}} = \sum_{k=1}^{\mathcal{N}} \left( \frac{C \left( \theta_k^n - \theta_n^{m} \right) d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right)}{\sum_{n \in \mathcal{N} \setminus \mathcal{N}} \left( d_k^{m-\beta} 10^{-1.2} \left( \lambda_n \left( \frac{\theta_k^n - \theta_n^{m}}{B_n} \right)^2 + \lambda_h \left( \frac{\phi_h^m - \phi_n^m}{B_h} \right)^2 \right) \right)} \right) = 0
\]  
(A.10)

\[
\frac{\partial \gamma^n}{\partial \theta^n_{\text{tilt}}} = \sum_{k=1}^{\mathcal{N}} \left( \theta_k^n - \theta_n^{m} \right) \gamma^n_{k,n} = 0
\]  
(A.11)

if user distribution is perfectly uniform i.e. \( \rho = \frac{\mathcal{N} \left[ \mathcal{N} \right]}{\mathcal{N} \left[ \mathcal{N} \right]} = 1 \) we can replace summation in (A.11) with the surface integral over whole area making it independent of user locations , i.e.

\[
\int_x \int_y \left( \theta_k^n - \theta_n^{m} \right) \gamma^n_{k,n} \, dx \, dy = 0
\]  
(A.12)
Now $\tilde{\hat{\eta}}^n$ obtained from above equation, maximises the sum SIR i.e. $\gamma^n$ in the coverage area, since $\eta = f(\gamma)$ where $f$ is convex and monotonically increasing function, hence proposition 1.
Appendix A. Proof of Center of Gravity Proposition
Proof of Optimality condition for Minimal Blocking in Homogenous WCS

The blocking will be minimum or maximum if all partial derivatives of $\bar{B}$ with respect to traffic in each cell are zero, i.e.

$$\frac{\partial \bar{B}}{\partial T_n} = 0; \quad \forall \ n = 1, 2, 3...N \quad (B.1)$$

Solving for first cell, i.e. $T_1$

$$\frac{1}{T} \frac{\partial}{\partial T_1} \sum_{n=1}^{N} \left( \frac{T^n_M}{\sum_{m=0}^{M} \left( \frac{T^n_m}{m!} \right)} \times T_n \right) = 0 \quad (B.2)$$

$$\frac{\partial}{\partial T_1} \left\{ \frac{T^{M+1}}{M!} \sum_{m=0}^{M} \left( \frac{T^n_m}{m!} \right) + \frac{T^{M+1}}{M!} \sum_{m=0}^{M} \left( \frac{T^n_m}{m!} \right) + \frac{T^{M+1}}{M!} \sum_{m=0}^{M} \left( \frac{T^n_m}{m!} \right) + \cdots + \frac{T^{M+1}}{M!} \sum_{m=0}^{M} \left( \frac{T^n_m}{m!} \right) \right\} = 0$$

By taking the derivative of each term we get:

$$\frac{\partial}{\partial T_1} \left\{ \frac{T^{M+1}}{M!} \sum_{m=0}^{M} \left( \frac{T^n_m}{m!} \right) \right\} = 0 \quad (B.3)$$
Appendix B. Proof of Optimality condition for Minimal Blocking in Homogenous WCS

As $M$ is supposed to be large number in OFDMA based systems i.e. large number of channels(sub-channels) are available per cell (this assumption is particular true for OFDMA based systems since much larger number of channels are available per cell compared to legacy TDMA-FDMA systems). Therefore, for mathematical traceability, we can use the Taylor series approximation here. The equation (B.3) can then be written as:

$$\frac{\partial}{\partial T_1} \left\{ \frac{T_1^{M+1}}{e^T_1} \right\} = 0$$

Taking the derivative:

$$\frac{1}{M!} \left\{ \frac{(M + 1)T_1^M e^{T_1} - T_1^{M+1} e^{T_1}}{(e^{T_1})^2} \right\} = 0$$

$$\left\{ \frac{(M + 1)T_1^M - T_1^{M+1}}{e^{T_1}} \right\} = 0$$

$$(M + 1)T_1^M - T_1^{M+1} = 0$$

$$(T_1^M)( (M + 1) - T_1) = 0 \quad \text{(B.4)}$$

(B.4) implies that either

$$T_1^M = 0 \quad \text{(B.5)}$$
or

$$(M + 1) - T_1 = 0 \quad \text{(B.6)}$$

Since (B.5) can not be true for reasonable values of $T_1$ and $M$, hence the valid conditions for optimal blocking can be found through (B.6) as follows

$$T_1 = M + 1$$
By second derivative test it can be shown that critical point represented by (B.6) is a minimum. Similarly by putting the partial derivatives of (4.6) with respect to traffic in other cells equal to zero, we get

\[
T_1 = M + 1 \\
T_2 = M + 1 \\
T_N = M + 1 \\
\Rightarrow T_1 = T_1 = \ldots = T_N = \frac{T_1}{N} = \bar{T}
\]  

(B.7)
Appendix B. Proof of Optimality condition for Minimal Blocking in Homogenous WCS
Difference between Area Spectral Efficiency and Effective Spectral Efficiency and

The difference between Area Spectral Efficiency (ASE) and Effective Spectral Efficiency (ESE) is as follows:

Area spectral efficiency is calculated measuring throughput of whole system and dividing it by the whole area.

For ESE, we do not measure throughput of whole system, rather we divide whole area in finite small bins, and estimate the spectral efficiency achievable in each bin, and calculate mean value of spectral efficiency over all bins.

The estimation of the spectral efficiency can be theoretical i.e. by mapping the SINR available in each bin to the spectral efficiency using Shannon equation. Or practically it can calculated using the threshold SINRs for finite set of modulation and coding schemes used in any standard e.g. LTE has 15 different modulation and coding pairs.
Appendix C. Difference between Area Spectral Efficiency and Effective Spectral Efficiency and

The *mean* can be calculated by giving all bins equal weights i.e. as expected spectral efficiency, or by weighting area bins of more importance (e.g. high streets, dense business areas) with higher weights than that of bins of low importance (parks, rural or unpopulated areas).

Note that essential units for ESE are b/s/Hz. The further denominator per cell, per site, per sector or per $km^2$ is optional and placed to indicate the area over which bins spectral efficiency is averaged. If this area is large enough that it reuses spectrum, e.g. per site. Then we multiply the average spectral efficiency per bin with frequency reuse factor over that area e.g. 3 if the site has three sectors and each of them using same spectrum. This can provide an over all measure of spectral efficiency and spectrum reuse efficiency in that area as a single figure.

Advantages of ESE: There are following main advantages of ESE over classic ASE

(a) ESE is a more effective and informative and flexible measure than ASE as it can weigh the area bins proportional to their importance to give more useful measure of quality of coverage.

(b) Unlike ASE, ESE does not require calculation of system wide throughput, i.e. ESE does not require dynamics simulations including scheduling and user distributions, fast fading models. Rather it can be calculated through static simulators i.e. by modelling the deployment architecture of the system to determine geometric SINR only.

(c) ESE is independent of the radio resource scheduling, user scheduling and user geographical distribution in the system that otherwise effects the throughput and hence the ASE measure.

(d) Advantage (b) and (c) also makes ESE independent of short term dynamics of system e.g. user mobility, temporary channel conditions.
Therefore ESE is more useful triggering measure for self organisation that adapts system over long term basis e.g. adaptation of sectors, antenna radiation pattern, frequency reuse, relay station locations etc.

(e) ASE does not include impact of area in outage. But ESE has a potential to include the impact of outage more soundly by weighing bins in outage more heavily.
Appendix C. Difference between Area Spectral Efficiency and Effective Spectral Efficiency
An optimization problem has three major aspects

i. **Objective function** or cost function i.e. the function that needs to be optimised (maximised or minimised)

ii. **Optimization variables** i.e. the variables in the optimization functions for which the values need to be determined that will optimize the objective function

iii. **Constraints** are set of rules formulated as function of optimization variables that allow them to take certain values while excluding others.

Based on the these three aspects optimization problems can be divided in different categories. Below we briefly describe the major types of optimization problems
### D.1 Types of optimization problems

Table in figure D.1 surmises the main types of optimization problems. A brief description of the key types is also given below:

#### D.1.1 Convex Optimization

Convex optimization or programming is applicable to cases when the objective function is convex and the constraints, if any, form a convex set. This can be viewed as a particular case of nonlinear program-
Convexity offers two major advantages

- **Global Optimality**: In convex function a local optimum is the global optimum.

- **Use of Lagrangian with Zero Duality Gap**: Convexity allows prompt use of lagrangian multipliers for constrained optimization. It is often possible to convert the primal problem (i.e. the original form of the optimization problem) to a dual form. This dual problem is obtained by using nonnegative Lagrangian multipliers to add the constraints to the objective function, and then solving for some primal variable values that minimize the Lagrangian. This solution gives the primal variables as functions of the Lagrange multipliers, which are called dual variables, so that the new problem is to maximize the objective function with respect to the dual variables under the derived constraints on the dual variables (including at least the non negativity).

However in general the optimal values of the primal and dual problems need not be equal. Their difference is called the "duality gap." For convex optimization problems, the duality gap is zero under a constraint qualification condition. Thus, when the problem is convex and satisfies a constraint qualification, then the value of an optimal solution of the primal problem is given by the dual problem.

### D.1.2 Nonlinear programming

Nonlinear programming studies the general case in which the objective function or the constraints or both contain nonlinear parts. This may or may not be a convex program. In general, the con-
The convexity of the program affects the difficulty of solving more than the linearity.

- If the objective function is concave (maximization problem), or convex (minimization problem) and the constraint set is convex, then the program is called convex and general methods from convex optimization can be used in most cases.

- If the objective function is a ratio of a concave and a convex function (in the maximization case) and the constraints are convex, then the problem can be transformed to a convex optimization problem using fractional programming techniques.

- **Quadratic programming** allows the objective function to have quadratic terms, while the constraints must be specified with linear equalities and inequalities. For specific forms of the quadratic term, this is a type of convex programming.

- **Fractional programming** studies optimization of ratios of two nonlinear functions. The special class of concave fractional programs can be transformed to a convex optimization problem.

- **Non-linear non Convex optimization** Unlike Convex optimization, non convex optimization is much more difficult and no general strategy is available for non-convex programming. Several heuristic methods are available for solving non-convex problems most of them are combination of gradient descent and genetic algorithms.

  In general these method involve the use of branch and bound techniques, where the program is divided into subclasses to be solved with convex (minimization problem) or linear approximations that form a lower bound on the overall cost within
Sequential quadratic programming (SQP) is a similar state of the art iterative method for nonlinear optimization. SQP methods are used on problems for which the objective function is twice continuously differentiable and the constraints are continuously differentiable.

SQP methods solve a sequence of optimization subproblems, each which optimizes a quadratic model of the objective subject to a linearization of the constraints. If the problem is unconstrained, then the method reduces to Newton's method for finding a point where the gradient of the objective vanishes. If the problem has only equality constraints, then the method is equivalent to applying Newton's method to the first-order optimality conditions, or Karush-Kuhn-Tucker conditions, of the problem.

D.1.3 Other Major Optimization Types

- **Stochastic programming** studies the case in which some of the constraints or parameters depend on random variables.
- **Combinatorial optimization** is concerned with problems where the set of feasible solutions is discrete or can be reduced to a discrete one.
- **Integer programming** studies linear programs in which some or all variables are constrained to take on integer values. This is not convex, and in general much more difficult than regular linear programming.
Appendix D. Brief Overview of Optimization Theory
Bibliography


[7] 3GPP, “Telecommunications management; self-healing OAM;


2046-2051.


Bibliography


[48] I. Siomina, "P- epich power and antenna tilt optimization in UTMS networks," in Proc. advanced industrial conference


[54] O. Yilmaz, S. Hamalainen, and J. Hamalainen, “Comparison of


249


[123] F. Bernardo, R. Agusti, J. Perez-Romero, and O. Sallent,
“A self-organized spectrum assignment strategy in next genera-
tion OFDMA networks providing secondary spectrum ac-
cess,” in *Proc. IEEE International Conference Communi-

learning approach for inter-cell interference coordination in
OFDMA cellular networks,” in *Proc. 8th Int Modeling and Op-
timization in Mobile, Ad Hoc and Wireless Networks Sympo-