A Survey of Self Organisation in Future Cellular Networks

Osianoh Glenn Aliu, Student Member, IEEE, Ali Imran, Member, IEEE, Muhammad Ali Imran Member, IEEE, and Barry Evans, Senior Member, IEEE

Abstract—This article surveys the literature over the period of the last decade on the emerging field of self organisation as applied to wireless cellular communication networks. Self organisation has been extensively studied and applied in adhoc networks, wireless sensor networks and autonomic computer networks; however in the context of wireless cellular networks, this is the first attempt to put in perspective the various efforts in form of a tutorial/survey. We provide a comprehensive survey of the existing literature, projects and standards in self organising cellular networks. Additionally, we also aim to present a clear understanding of this active research area, identifying a clear taxonomy and guidelines for design of self organising mechanisms. We compare strength and weakness of existing solutions and highlight the key research areas for further development. This paper serves as a guide and a starting point for anyone willing to delve into research on self organisation in wireless cellular communication networks.

Index Terms—Cellular networks, self configuration, self optimisation, self healing.

I. INTRODUCTION

The future of wireless cellular communication systems is marked by a drastic change in user behaviour triggered by the rampant growth of bandwidth hungry applications such as video streaming and multimedia file sharing. Particularly with the advent of a myriad of smart hand held devices, demand for broadband on the move is undergoing an unprecedented rise. This trend in user behaviour manifests itself as a tremendous pressure on cellular communication systems in terms of demands for capacity, Quality of Service (QoS), and energy efficiency.

While the surge in user applications is only bounded by imagination, the capacity of cellular systems, to support these applications, is tightly bounded by fundamental physical limits. This problem is further aggravated by considering financial constraints from the operator’s point of view, as higher capacity and QoS comes at the cost of higher capital expenditure (CAPEX) and operating expenditure (OPEX). Since users may be reluctant to pay proportionally higher bills for improved services, minimizing CAPEX and OPEX in order to make the business model commercially viable, whilst seeking to provide better QoS and capacity is a crucial consideration for today’s operators. As a result, operators strive 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 manner has triggered research to add intelligence and autonomous adaptivity, dubbed holistically as Self Organisation (SO), into future cellular networks. This initiative is mainly motivated by the following factors.

1. Although optimal capacity of the wireless channel is known to have physical upper bounds, the inherently unpredictable nature of spatio temporal dynamics associated with wireless cellular systems, makes even this optimal performance in terms of capacity and QoS unachievable with fixed legacy designs, as it lacks flexibility to intelligently adapt to the dynamics cellular systems face. Therefore, due to the mobility of users and the varying nature of the wireless channel, systems suffer from the under utilisation of resources resulting in either low resource efficiency or over utilisation which results in congestion and poor QoS, at varying times and locations.

2. With the recent advent of Femtocells, and increasing deployment of outdoor relays or Pico cells in the search for capacity and QoS, the number of nodes in future cellular systems will be too large to be configured and maintained for regular operation, with classic manual and field trial based design approaches. This is particularly true for Femtocells as they will be deployed in an impromptu fashion with plug and play capabilities and can thus cause severe interference resulting in degradation in performance of neighbouring macro cells, if not equipped with self organisation.

3. Given the huge scale of future wireless systems, the classic approach for periodic manual optimisation required during the life time will cease to be efficient. Also the increased complexity of systems will lead to greater human errors which will result in longer recovery and restoration times.

4. Finally, in addition to improved performance, SO can reduce the OPEX significantly as it eliminates the need for expensive skilled labour required for configuration, commissioning, optimisation, maintenance, troubleshooting and recovery of the system.

In summary SO is effectively the only viable way of achieving optimal performance in future wireless cellular networks in a cost effective manner. Standards for Long Term Evolution
(LTE) and LTE-Advanced [1] have therefore identified self organisation as not just an optional feature but an inevitable necessity in these future wireless systems [2].

Self organisation has been investigated in different types of communication networks. Compared to wireless cellular networks, work on self organisation in wireless sensor networks, ad hoc networks and autonomic computing networks is more mature and subsequently extensive surveys on these regimes has been published in [3], [4] and [5] respectively. However, in context of wireless cellular networks, self organisation is a relatively new yet rapidly growing area; and in the last decade significant research effort has been channelled at exploiting its full potential. It is therefore important and timely to put into perspective the major contributions from research projects deliverables, technical publications, and technical reports by standardisation bodies dedicated solely to self organisation in wireless cellular communication networks. To the best of our knowledge, this paper is the first attempt to provide a consolidated review of recent developments on self organisation in cellular networks along with a comprehensive tutorial on general characteristics of self organising solutions and the methodologies suitable for designing self organisation in cellular networks.

While the need for SO has been well investigated in [1], [6]–[8], designing self organisation functionality into cellular systems is a big challenge that requires extensive research. Many worthwhile efforts have been made to embody this objective in the last decade by individual researchers, projects and standardization bodies. The aim of this paper is not only to provide a categorical survey of the existing research works for SO in cellular communication systems by presenting a set of useful taxonomies, but also to identify potential methodologies and the open research issues for designing SO in future systems. In this process we also discuss a set of important features that make an algorithm or system self organising. These features can be used to assess the degree of SO an algorithm or system can possess and where applicable be used as a generic barometer in the survey process.

The contributions in this paper and its organisation are as follows

- In section II we present different definitions of SO presented in the literature so far. The aim is to clarify the ambiguities associated with self organisation. We present a framework consisting of a set of features that fully characterise SO from the view point of practical design and potential applications. Building on this framework we conclude this section by constructing a concise definition of SO.

- In section III we elaborate the general scope of SO in the context of wireless cellular communication systems. Major use cases for SO are identified and potential taxonomical approaches are discussed. The most widely accepted taxonomy that divides SO further into Self Configuration, Self Optimisation, and Self Healing has been chosen to organise the literature survey presented in this paper. The next three sections, IV, V and VI are devoted to a discussion of the state of the art literature on each of these three phases of self organisation.

- In section VII, we discuss various approaches and techniques that have been used so far and other novel approaches with significant potential for designing self organising systems. A brief time-line of SO is presented highlighting the key related contributions, projects and standardisation activities, in chronological order in section VIII.

- In section IX we conclude by listing a number of unsolved problems and suggest potential research directions capable of addressing these challenges.

II. UNDERSTANDING SELF ORGANISATION

The concept of SO is mainly inspired from nature where certain biological systems exhibit unique organised behaviour in order to achieve a desired objective while autonomously and intelligently adapting to the dynamics of their immediate environment. A school of shooling fish, swarming insects, herding sheep, synchronously flashing fire flies, human body’s complex immune and endocrine systems and flocking birds are some of the examples of such natural self organised behaviours. Since its conception, the widespread use, and in certain cases misuse, of the term SO has led to its drift from a technical issue to a philosophical debate as different authors have different interpretations to the notion of self organisation, particularily when it comes to designing SO in man made systems. We will not join this debate but rather focus on providing the reader with an understanding of underlying principles and characteristics that should be designed into and subsequently observed in systems in order to be called self-organised.

A. Definition

Self Organisation (SO) has been defined in various fields including biology, computer science and cybernetics [9]. In the very focused context of cellular systems, the concept of SO is not new, yet no single widely accepted definition exists. Here we discuss the key notions of SO that have emerged in the literature in the context of wireless cellular systems.

Haykin gave a visionary statement on the emergence of SO as a new discipline (he referred to it as cognitive dynamic systems) for the next generation of cellular systems [10]. Such intelligent systems would “learn from the environment and adapt to statistical variations in input stimuli to achieve highly reliable communications whenever and wherever needed” [11]. Spilling et al. [12] introduced the idea that SO should be seen as an adaptive functionality where the network can detect changes and based on these changes, make intelligent decisions to minimise or maximise the effect of the changes [12]. Yanmaz et al. viewed it from a perspective of cooperative networks and defined it as a phenomenon where nodes work cooperatively in response to changes in the environment in order to achieve certain goals [13]. This notion further characterises the behaviour of individual entities of the systems and identifies that individual behaviours should emerge on a ‘system wide scale’. Prehofer et al. further gave a detailed insight into the design principles and summarised self organisation as an emergent behaviour of system-wide adaptive structures and functionalities based on the local interactions of entities in the system [14]. It should be emphasised from
this definition that though most SO systems in nature result from an organised emergent pattern through interaction of local entities, localisation and distributed control are still not exclusively ‘necessary and sufficient’ features for self organisation.

Elmenreich in [15] started with a premise on how best to cope with the complexity associated with dynamic systems and identified self organisation as a solution. A clarification of the main idea behind SO rather than giving another definition to the growing list was elaborated. His idea highlights self organised systems as consisting of a set of entities that obtain a global system behaviour due to local interactions among these entities without the need for a centralised control. They also emphasised that self organised systems are neither a new class of system nor always have emerging structure as their primary property.

In the above definitions the occurrence of certain keywords are worth further consideration. Adaptive behaviour is the ability for any solution or algorithm to change from its current setting to a new setting in direct concordance with a change in state of the system. Distributed control applies to systems whose main functions are controlled locally from different points, as against having one central control point for the entire system. Emergent behaviour are patterns that can be observed in a system without being explicitly programmed to exhibit that behaviour. In summary, although the primitive feature of SO system is known to be adaptability, it is pertinent to acknowledge that autonomy, distributed, dynamic and emergent behaviour are also key attributes associated with SO that raise it above simple adaptability, as inferred from the definitions discussed above.

In order to establish a more solid definition of SO we need to first unsubscribe to the notion that SO must always emerge from intelligence. Though this is intuitive, it is also restrictive and difficult to practically design for by implementing intelligence. We rather suggest that from a view point of design and operation, the desired intelligence can be designed to emerge from SO. To support this argument we present the following case studies of self organising systems in nature.

1) Common Cranes: A flock of common cranes adapts its flight attributes 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’s flight efficiency [16]. This is done by dynamically maintaining an optimal delta formation during group flight. Most importantly, it does so

- without a leader and without global information exchange or coordination among all birds of the flock i.e. no explicit global signalling and no central control (Perfect Scalability)
- without diverging from the long term direction of flight or breaking apart even in the face of changing wind and weather conditions. (Perfect Stability)
- without loosing their ability of acutely executing individual as well as collective maneuvering to avoid predator attacks and large hurdles (Perfect Agility)

Such systems (Common cranes, school of fish or ants) behave in a truly self organising manner i.e. they are adaptive, autonomous and an emergent organised behaviour. It can be inferred from close observation of these and other similar natural systems that it is self organisation that emerges from certain characteristics of the system that in turn may manifest some degree of intelligence at a system wide level. From the view point of characterising and designing self organising systems, based on the observations in nature, these characteristics of adaptive systems can be identified to be scalability, stability and agility.

In the next subsections we discuss these key characteristics explaining how combined with adaptiveness they can manifest SO for practical wireless cellular systems. These characteristics can be used, not only to establish a more concrete and technically sound definition of SO but can also be used in assessing the degree of SO in solutions presented in the literature.

B. Characteristics of Self Organising Solutions

1) Scalability: Scalability strictly means the complexity of a solution should not increase unboundedly along with the scale or size of the system. The characteristics of scalability in natural systems emanates from the fact that global behaviours mostly emerge only from local and simple behaviours; and therefore the system remains operational if a reasonable number of entities leave or enter the system. 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, the following two main factors should be considered in the design process:

- Minimal complexity: algorithms should be austere in terms of time, space and other hardware resources required to implement. The lower the complexity of the solution the better its scalability.
- Local cooperation: algorithms 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.

As an example, consider an adaptive algorithm used in cellular system that changes the antenna tilts of all nodes in the system with the 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 self organising. Thus to be perfectly scalable, the complexity should not increase as the number of nodes increases. Such a distributive model aids perfect scalability in large systems. With the advent of machines with unfathomable computational power, scalability seems to become a relative notion and even centralised complex systems can be afforded. However in the context of an ideal SO system, scalability achieved through minimal complexity and local control are highly desirable as it helps to achieve the other two characteristics of SO as discussed below. Further study on the tradeoff between highly scalable solutions and the control of individual entities in such a system as well as the need to avoid global state information can be found in [4].
2) **Stability**: A simplistic but generic definition of stability pertinent to our context of SO is: an algorithm or adaptation mechanism that is able to consistently traverses a finite number of states within a acceptable finite time. This means that transiting from its current state to any desired future state within a finite and feasible time frame is certain and valid for a specific task. It does not continue infinite oscillations but ultimately reaches a stable state\(^1\) in a bounded time while transiting from one state to another. Designing stable algorithms is very specific to the nature of the problem and is a separate field of research in itself. For the scope of this study, this simple notion of stability as explained above also ensures robustness i.e. when the system is faced with an unusual event or operational environment, forcing it into an undesired state, the system should be able to return to one of its desired states within a finite time frame. It also ensures that the system does not oscillate between states infinitely (ping pong effect). Although robustness is a fundamental requirement of SO, it need not be considered as separate criterion for SO, but rather as a measure of the stability of a SO solution. Perfect stability also requires that the system should be elastic and self healing. That in turn, most importantly, means that single point failures or chaotic design that mostly results from a centralised control should be avoided. The reader can refer to [17] for details on stability analysis for autonomous swarm aggregation that are based on biological principles and exhibit self organisation functionality.

3) **Agility**: An adaptive system can be scalable and stable but still not perfectly self organising as it might be sluggish in its adaptation. Agility is another key characteristic of self organising systems (as observed in the case studies above). Agility describes how supple or acutely responsive an algorithm is in its adaptation to the changes in its operational environment i.e. in order to be self organising, algorithms should not only have the capability to adapt and 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 the subsection on stability. Perfect agility is however 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. Depending on the desired objectives, a threshold can be set by the operator for a certain level of desired agility. 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 way to enhance agility as it reduces the delays incurred due to feedback to a central location, processing of feedback, decision making and propagation of the decisions back to the point of actuation. Further reading on agility or different dimension of adaptation and required time frame for adaptation is found in [18].

A rather implicit characteristic of SO networks is intelligence. In engineering systems, such intelligence can be designed to emerge from the elementary and local self organising functions as discussed above in the context of case studies of SO systems in nature. Alternatively, intelligence can also be manifested either through classic learning using artificial intelligence techniques; or by pre-designing a set of optimal solutions for all potential states of the operating environment of the system. All of these approaches will be discussed in detail in section VII on methodologies for designing self organisation.

From the discussions in this section so far, we can infer that an adaptive and autonomous 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 the face of all potential dynamics in its operating environment.

In contrast to conventional adaptive systems, achieving scalability, stability and agility is now more challenging. It is thus important to ensure that SO systems are stable, agile and scalable enough to maintain the global system objectives in the face of all potential dynamics of its operating environment. In later sections, we would further discuss examples of selected solutions in literature and analyse their stability, scalability and agility after a firm understanding of self organisation in wireless cellular communication systems.

It is pertinent to clarify the meaning and difference between frequently used terms synonymous to self organising networks. These include adaptive networks, autonomous networks and cognitive networks.

*Adaptive networks* are systems that change their configuration in direct response to changes in the system. These changes are usually triggered once certain fixed conditions are met. They do not have to be scalable, agile or exhibit any form of intelligence. Adaptive networks have their origin from control systems where adaptation is based on a feedback loop.

*Autonomous networks*\(^2\) are simply adaptive networks with no human or external intervention. Such systems are not necessarily stable, agile or scalable according to the definitions provided above. In relation to self organised networks, we can deduce that all self organised solutions are adaptive and autonomous but not all adaptive or autonomous networks are self organised. A detailed study of autonomous adaptation describing why, when, where and how adaptation is performed can be found in [18].

*Cognitive networks* are autonomous networks enhanced with the capability of learning and adaptation of system parameters based on interaction with its environment. The cognitive nodes are able to plan, observe and execute actions in concordance with the cognition cycle [11]. The key part of cognitive networks is this interaction with the operating environment and its ability to learn from the process. For further reading on cognitive networks the reader can refer to the seminal paper [19].

Self organised networks however, are not just adaptive and autonomous, but are able to independently decide when or how to trigger certain actions based on continuous interaction with the environment. They are able to learn and improve performance based on feedback from previous actions taken.

\(^1\)This simplistic notion of stability can also be extended to stateless/continuous systems by discretising their operational environment into finite states.

\(^2\)Autonomous networks are quite different from autonomic networks used in computer networks. Although similar to self organising cellular networks in principles, they are technically referred to as autonomic communication systems in computer networks [5].
It is thus important to emphasise that SO is a functionality that can be observed in any network including cognitive networks.

In this section we have provided a firm definition and understanding of the term self organisation starting with definitions used in other disciplines. We deduced that adaptivity, autonomous and learning are key in defining such systems. Furthermore we have elaborated stability, scalability and agility as characteristics that are desired even more in any self organised system. System exhibiting all these characteristics can be summarised under the following 9 specific categories

1. Coverage and capacity optimisation
2. Energy saving
3. Interference reduction
4. Automated configuration of physical cell identity
5. Mobility robustness optimisation
6. Mobility load balancing optimisation
7. Random access channel (RACH) optimisation
8. Automatic neighbour relation function
9. Inter-cell interference coordination

All the aforementioned use cases can be classified under one of the four main system objectives i.e. Coverage expansion, capacity optimisation, QoS optimisation and Energy efficiency all driven by the basic motive of cost optimisation as shown in Fig. 1. It must be noted that these use cases and their main objectives have strong coupling with each other making the optimisation of one difficult without compromising the others [23]. This coupling is further discussed in the following sections.

B. Taxonomies for SO in Wireless Cellular Systems

With a clarification of the characteristics of SO and applications relevant to cellular communication systems, we present a taxonomy under which such systems can be studied. Three main classification regimes have been identified

1. Time scale based classification: Cellular communication systems have a set of different scale dynamics as explained in Fig. 2. Therefore, different self organising algorithms designed to cope with these dynamics have to operate on respective time scales. e.g. adaptive modulation and coding scheme algorithms operates on a micro second scale [24] whereas a power control based load balancing algorithm might operate on time scales of

![Diagram of Taxonomy](https://via.placeholder.com/150)

Fig. 1. SON Taxonomy

![Diagram of Time Scale](https://via.placeholder.com/150)

Fig. 2. SON Time scale of operation

---

**Reference:**

ALIU et al. A SURVEY OF SELF ORGANISATION IN FUTURE CELLULAR NETWORKS

---
minutes and hours [25] whereas adaptive sectorization algorithms might operate on the scale of days and months [26]. This different diversity of times scales can be used to classify SO. Although, this is the simplest taxonomy as it does not describe much about the scope or objectives of the algorithms.

2 Objective or use case based classification: Each SO algorithm can be classified using categories of use cases. But the problem with this approach is that one algorithm can have multiple use cases, e.g. same self-organising algorithm that aims for load balancing can optimise capacity as well as QoS.

3 Phase based classification: There are three clear phases in the lifecycle of a cellular system i.e. deployment, operation, maintenance or redeployment. Holistically, self-organising solutions/algorithms can be classified with respect to each of these phases. The algorithms that self-organise these three phases can be classified into self-configuration, self-optimisation and self-healing.

Fig. 1 illustrates these possible taxonomies based on objective and phase based classifications and their interrelationships while Fig. 2 depicts the time based classification. From the classification types highlighted above, the phase based classification is commonly used by standardisation bodies [27] and major research projects [6] as it isolates different stages in the system lifecycle and solutions can be easily deployed focusing on individual stages. Therefore, we categorise this survey on SO in three parts corresponding to the phases: self-configuration, self-optimisation and self-healing.

For a more detailed analysis of SO under these phases, we have also made use of the other two modes of classification where applicable. We will also use the framework for characterisation of SO presented section I as a general barometer to assess the degree of SO in the proposed solutions where applicable.

IV. SELF CONFIGURATION

Configuration of base stations/eNodeBs (eNBs), relay stations (RS) and Femtocells is required during deployment/extension/upgrade of network terminals. Configurations may also be needed when there is a change in the system, such as failure of a node, drop in network performance or a change is required in service type. In future systems, the conventional process of manual configuration needs to be replaced with self-configuration. To fully appreciate the importance of self-configuration, a clear understanding of future cellular systems is necessary. Fig. 3 shows the topology of LTE networks. The main difference from legacy network architectures is that future networks will be IP based and will have a flat architecture. The reader can refer to [28] for details on LTE system architecture. This decentralised structure requires most of the configuration functions to be done locally at each node. Given the huge number of nodes and scale of the system, this will only be feasible if done via self-configuration. This is particularly true, in the case of impromptu deployment scenarios e.g. for Femtocells or relay stations where no skilled manual support is available.

A general framework for self-configuration of future LTE networks has been presented in [29]. This paper addresses the problems associated with autonomous deployment of a new site i.e. without the need for any human intervention. It is foreseeable that nodes in future cellular networks should be able to self-configure all of its initial parameters including IP addresses, neighbour lists and radio access parameters. In the following subsection, we discuss the literature on each of these three major self-configuration use cases.

A. IP Address Self Configuration

The process of obtaining network layer configuration settings (IP settings) in the IP based future cellular network can either be via a dynamic host configuration protocol (DHCP) as used in computer networks or via a Bootstrap protocol (BOOTP) agent. An alternative option is suggested by Sophia et al. in [30]. These authors propose that each node should be added into a multicast group with a uniform subnet mask so that configuration information can easily be sent as a multicast to the nodes in that group. This is based on the Internet Group Management Protocol (IGMP). With this approach the neighbour list is generated before the node becomes operational.

The authors in [31] provide a detailed overview of the self-configuration process in eNBs which can also be extended to the Femtocells impromptu deployment scenario. It is proposed that a newly deployed node obtains its IP address, and that of the Operation and Maintenance (O&M) centre and then sets up a link to the core network through an access gateway. This link is only established after all security checks and authentication are made from the core network. The node then downloads the necessary software, and operational parameters, whilst other parameters are chosen autonomously based on already existing configuration settings of neighbouring nodes.

Fig 4 gives a summary of the self-configuration flow for a newly deployed node designed to operate in a plug and
play mode. Once plugged in, the node scans its neighbourhood for neighbour cell ID and generates a neighbour cell list using any of the schemes described in section IV-B. From the list of neighbours, a sponsor node is chosen. The concept of using a sponsor node is explained later in the section VII-A2 on distributed learning. Communication is then established with the configuration server while initial location dependent radio access parameters are obtained from the current setting of the sponsor node. Based on these initial settings, the node is able to autonomously determine which configuration parameters optimise its performance for its expected traffic and location. After this pre-configuration and backhaul links have been established, the node is now in a full operational state. We now describe works in literature on the neighbour cell list generation.

B. Neighbour Cell List Self Configuration

An important functionality or use case of SO in LTE networks is the automated configuration of neighbour cell list and the automatic updating of neighbour relation function. Each node making up a cell has a unique cell ID and via some neighbourhood functions needs to create a neighbour cell list. The list update is required when a new cell is deployed or is temporarily out of service in the network.

Authors in [29] propose an algorithm for generating a neighbour cell list and updating it using a centralised as well as decentralised approach. The criteria for the selection of neighbours or the initial generation of the neighbour cell list can be based on the geographical coordinates of the cell sites as demonstrated in [32]. This is achieved from the O&M layer, where knowledge of the entire network layout, that shows the location of previously deployed nodes can be used to generate a list of potential neighbours for newly deployed nodes. However, other than being less scalable this approach is also not accurate and hence unstable as geographical coordinates cannot determine handover regions. Geographical coordinates can be used to estimate only the distance dependent pathloss. Nevertheless, characteristics of the wireless communication channel as well as scenarios where sites no longer in operation (or idle mode) would still be erroneously considered as neighbouring site. Furthermore this approach totally neglects effect of other parameters such as antenna patterns, which play a vital role in determining the actual cell coverage and thus potential neighbours.

The authors in [33] address the same problem of generating a neighbour list for self configuration and propose an algorithm that incorporates antenna patterns and transmission power in addition to geographical coordinates. In the proposed algorithm, the neighbouring cell list is generated by selecting cells that have overlapping coverage. Although this algorithm is an improvement over the scheme proposed in [32] the coverage estimation used is still theoretical and does not take into account factors of the real radio environment, such as shadowing and propagation loss etc. This problem is solved by authors in [34] as they propose a more accurate method of generating the neighbour list based on real time scanning of signal to interference plus noise ratio (SINR) from adjacent cells. The performance of this algorithm has been evaluated for different cell types (pico and macro) and the results show a tradeoff in using a low SINR threshold that results in a very high number of neighbours and a high SINR threshold that results in a limited number of neighbours. This is the best self organising approach in the literature to date. However, it is still not yet a perfect SO solution for the following reasons. Firstly, the absolute real time nature of the solution makes it less agile. Secondly, in the case of the radio environment with highly mobile scatterers the solution may become unstable as the neighbour list will be continuously changing. Finally, if a valid neighbour cell is temporarily shadowed by a scatterer it may not be included in the neighbour list, and hence will never be considered for handover. This can be due to the fact that at the instant the SINR measurements are made, an obstacle might be causing a blockage and seconds later is no longer present. This would give a wrong representation of the actual neighbours. In the future work section we will present directions for improvement in neighbour list discovery algorithms.

C. Radio Access Parameter Self Configuration

Having established a neighbourhood function that identifies cells and their respective neighbours, there are a large number of radio access configuration settings in future cellular networks. Below we discuss literature reported on self configuration of major radio access parameters.

1) Frequency allocation self configuration: For impromptu deployment, either for relay, pico or Femtocells, determining the (Media Access Control) MAC layer frequency channel
that causes least interference to existing nodes autonomously and which still provides enough bandwidth to achieve the desired throughput is still an open research issue. While many authors have addressed this issue, the majority of these are limited to the realm of cognitive radio where these algorithms usually aim for an opportunistic spectrum reuse [35], [36]. Unfortunately, these algorithms are not straightforwardly extendable to Femto, pico or relay deployment scenarios in the context of future cellular networks as the nodes are primary not secondary users and spectrum allocation has to be guaranteed not opportunistic. A pertinent paper in this regard is a decentralised medium access protocol proposed in [37] where authors show that the proposed MAC protocol is capable of operating in the worst case scenario of no frequency and eNB location planning. It can thus pave the way towards self organising or more precisely self configuring cellular systems by essentially eliminating the need for pre-planned deployment.

2) Propagation parameter configuration: The type of antenna used, its gain and the transmit power are major factors influencing the performance of the entire network. Although omnidirectional and sectorised antennas with fixed radiation patterns have dominated cellular systems to date, future systems may use smart antennas with the capability to reconfigure their radiation patterns electronically. Provision for this upgrade has been made in the standards by allowing beamforming techniques and antenna diversity via MIMO techniques. With the advent of smart antennas, the role of antenna configuration may go beyond just self configuration to self optimisation and self healing of cellular systems to maintain the major system objectives. Even with current antenna technology, the azimuth and tilt angle of the antennas have significant impact on the system performance as it determines the signal propagation direction and hence controls the interference and capacity of the system. As state of the art antennas allow remote electrical tilting in addition to mechanical tilting, there is a margin for self configuration and even self optimisation of antenna tilt. A significant number of papers in the literature have addressed this issue, most of which have focused on improving the coverage and capacity in general for GSM [38], CDMA [39] and HSDPA [40]. Authors in [41] propose a scheme for antenna tilt angles and power configuration settings from a centralised server that has a global knowledge of all current system settings. Self configuration of future wireless cellular systems in this way would require increased signalling with the centralised server and its complexity would increase with the number of nodes making it less scalable.

Authors in [42] and [43] address the problem of traffic hot spots via antenna tilting both for CDMA and HSDPA respectively. The tilting mechanism proposed in both of these studies are based on the basic idea of tilting down the overloaded sector’s antenna to reduce its effective coverage area in order to shift its load to neighbouring cells. These forms of totally uncoordinated tilt adjustments are not only suboptimal in view of system wide performance but they are also shown to trigger excessive handoffs from the overloaded sector to the adjacent sectors [42]. Such soft hand offs can be manageable in CDMA, but in OFDMA they are hard hand offs and undermine the practicality of the approach; as handover involves a change of carrier frequency. These solutions might not be stable in their usual operation as the post effects of such totally uncoordinated individual antenna tilting can spread in the cellular network and can cause oscillations.

Temesvary demonstrated in [44] a simulated annealing approach for tilt optimisation based on terminal measurements of the SINR obtained from Channel Quality Information (CQI). An extended local search is performed and each solution obtained to the cost function is accepted with a given probability. As the search continues, better solutions are accepted and the algorithm runs until there are no new improvements in the cost function. The results presented in the paper can be seen as self organising as they give comparable solutions to an exhaustive search method, they are also stable and agile as they reach desired states from any initial configuration of newly installed nodes.

The authors in [45] take a game theoretic approach by modelling a utility function for the system where each cell tries to maximise this function through antenna tilt in a non-cooperative game against neighbouring cells to achieve a Nash equilibrium. However, it is well known that the existence of the Nash equilibrium does not imply convergence. A remedy to this problem is proposed in [46] where a discrete time gradient play at each stage of the game is used to ensure convergence.

Antenna tilt configuration has been widely studied in UMTS networks but there is limited literature focused on LTE networks. Yilmaz et al. tried to fill this gap by demonstrating the effect of mechanical tilt and electrical tilt on the downlink performance in LTE networks [47]. Simulation results reveal that electrical tilting provides better performance in interference limited systems and both electrical and mechanical tilting produce similar performance in a noise limited environment. However, the authors do not discuss how these tilt angles can be optimised in a self organised fashion.

Lu et. al. focus on the uplink power control and antenna tilt using real 3D models. They show that a tilt angle of 4° to 8° was optimum for cell radius of 300m [48]. Athley et al. evaluated the effect of electrical and mechanical tilts on capacity and coverage in LTE networks and extended Yilmaz’s results in [47] by finding the optimal combination of mechanical and electrical tilt. It was shown that electrical tilt is optimal for cell edge performance and mean throughput. However, for achieving peak throughput, a loss of 25% is observed using electrical tilt only and a loss of 7% for mechanical tilt only, as compared to using a combination of both [49]. Despite the simulation results provided by the authors listed above, it is our opinion that implementing electrical tilt remotely requires less effort and is better suited for self organising networks. Problems related to self configuration and optimisation of propagation parameters that need further investigation are highlighted in the future work section.

D. Effect of Operators Policy on Self Configuration

Policies are rules given from an organisational level to determine the objectives of algorithms being implemented on network elements. The policy of the operator is also vital in the
self configuration process. The operator policy could either be coverage extension, capacity optimisation, energy efficiency or fairness among users or any combination of these with predefined priorities.

In self configuring networks the initial configuration parameters can be divided into two main groups as presented in [50]. The first group are those parameters whose settings have no effect on neighbouring cells and can thus be set independently and might not require sophisticated real time algorithms. However, the second group of parameters whose settings determine the performance of neighbouring cells need to have knowledge of current configuration setting of the neighbours. Settings of both groups of self configuration parameters would be governed by operator policy. The authors in [51] give a detailed classification of groups and stages in the initial configuration phase. They present a policy based framework for self configuration and propose that a subset of parameters can be configured in real time through a Dynamic Radio Configuration Function (DRCF). The authors conclude that operator policy is vital in formulating DRCF and thus affects configuration and operation of new cells. This policy therefore also determines the optimum time required to self configure new cells and neighbouring cells that have been in idle mode (during low traffic intervals e.g at night).

Meng et al. proposed a self configuration management mechanism that relies on data in the Resource Information Base (RIB) and Policy Information Base (PIB) [52]. Based on data analysis, the choice of operator policy stored in a policy information base would determine if coverage is more important in a given urban area or they are more concerned with high data rate for few users in a rural area. These policies are key in determining the type of self configuration actions taken on new nodes inserted in a region.

Self configuration is a key component in the self organisation cycle especially at the preoperational stage of any cellular network. This section focussed on IP address self configuration, automatic generation of neighbour cell lists and radio access parameter configuration. It is certain that the objective of coverage extension and capacity optimisation would ultimately lead to deployment of more macro, pico sites for outdoors and extensive amount of Femtocells for indoor environment. The surge in the number of indoor base stations where operators would have limited or no access in effecting configuration changes, makes evident the need for a self configuration functionality. Plug and play nodes are thus desired with huge potential in CAPEX saving. Key papers in self configuration in LTE networks include [34], [44], [53], [54]. However, solutions for self configuration are still at its early stages and more mechanisms in the physical layer are still being investigated to ensure feasible plug and play solutions for both indoor and outdoor eNBs.

V. SELF OPTIMISATION

After the initial self configuration phase it is necessary to continuously optimise system parameters to ensure efficient performance of the system if all its optimisation objectives are to be maintained. Optimisation in legacy systems can be done through periodic drive tests or analysis from log reports generated from network operating centres. However this approach will not be tenable in future networks. The need for the network to continuously self optimise its operation has been highlighted in [21] and a number of papers have already embarked on providing solutions to the problems in this area [55]. Below we present a survey of the literature by further classifying these works with respect to use cases of SO in section III-A.

The framework for our discussion on self optimisation can be seen in Fig 5. We focus on load balancing, interference control, coverage extension and capacity optimisation. Load balancing would be discussed under resource adaptation, traffic shaping and coverage adaptation. We go in details with coverage adaptation based load balancing discussing mechanisms of achieving this (power adaptation, antenna adaptation or via relay nodes).

A. Self optimisation for Load Balancing

The need for load balancing mechanisms to mitigate the effects of natural spatio-temporally varying user distributions was realised immediately after the advent of commercialised cellular communication systems [56]. A large number of research papers since then have addressed this problem and have proposed a variety of very useful load balancing strategies some of which even reach the required level of scalability, stability and agility and achieve the desired SO behaviour [57], [58]. However, the problem lies in the fact that, most of these schemes were specific to the particular generations of cellular systems evolved so far and only a few are applicable to the emerging systems (e.g. LTE and LTE-A) due to the differences in the MAC and physical layer. In the following subsections we discuss the key papers by broadly classifying them in four general categories based on the main underlying approach they take towards Load Balancing (LB), i.e. 1) Resource Adaptation based LB, 2) Traffic Shaping based LB 3) Coverage Adaptation based LB, and 4) Relay assisted LB.

1) Resource Adaptation (RA) based LB: In this scheme, the main underlying principle is to adapt the amount of resource 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 other cells that are less loaded or from a common pool of free channels. References [57]–[60] present various load balancing schemes building on this main idea. Authors in [59] presented and evaluated the performance of channel borrowing algorithms where an over loaded cell borrows channels from selected cells that are least loaded. The authors further refined the idea in [59] and presented in [57] an improved version in terms of scalability by limiting the borrowing process within the hierarchical tier based local structures or clusters, although at a cost of loss in performance due to lack of global optimality in this distributed approach. The scope of these schemes has been limited to legacy GSM type systems where neighbouring cells use different frequency channels. In the case of emerging OFDMA based cellular network (LTE and LTE-A), frequency reuse of one prevails, and such channel borrowing will cause intra-cell interference that is not otherwise present in OFDMA based systems. However, some authors have attempted to
extend the channel borrowing concept to make it applicable to CDMA based cellular networks by introducing the idea of virtual channel borrowing for CDMA based systems [61]. The basic concept of channel borrowing has also been extended for emerging cellular systems by generalising it in the form of bandwidth management strategies [58], [60]. While the work presented in [58] remarkably exhibits all the desired features of SO i.e., scalability, agility and stability, due to its cell by cell implementation style, its lack of pragmatism for wireless cellular systems with universal frequency reuse undermines its applicability. Arguably this is the reason such schemes have not been further investigated in the context of LTE and LTE-A.

2) Traffic Shaping based Load Balancing: In these 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 on going calls in order to effectively match the offered traffic load with the available resources for optimal Load Balancing (LB). A number of papers have focused on optimising such traffic shaping strategies to minimise trade off in terms of hand over overheads and interference [56], [62]–[66]. These papers all refer to the pioneering paper [56] that presented a simple but seminal idea of selecting least loaded cells among candidate neighbour cells for handover based on LB. A comprehensive review of LB balancing algorithms using the traffic shaping approach and other approaches have been presented in [66] along with a general mathematical framework that models the underlying principles of a variety of LB strategies. Although call admission is also a possible way of traffic shaping [67], [63], [65], due to the randomness of arrival times and user mobility, its use for optimal load balancing is a relatively complex approach that necessitates extensive cooperation among cells to ensure QoS, so that the call rejected by one cell is accepted by others. On the other hand, handover parameter adaptation to trigger forced handovers of ongoing calls to shed extra load to neighbouring cells, is a more straightforward approach [56], [62], [64], [68]. Authors in [68] presented a handover based approach for LB exclusively for LTE. The algorithm requires all cell loads to be known at the 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 a centralised nature, that might compromise scalability and agility of the algorithm, are not thoroughly explored in this work. It is important to mention that, in the case of LTE and LTE-A, HO based traffic shaping is not as feasible an approach as it was for CDMA based UMTS. This is due to the fact that in OFDMA based systems such as LTE and LTE-A, the luxury of soft and softer handover is not available and hard handover usually involves a change of carrier frequency incurring extra complexity and overheads. This makes the majority of the existing solutions based on the traffic shaping approach less attractive for LTE and LTE-A.

3) Coverage Adaptation based Load Balancing: These schemes rely on mechanisms to change the effective coverage area of the cell to match the traffic offered with the resources available either through power adaptation, antenna adaptation or a hybrid of both techniques. This approach has been most extensively studied in the literature because of its flexibility, and effectiveness and in the following subsection we discuss the key papers by further classifying them into the three sub categories listed above.

a) Load Balancing via Antenna Adaptation

In these schemes the basic idea is to reduce the coverage area of the over loaded cells either through tilting down the antenna [69] or by changing its radiation pattern [70]–[78]. Authors in [70]–[72] consider a scenario with traffic

![Diagram](image-url)

Fig. 5. Self Optimisation Chart
hot spots that cause congestion in some cells and propose
a solution that contracts the antenna patterns of the congested
cells around hot spots, whereas neighbouring cells expand
their radiation pattern to fill in any coverage gaps. These
three papers by the same authors, assume negotiation among
only neighbouring eNB to fill in any coverage gaps and
thus have basic scalability, but at the cost of the stability
as coverage gaps may be left uncovered with such limited
local cooperation. In [73], the same authors attempt to address
the coverage gap issue by using a bubble oscillation model
where air in the adjacent bubbles fills any gap among adjacent
bubbles via oscillation of the bubbles. Authors use the analogy
of air for antenna radiation patterns and uncovered users
as the gap between the bubbles, suggesting that all users
can eventually be covered through oscillation of antenna
radiation patterns. Whilst the use of this model is novel and is
realistic in its physical context, oscillation of antenna radiation
patterns in cellular systems can have adverse impact on the
handover process, that might not be a problem in CDMA
based cellular systems, where a user can communicate with
multiple eNB simultaneously (soft handover). There may be
a serious compromise on agility and stability in general (non
cooperative) OFDMA based cellular systems because of the
hard handovers involved as discussed above. In [75] again, the
same authors present a solution in the context of WCDMA
instead of CDMA.

Authors in [77] 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 eNBs in the network.
Although significant improvement in terms of service prob-
ability 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 [76] propose a radiation pattern adaptation scheme for
adaptive sectorisation in cellular systems. The key feature
of this scheme is that it does not track individual mobile users
but rather relies on statistical information. Furthermore, it
does not require extensive inter site cooperation as it can
be implemented on each eNB individually making scalability
more effective. This scheme makes use of spatial information
and mobility patterns of the mobile users to depict user
geographical distributions 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 sectorisation methodology. Since the empirical or
offline statistical information of user distributions has been
used in this scheme, the agility of the scheme may only
measure up to the requirement of very long term dynamics,
making it less pragmatic for online operation to cope with
relatively short term load variation among cells.

Authors in [74] compare the impact of cooperative beam
shaping with the cooperative tilt adaptation for LB. The tilt
based approach has relatively less margin for performance
improvement but it is more pragmatic as it is implementable
with conventional widely commercialised parabolic antennas
[69]. The radiation pattern adaptation approach has been
shown to have a huge margin of performance improvement
but requires smart antennas that can change their radiation pat-
tern electronically. Another limitation of the radiation pattern
adaptation based approach is that, an expansion in beamwidth
of neighbouring cells is required to fill up the coverage
gaps, when an over loaded cell narrows its beamwidth. This
expansion in beamwidth will cause an inevitable decrease in
the gain of the antennas, hence the QoS for users in that cell
deteriorate.

b) LB through Power Adaptation (PA)

In these schemes, the coverage is adapted for LB through
control of transmission power of the signal carrying the cell
signature that is used by the users for cell association. This
is different from the handover parameter control used for traffic
shaping discussed in subsection V-A2. In contrast to traffic
shaping through forced handovers, coverage adaptation through
power adaptation does not only affect the ongoing calls, but
effectively changes the coverage area and thus changes the
association of all users in the coverage area. Only a few papers
have reported work in this area [79]–[84].

Authors in [79] present a centralised scheduling algorithm
wherein users may switch eNB in every time slot with the joint
objective of throughput maximisation and LB simultaneously.
Authors in [79] also assess the heavy overhead of this solution
that makes it void of scalability and agility, and therefore
propose a lighter version of the algorithm. They suggest
separating LB from throughput maximisation and thus the time
time scale of the eNB switching for LB alone can be increased to
several time slots such that it is just enough to cope with
user mobility, rather than in every single slot as required for
throughput optimisation. This solution reduces the signalling
overhead, making it relatively more scalable, however central
control is still required in this solution.

Authors in [80] determine such a combination of pilot
power levels in the network that guarantees full coverage and
maximises the capacity ratio of the bottleneck cell. Whilst the
algorithm has been shown to yield reasonable performance
improvement through realistic planning tools, it is more suitable
for offline planning of pilot powers, than an online LB scheme
executable during the operational phase of the network. This
is because it requires long term traffic statistics for the whole
system to determine the bottleneck cell, as well as the load
in each cell to determine pilot power vectors for all cells.
Furthermore, it does not explicitly address the effect of soft
handover on the capacity of the network which is important
in the context of CDMA for online LB operations. The scope
of [80] is also effectively limited to CDMA based system.

Only [81]–[84] consider an OFDMA based systems and
propose algorithms for dynamic association or coverage adap-
tation. References [81] and [82] use coverage adaptation
or dynamic association, as termed by the authors, for joint
objective of load balancing and interference avoidance through
fractional frequency reuse. While this scheme shows signifi-
cant gain in terms of designed utility as an indicator of system
wide performance, the underlying assumption of network wide
feedback and channel estimation, at each MS and eNB at
each scheduling instant and the need for a central control
entity, makes this solution lacking in scalability and agility
required for self organisation. Authors in [83] proposed a similar algorithm that is again purely centralised and thus lacks scalability and agility.

Only the solution presented in a recent work [84] is fully scalable as it is can be implementable in a fully distributed fashion. The basic idea is that each eNB periodically broadcasts its average load and the MS uses this information along with the signal strength, to make the decision for cell association. This is contrary to legacy cellular systems where the cell association decision is only made on the basis of received signal strength. This distributed algorithm has been shown to achieve the global optimal iteratively but with the two crucial assumptions of 1) spatial load being temporally stationary and 2) the time scale at which the eNB broadcasts its load is much larger than the time scale of call holding. The first assumption has implications in the sense that even in the given spatial vicinity, loads are bound to vary over time in real wireless cellular system environments due to mobility etc. The violation of this assumption will directly compromise the stability of the proposed solution [84]. In order to make the second assumption valid, eNBs have to keep their broadcast time very large, thus compromising the agility of the solution.

c) LB through Hybrid Approaches

In addition to the papers discussed above, some authors have used multiple approaches to LB simultaneously. Authors in [85] presented analysis for the simultaneous use of traffic shaping through both call admission control and handover hysteresis control and coverage adaptation through both antenna adaptation and TA for general TDMA/FDMA systems. While simulation results report 3-11% network wide gain in performance the proposed methods lack scalability because of the requirement for a central control unit that needs to exchange excessive signalling with all users in the network to obtain spatial traffic estimates. Furthermore, the dependency of the proposed methods on the use of mobile positioning jointly with cell assignment probability maps (generated by the network planning process) for spatial traffic estimations makes it less agile in dealing with the spatio temporal dynamics of cellular communication systems. It is thus more of an offline design methodology, more useful during the deployment phase, than an online LB mechanism implementable in the operational phase. Authors in [86] proposed a similar centralised LB algorithm for CDMA based systems that use both antenna adaptation and power adaptation together. As highlighted by the authors, this is time consuming and hence not agile enough to be used for real time LB. Rather its use as a self healing scenario has been proposed by the authors. Furthermore, it is important to mention that this algorithm is also centralised and hence lacks scalability.

4) Relay Assisted LB: In addition to the approaches discussed above, relay assisted cellular networks have also been extensively investigated as possible LB tools [67], [79], [87]–[97]. This is due to the fact that relay nodes can help achieve LB via at least three different means. 1) Through coverage adaptation, by improving coverage and signal quality at the point of interest e.g hot spot [92]. 2) Through resource adaptation by local or opportunistic reuse of spectrum [94]. 3) Through traffic shaping by relaying or routing traffic from over loaded cells to less loaded cells. The last of these three approaches has been most extensively investigated in literature [87]–[92], [95]. However, most of these approaches assume CDMA cellular systems with ad-hoc relays operating on an out-of-band spectrum e.g. ISM. As highlighted by authors in [90] 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 the ad hoc relays. It is concluded in [90] that for dynamic load balancing the number of such channels required is much more than those required for load sharing (improving the call blocking probability of a hot spot to 2%).

Authors in [95] and [96] are the only ones to consider relay stations with inband spectrum and of a non adhoc nature for LB in a fully architecture based OFDMA cellular system. The main idea in [95] is that all users establish their association with the base station or relay station dynamically to maximise a utility function designed to reflect system wide performance. It is further proposed that relays also establish their association with the eNB’s to maximise the same utility. Stability of the system is ensured by confining the reassocation process to one relay node and UE per time slot to avoid a 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, a central control unit is still required to receive, process, and feed back the dynamically changing system wide utility to and from all UE and relay nodes in the network. This may have an adverse effect on the agility of the solution in a practical system because of the delays incurring from large amounts of data processing and its relaying to and from a central unit.

Authors in [96] exploit the idea of dynamic clustering through cooperation among neighbouring cells thereby avoiding need for central control unit to keep solution scalable. They introduce a dynamic clustering approach, where overloaded cells form clusters by selecting from the 6 neighbouring cells those with the 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 a hypothetical scenario where only one central cell is over loaded and a more realistic evaluation is indicated as 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 as it may have stability issues when the same cells are neighbours to more than one overloaded cell resulting in a ping-pong effect in the clustering process.

B. Self Optimisation for Interference Control

Self optimisation provides a diversity of ways to control interference and thus improve capacity. The simplest direct approach would be to enable certain cells to switch off or go into idle mode when not in use. Authors in [98] provide a self optimising interference control scheme using this approach. The basic idea is that based on previous observations, the node establishes a pattern and after a given period of time in idle mode, the node hibernates. This reduces interference caused to other cells and also yields energy efficiency.

3GPP has adopted OFDMA as the multiple access technique on the downlink and SC-FDMA on the uplink as the
air interface for LTE and LTE-Advanced. Therefore, unlike UMTS intra-cell interference is not a problem anymore in LTE and LTE-Advanced as data is transmitted in orthogonal frequency or time slots. Inter-cell interference is still a major cause of low spectral efficiency and reduced system capacity. Inter-cell interference can either be mitigated through interference randomisation, cancellation or coordination. Interference randomisation is achieved through frequency hopping or spectrum spread techniques to ensure interference is equally distributed among the users. Interference cancellation exploits diversity [99] and its performance depends on the Degrees of Freedom (DoF) of the user i.e. A user receiving its desired signal and unwanted signals from 5 diverse sources, has a DoF of 5 and can theoretically cancel out all of the unwanted signals (interference). However, when the DoF is less than the sources of interference (which is the case in most multicell scenarios) interference is only reduced and not effectively cancelled. Inter Cell Interference Coordination (ICIC) has two further possible approaches i.e. static or dynamic. In the static approach interference is simply avoided through pre-planned orthogonal frequency allocations among neighbouring cells. In dynamic coordination, interference is dynamically managed by maintaining orthogonal resource allocation among neighbouring cells during the radio resource scheduling process. Dynamic ICIC schemes operate on short time scales and have more potential for performance improvement but at the same time are less scalable due to the need for cooperation and heavy signalling. On the other hand, static coordination schemes can operate on much longer time scales and are more suitable for the long time scale dynamics of the system as they are far less agile, hence less capable of coping with continuously changing dynamics of cellular systems. ICIC schemes have been widely studied compared to the other two inter-cell interference mitigation schemes. Below we briefly discuss the literature on both types of ICIC in context of self optimisation.

1 Long term Self Optimisation through ICIC

a Self optimisation via Integer Frequency Reuse: The most classic type of ICIC is via Integer Frequency Reuse (IFR) schemes where spectrum is divided into integer numbers of bands, and each band is reused in different cells such that the distance among the co-channel cells is maximised. IFR schemes provide a trade off between two major SO objectives of capacity and QoS i.e. the tighter the frequency reuse (closer to 1) the higher will be the capacity, but lower the QoS particularly for the cell edge users (critical users) and vice versa. As different IFRs offer different degrees of trade-off between the capacity and QoS, that would remain fixed in classic cellular systems, it is desirable to have a self organising framework that would make the most of the diversity of IFR schemes to yield a holistic solution. A reasonable attempt to achieve this has been proposed in [100]. Imran et al. proposed a framework for a self organising frequency reuse, sectorisation and relay station deployment for long time scale self optimisation in future cellular systems. The problem was formulated as a multi objective optimisation of three system objectives: spectral efficiency, fairness and energy efficiency. A utility function was defined to reflect operators policy that could be adapted to achieve desired prioritised objectives and to cope with changing spatiotemporal dynamics. This work is a good example of a solution that requires central
control and yet has most of the desired features of SO i.e scalability and agility. This is due to the fact that the amount of signalling required (5 bits from central control unit to eNB) is minimal that the solution maintains scalability even for a very large size cellular systems. Moreso the changes that trigger such signalling is only required on long term scale (days to months). The proposed solution thus remains agile due to the frequency of the signalling required.

b Self optimisation via Fractional Frequency Reuse: Integer frequency reuse of 1 is always optimal in terms of capacity but poor cell edge performance is a big hindrance to achieving this optimal performance. A way around this problem has been devised in the form of Fractional Frequency reuse (FFR) schemes in which a frequency reuse greater than 1 is used in cell edge areas whereas cell centers use a frequency reuse of 1. This ensures fairness and improves data rate and coverage of cell edge users as validated by results presented in [101]. A number of variants of FFR have been proposed and evaluated in [102]–[106] and shown to offer different levels of the capacity and QoS enhancements. Although the tradeoff between capacity and QoS is more flexible in the case of FFR compared to IFR [107], the recurring challenge in this tradeoff remains confined to a given FFR scheme. For self optimisation a framework for self organised FFR (SOFFR) as an extension of the one proposed for IFR in [100] is desirable. Authors in [108] attempted this but with more focus on self configuration of FFR patterns in the deployment phase of LTE networks than self optimisation during the operation phase. They use a novel hexagonal division approach with a clustering algorithm and showed a trade off between spectrum utilisation and the cascading effect of configuration changes over the entire network. Recent work by Rengarajan et al in [109] present a dynamic FFR scheme that can self optimise according to the system dynamics. The basic idea in this scheme is to dynamically create an FFR scheme that maximises system utility, through exchange of defined interference cost among neighbouring cells. With so many new dynamic FFR schemes already being proposed in the literature, an efficient self organised FFR scheme for relay based LTE system showing the effect of relaying on dynamic FFR patterns and how FFR patterns can change when new nodes are inserted or relay nodes go into idle mode, would provide new solutions for self optimisation of wireless cellular networks.

2 Short Term Self Optimisation through Dynamic ICIC

Authors in [110] formulate the dynamic ICIC problem as an optimisation problem. An integer linear programming technique is used to breakdown the problem into subproblems to reduce computational complexity. This approach has been demonstrated to improve performance of critical users [110] as well as maximises throughput [111]. A seminal paper by Stolyar et al. in [112] provides an interesting dynamic frequency reuse scheme complemented with an intuitive explanation on how each cell can selfishly try to minimise its users power usage based on their number of subcarriers and power requirements. This results in edge users being allocated “good subcarriers” with high power requirements and avoiding the use of those subbands for their own cell edge users in order to minimise total power usage. Simulation results were shown for constant bit rate traffic types and in [113] two different algorithms were demonstrated for best effort traffic type. This simple localised rule results in the entire system settling for a minimised power usage for all user - subband allocation. The algorithm is localised (scalability), does not suffer oscillations (stability), adapts readily to an evolving set of user requirements (agility). The authors in [114] designed a self organised spectrum assignment based on reinforcement learning that can help operators to define policies for the system to self organise its spectrum assignment and to achieve the best tradeoff between spectral efficiency and QoS thus creating more spectrum access for opportunistic secondary users.

Self optimising cellular networks by control of interference has been extensively studied in literature. The trend is rather than avoid interference we can employ coordination to minimise interference exploiting the dimension of frequency and geographical location of users. Coordination can either be long term employing IFR and FFR schemes or short term through dynamic FFR schemes. The current status in standards (3GPP Release 11) is the introduction of enhanced ICIC (eICIC) as current ICIC schemes do not efficiently address interference in Home eNBs which would form a pivotal part in future wireless cellular networks.

C. Self Optimisation of Capacity and Coverage via Relaying

Relaying has been identified by 3GPP as a means of providing cost effective throughput enhancement and coverage optimisation [115]. Different relay architectures have been proposed in [116] for LTE-Advanced. A prompt deployment of Relay Stations (RS) is desired to meet the two main objectives of capacity optimisation and coverage extension. Peng et al. in [117] demonstrated a self organising relay station architecture for multihop cellular systems. The proposed scheme is based on the premise that RS would be plug and play. This enables them to independently explore the environment and determine in cooperation with other nearby RS on which eNBs to establish a connection with. Results show lower call blocking rates and improved performance in traffic congestion and unexpected node failure compared to conventional cellular systems.

Assuming fixed relay stations, authors in [118] proposed a framework for self organising load balancing via a scheme termed Self organising Cooperative Partner Cluster (SCPC). Each node (RS and eNBs) selects partners (neighbouring nodes) for coordination of information on individual load status and available resources. Measurement reports and triggers for the algorithm are based on a load balancing policy and performance evaluation feedback. A key part of this work is that it demonstrates an optional feature to select a single, multiple or combination of various load balancing policies in accordance with the dictates of the network environment.
From the discussion in this section, we can infer that self optimisation of cellular networks has seen the greatest amount of research inputs compared to other phases (self configuration and self healing). The continuous operation of systems to ensure optimum performance in the face of changing dynamics such as environment, node mobility, varying data demand etc., requires autonomous intelligent response to maintain global system objectives. From the numerous use cases identified in this section, load balancing and interference coordination have been investigated more extensively in literature. It is foreseeable that with the rising need for energy efficiency, more efforts still need to be channelled towards increasing the operational energy efficiency in wireless cellular systems to minimise energy wastage. Reliable provision of differentiated QoS aware service can also be enhanced through self optimised solutions. It should be noted that most of the work done in the context of self optimisation, focus on conventional deployment architectures. Future networks would however feature variety of deployment architectures ranging from Femto and Pico cells to relay station. Therefore more work is required on having self organised association of such nodes, and operation in both active and idle states for reduced interference and energy efficiency gains.

VI. SELF HEALING

Wireless cellular systems are prone to faults and failures due to component malfunctions or natural disasters. These failures could be software related or hardware oriented. In legacy systems, failures were mainly detected by the centralised O&M software. Events are recorded and necessary alarms are triggered. When alarms cannot be cleared remotely, radio network engineers are usually mobilised and sent to cell sites. This process could take days and maybe weeks before the system returns to normal operation. In some unique cases, some failures are even undetected by the O&M monitoring software and cannot be addressed except when an unsatisfied customer lodges a formal complaint. In future self organised cellular systems, this process needs to be improved by incorporating self healing functionality. Self healing is a process that involves the remote detection, diagnosis and triggering compensation or recovery actions to mitigate the effect of faults in the networks equipment.

In Fig. 6, we illustrate a scenario where due to a fire outbreak, natural disaster of any other type of fault, a node is completely out of service. Using the self healing logical flow presented in Fig 7, neighbouring nodes are able automatically detect a fault, classify the type of fault and trigger necessary compensation actions. Compensation can either be through relay assisted handover, power compensation or antenna tilt reconfiguration of neighbouring nodes. Logical views and steps involved in the self healing procedure has been suggested in [119] for 3GPP. A general approach towards self healing is suggested that consists of basic elements of monitoring, diagnosis and compensation. Learning is also implicitly part of this approach as incorrect diagnosis based on erroneous correlation of alarms can be logged for a more intelligent diagnosis in future.

In Fig 7 we present a general approach in the form of a logical self healing flow. This has been distinctively divided into three phases: monitoring, diagnosis and compensation. The system continues monitoring the network for any anomaly or if any predefined conditions for self healing conditions are satisfied. If such conditions are met, data are analysed using Bayesian networks or any expert system to specify the type of faults (with a certain probability of accuracy) and recommend the optimum compensation action to be taken as well as which neighbours are required to compensate the faulty node. The neighbouring nodes also periodically listens to signalling from the faulty node via the X2 interface to establish if the node has been restored. Once restored, the neighboring nodes can return to their pre-compensation mode so as not to degrade overall system performance by interference. An important feature to be observed is the loop that continuously monitors the faulty cell to determine if it has been restored to normal operation. This is necessary to trigger neighbouring compensating cells to revert to their status quo.

Whilst a general framework for self healing has been portrayed, specific self healing functionalities are still emerging in literature though at a slow rate. A preliminary idea of using relay stations as self organising agents or more specifically as self healing agents was proposed in [120]. The concept of small scale free world was extended to relay based Radio Access Network (RAN) to show how the network can, amongst other features, be made robust and reliable. Relay nodes are used to reroute traffic to a new neighbouring eNB when the serving eNB fails or is congested. The three
A cell is said to be “in outage” when there is a dip in performance due to an abnormally operation. Future networks would employ intelligent cell outage detection schemes. Authors in [121] described three stages of cell outage. Degraded cells which are still operational but sub optimally; crippled cells where there is a major fault in the cell and SINR of users affected is very low thus a dip in the capacity of that cell; catatonic cells describing complete outage or system failure such as in the event of a disaster. Khanafer et al. in [122] considered an automated diagnosis that can be implemented in UMTS networks using Bayesian analysis. In the event of a fault, the system troubleshoots the fault and predicts the cause with a given probability. The accuracy is based on the initial training data and classification of fault types. In [121] a cell outage detection algorithm is also demonstrated which is based on monitoring changes in a visibility graph which will be generated in the neighbour cell list. A binary classification scheme based on pattern recognition was also described. The performance of this outage detection scheme is based on the accuracy of the classification algorithm. It was also pointed out that even with a perfect classification algorithm, there would still be undetected outage events, especially for low user traffic. In the detection of faults, an algorithm could raise a false alarm and trigger a compensation action even when there is no fault. It must be noted that in using a classification algorithm, not all observed patterns should trigger self healing process. When certain anomaly are detected and the particular type of fault identified (with the aid of expert systems), it is necessary to have a knowledge plane that determines if all necessary conditions (duration of anomaly, effect on system performance etc.) are met before the self healing process is initiated. These conditions are referred to as Trigger Conditions of Self Healing (TCoSH).

B. Cell Outage Compensation Scheme

The cell outage compensation actions are entirely based on the faults detected. TCoSH have to be met before compensation actions are initiated. Some faults can be automatically cleared whilst others have to be manually cleared which might require a visit to the cell site. In the former, compensation actions have to be triggered automatically. Compensation actions involve self reconfiguration of system parameters as explained in section III. Compensation actions in the event of complete outage as detected by neighbouring cell would initiate reconfiguration settings to decrease their antenna tilt, increase transmit power or even relay nodes can change association to new nodes and increase their transmit power to increase coverage or a combination of these actions. In [123] Amirijoo et al. present a cell outage management description for LTE systems. Unlike previous works they give a total overview of both detection and compensation schemes highlighting the role of the operator policies and performance objectives in design and choice of compensation algorithms. The importance of time scale for triggering compensation actions and possibilities of false detection were also highlighted.

The section has provided a better understanding of the importance of self healing phase in self organisation. To ensure high reliability of future networks, we have discussed the need to continuously monitor the system for any anomaly, analysis and diagnosis if anomaly is detected and trigger the right compensation action autonomously. Key advances in cell outage detection and compensation schemes in literature were discussed. However, the area of self healing has not gained much attention as compared to self optimisation functionality in SON and a substantial amount of work is required before full deployment of self healing functionality is a reality in future wireless cellular systems.

VII. METHODOLOGY

In this section we identify the possible approaches that have been taken, or could be taken, in designing SO functionality in cellular communication systems. The designing approach towards SO can be divided into two main groups: Learning and Optimisation. Learning could be in the form of supervised or unsupervised learning. Optimisation based approaches could be described as Deterministic or Stochastic optimisation techniques. Detailed illustration of possible methodologies to achieve self organisation are shown in Fig. 7. The figure also relates these approaches to their heritage from the literature. The associated references are by no means exhaustive. Nonetheless, they still present the main picture of ongoing attempts to design SO in wireless cellular networks and should help readers to find details on the use of these methodologies. In the following subsections we discuss some of these methods in more detail.

A. Learning Algorithms

As discussed above, self organising systems in nature reflect certain levels of intelligence and the most rudimentary step to acquire intelligence, is learning. Therefore it is reasonable to investigate learning algorithms to design SO in wireless networks [9]. An ideal reference in emulating learning and subsequent intelligence is the human brain. However perfect modelling of the brain is no trivial task and is still an ongoing study in the field of Artificial Intelligence (AI). The definition of intelligence is relative, but as described in the AI literature: Intelligence is simply the ability to understand and learn from previous events [124]. One of the major desired features of self organising cellular systems that would elevate its performance from conventional adaptive systems is the ability to learn from previous actions and improve performance i.e. reflecting intelligence. Humans learn either with the help of a tutor or independently. Similarly for machine learning, we can group all learning algorithms as supervised and unsupervised learning algorithms.

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 [125].
Albeit, when such systems are faced with new set of inputs they could choose incorrect actions which would lead to poor performance.

Unsupervised learning sometimes referred to as self-organised learning [125] involves the 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. Some important learning algorithms investigated include

1) 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 attempts to minimise the penalty received in each iteration and thus improves/learns to take actions that minimises a cost function. The feasibility of RL techniques lies in their continuous interaction with the environment. This would be easily implementable in future networks whose nodes would have cognitive abilities. However when the input-output space is continuous, an abrupt change to a different action when the learning threshold is narrowly exceeded periodically which leads to instability. Thus various modified RL schemes are often used. A number of authors have found RL suitable for developing self optimisation algorithms [114], [126] and [127].

2) Docitive Learning (DL): Docitive Learning (DL) is a new emerging paradigm recently addressed in [128]–[130] and promises an improvement over classic 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 be dynamically taught to choose its supervisor amongst a set of available nodes based on the correlation in its own operational environment and the environments of other accessible nodes. Thus the learning node can choose node(s) with a high degree of correlation as a role model rather than a classic supervisor, as is the case with conventional supervised learning. Thus the learning already acquired by the nodes with a high degree of correlation with the learning node, can be used directly, saving a lot of time and avoiding much complexity. DL techniques are very much promising and seem very feasible to implement. However, it is still largely an unexplored realm.

3) Game Theory: Game theory is a widely used unsupervised approach towards learning that involves individual agents trying to maximise their own selfish objectives. This usually would be at the detriment of other nodes involved in the game. An equilibrium point (Nash equilibrium) could be reached where each node cannot unilaterally take an action to improve its state, for the given state of other nodes in the system. Game theory has been used in modelling communication systems. Mackenzie in [131] showed how it can also be applied in modelling self configuring functionality through a power control game. There is however limited literature in applying game theory in self optimising LTE based communication systems as more work needs to be done in proving the convergence to an optimum solution for all agents (or nodes) modelled in the game (network). Operators are quite skeptical about applying game theoretic approaches for solutions that would involve zero human intervention. The practicality of defining rules for cooperative and non cooperative games across different operators who may have different policies and try to maximise different cost functions is very challenging.

4) Self organising Maps (SoM): SoM are special class of artificial neural network algorithms [132] that rely on the probability density function of the input data to act as efficient clustering schemes [133]. Laiho et. al. in [134] demonstrated an algorithm based on SoM that can be used to monitor network performance and user behaviour. It forms a cluster of cells that have similar features and can prompt similar configuration changes for all cells in the cluster. This cluster of cells may be located at different geographical locations in

---

*D.docitive is from the latin word docere which means to teach or transfer knowledge.*
the network. The concept of SoM has also been applied at the scheduling management level in [133]. The basic idea in this paper is to divide the cells with similar traffic requirements into clusters by using SoM and to use an appropriate scheduling scheme for each cluster. Significant improvement in performance using this cluster based scheduling scheme is shown as compared to a scenario where the same scheduling scheme is used in the whole network. The algorithm does not require global coordination after the clusters are formed, but in the dynamic process of making clusters, the amount of global coordination required, the convergence time and complexity of clustering process itself, especially in large networks, renders it devoid of the essential features in SO such as agility and scalability. More efficient application of SoM in self monitoring systems could be worthy of investigation. The feasibility of SoM is apt especially on the management plane for detecting any system anomaly by observing its performance and interrelation among different deployed SON solutions. It can be easily deployed at a centralised unit or localised clusters.

5) Cellular Automata: Cellular automata is a well known discipline in cybernetic systems for developing autonomous systems using inspiration from biological complex and yet autonomous systems that are all made up of large numbers of small cells. The basic idea is that a system that needs to be automated is modelled as an aggregation of a large number of small cells. And each cell follows simple rules defined by a mathematical function and updates their individual states based on its current state and that of the neighbouring cells [135]. The reader can refer to [136] for details in modelling dynamic systems using cellular automata. Authors in [137] have taken this direction and have developed SO channel assignment schemes using cellular automata theory. Since SO systems in nature also have similar operational principles to CA, so we can infer that CA is one of the most natural approaches towards designing SO [138]. The two main challenges in the feasibility of deploying CA based solutions are the system convergence and determination of the exact system response to irrational changes in environment. It is thus necessary to investigate their reliability for self organised systems. Practical application of cellular automata is has not yet been extensively investigated for wireless cellular systems.

while neglecting others may lead to unstable or non pragmatic solutions. Below we discuss both kinds of approaches.

1) Distributed Optimisation: If the system wide optimisation objective of the SO, can be broken down to local optimisation functions to be addressed by local entities in the system, optimisation of which is dynamically maintained by each of the entities independently or semi independently, an adaptive scalable, stable and agile solution can emerge i.e. SO can be achieved that can manifest automaticity just as in a school of fish. While this approach seems deceptively straightforward, its hidden difficulty lies in the transformation of system wide objectives into local functions that are simple (agility), will not require global cooperation (scalability) and yet still merge into a consistent global behaviour (stability).

2) Multi-objective Optimisation: Cellular communication systems are inherently complex as a change in one system parameter intended to control one objective has an effect on other objectives. For example, increasing the transmit power in a cell would increase its coverage but also increase the inter-cell interference which in turn will decrease capacity and may affect energy efficiency as well. Therefore, as discussed above, in order to ensure stability of the SO solution classic single objective optimisation techniques have to be extended to take into account multiple objectives either as simultaneous optimisation objectives or as well defined feasibility constraints. The former approach is more desirable as it can optimise multiple objectives simultaneously yielding a holistic SO solution, but it is equally more difficult to produce.

In this section we have presented a range of methodologies that can be applied or have potential to be used in incorporating self organisational functionality in cellular networks. The role of learning algorithms was explained with details on the feasibility of applying Reinforcement learning, docitive learning, game theory and self organising maps. Distributed optimisation schemes have also been employed in literature with interesting results. We thus infer that achieving SO functionality in cellular networks could involve an interdisciplinary research with tools from machine learning/artificial intelligence. However, the method used either via learning algorithms or through optimisation techniques is not the vital part but the nature of the solution presented.

VIII. SO FROM CONCEPT TO REALISATION

The results from pioneering papers in the open literature, major research projects dedicated to SO and how these have contributed to evolving standards and activities in 3GPP are summarised in this section.

A. Evolution of SO in the Cellular Communication literature

Although much work has been done on self organisation for wireless ad hoc, mesh and sensor networks relatively few works focussed on self organisation in wireless cellular networks. In the focussed context of cellular systems, the term self organisation was first used by authors in [139] although in a limited sense of adaptiveness in the power control in GSM. Both [139] and [140] presented a simple adaptive power control algorithm to show how this kind of adaptability can mitigate the effect of BS positioning error.
Contributions such as [139]-[133] suggested the need for self organisation in cellular networks and presented a number of adaptive algorithms under the umbrella of self organisation but did not present an explicit design strategy to build real self organisation into cellular systems.

A second generation of papers in self organisation in cellular networks ( [14] and [120]) embarked on an holistic approach towards self organisation and postulated a set of germane principles and paradigms to be considered in the design process for self organising systems but stopped short of demonstrating these principles through a pragmatic application to specific problems in cellular systems. More recently, papers such as [13] - [100] focussed on designing self organising solutions for various problems in cellular systems but some of the solutions presented, cede the useful level of adaptability at the cost of compromising one or other important feature of SO i.e. scalability, stability or agility. Table I gives an analysis of some selected solutions and how the solutions proposed addressed the stability, scalability and agility. As in all generic solutions, stability requirement was satisfied in all papers studied. However the main intricacy comes from achieving scalability, stability and agility simultaneously. In [44] and [113], self organising solutions proposed are stable, scalable and agile with details summarised in Table I. However in [114] the self organising solution presented requires feedback from all cells in the entire network at the central controller. This requires alot of signalling overhead and increased complexity as the size of the network increases. Such solutions whose order of complexity increases with the size of the problem especially in some centralised systems. Management of such systems or seamless upgrade with emerging technologies becomes unfeasible. More analysis of other selected papers are presented in the table.

A resourceful textbook on autonomic network management principles can be found in [141]. It describes different solution models spanning concepts, architecture, services and security. Two dedicated chapters deal with proposals for self organising network architecture as well as autonomies in radio access networks.

B. Major Research Projects on SO in Cellular Systems

The interesting research problems highlighted in individual papers and increased understanding of the need for self organisation in future wireless cellular systems has prompted research via a number of large research projects in recent years dedicated to investigating, analysing and evaluating the possible deployment of self organisation functionality. Here we present a summary of the aims and achievements of several such projects on SO.

European Celtic Gandalf project [142] identified the heterogeneous nature and increased complexity of future wireless networks, and studied SO as a potential solution to overcome the complexity expected from concurrent operation of 2G, 2.5G, 3G and future network solutions. It evaluated the effect of automation in network management and joint radio resource management in WLAN/UMTS networks with automated fault troubleshooting and diagnosis. A major deliverable of this project was developing an auto troubleshooting toolset using Bayesian networks. The European "end-to-end efficiency (E3)" project [143] focused on integrating current and future heterogeneous wireless systems into cognitive systems for self-management, self-optimisation 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 future networks.

Another major project dedicated to SO is the European SOCRATES project [6] studies the concepts, methods and al-

<table>
<thead>
<tr>
<th>Paper</th>
<th>Scalability</th>
<th>Stability</th>
<th>Agility</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>[44]</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>Minimal iteration count parameter $I_{AMS}$ to prevent instability. Algorithm tested for different network sizes and applicable for larger networks as its complexity does not increase with size of network.</td>
</tr>
<tr>
<td>[112]</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>CQI feedback over the last 50 slots needs to be sent out every slot which increases complexity for larger sub bands and is thus not scalable. By intuition algorithm may be stable but not proven. Not agile as it assumes all users are indistinguishable (regardless of traffic type).</td>
</tr>
<tr>
<td>[113]</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>Algorithm scalable as it focuses only on local minimisation of system utility. Stable as power allocation &quot;jitter&quot; has little effect on system utility resulting value. MGR algorithm is agile as it readjusts sub-band power levels to near optimal values quickly from almost any initial power setting.</td>
</tr>
<tr>
<td>[114]</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>Not scalable as it uses a centralised controller which gets feedback from all cells. A &quot;decision maker&quot; checks for convergence for a final chunk-to-cell assignment which ensure stability. Agile as RL-DSA algorithm continuously interacts with environment.</td>
</tr>
<tr>
<td>[100]</td>
<td>YES</td>
<td>N.A.</td>
<td>YES</td>
<td>Solution complexity does not increase irrespective of any FRD configuration used. Solution proposed is intuitively stable but not addressed. Simple rules are used to adapt utility for changing spatiotemporal dynamics.</td>
</tr>
</tbody>
</table>

N.A. = Not Addressed
algorithms for SON with specific focus on LTE radio networks. An exhaustive list of use cases SO functionality can achieve in cellular systems were identified. It is further concluded that most use cases are not independent from each other as the parameters controlling them are not mutually exclusive making joint optimisation difficult if not impossible. They have thus proposed a SON coordinator that incorporates operators policies with control actions to ensure individual SON functions work towards the same goal. They have also gone into great details to describe a framework showing the interoperability between these major use cases for major tasks involved in self organised wireless networks which include self-configuration, self-optimisation and self-healing functionalities in future radio access networks. Details of the final deliverable from the project can be found in [144].

A recent project focusing on the growing management complexity of future heterogeneous networks is UniverSelf [145]. This project acknowledges the recent advancements in autonomies but also the lack of similar strides in the management and reusability of such autonomic mechanisms. The goal is to have a global autonomic management system via a unified management framework for all existing and emerging system architectures and to incorporate the intelligent and self-organising functionalities right into the network equipment (intelligence embodiment). To achieve this, the project designs and evaluates self-organising algorithms with cognitive capabilities. The project employs a use-case based approach in order to address valid problems that the operators face and/or are expected to face in the future. Test beds will finally be deployed for experimental validation.

A dedicated collaborative research on self organising cellular multi-hop networks by India-UK Advanced Technological Centre (IU-ATC), Theme 9 [146] investigates the fundamental performance bounds and system capacity of self-organising cellular multi-hop networks, and then developed a set of practical algorithms to demonstrate the self organising functionalities. The collaborative research aims to demonstrate that self organisation can help in achieving higher energy efficiency, reduce CAPEX/OPEX and is essential in disaster/emergency scenarios. Fundamental performance bounds and system capacity of these networks are investigated with a set of practical self organising algorithms for a variety of broadband wireless services and applications. A test-bed of a self-organising wireless LAN to show feasibility of algorithms developed is also being demonstrated.

With the introduction of indoor base stations and the need for self organisation among them, European project BeFemto [7] aims to develop evolved Femtocell technologies based on LTE-A that would enable a cost-efficient provisioning of ubiquitous broadband services and support novel usage scenarios such as networked, relay and mobile Femtocells. The project focuses on novel concepts and usage scenarios such as self-organising and self-optimising Femtocell networks, outdoor relay Femtocells as well as mobile Femtocells. This project is expected to have a major impact on the standardisation of the next generation Femtocell technologies based on LTE-A.

C. Standardisation Activity on Self Organisation

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 organising functionality was a solution towards improving operational efficiency of future networks [22]. Future network architectures are flat and simplified in order to reduce 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 up working groups aimed at coming up with specifications for agreed descriptions of use cases and solutions with emphasis on the interaction of self optimisation, self configuration and self healing. As at the time of writing this survey, Release 10 of the 3GPP technical specifications has been completed and Release 11 which would focus on enhancements is been discussed in respective working groups. This would include discussions on enhanced intercell interference coordination (eICIC) schemes.

The main documents describe concepts and requirements for self organising networks [147] as well as specific documents on self establishment of eNBs [148], Automatic neighbour management [149] self optimisation [150] and self healing [119] have also been produced and released. A key white paper that describes in simple terms the details of SON mechanisms as detailed in 3GPP standards Release 8, 9 and 10 can be found in [151] and [152]. They focus on self configuration and self optimisation to lower the cost of system and increase its flexibility through incorporating intelligence and automation.

IX. Future Research Directions

In this section we present potential challenges and directions for future work to implement multiple self organising functionalities in future cellular systems.

A. Docitive learning based Self Configuration

Most of the state of the art self configuration mechanisms rely on one of the following two approaches: Centralised static approach or Distributed dynamic approach. In a centralised static approach each new node configuration parameters are determined by a central entity. While this approach is stable and can be agile to a limited extent if the backhaul is well dimensioned and the network has a small number of nodes, however this is not the case for really large networks. On the other hand for the distributed dynamic approach, configuration parameters are determined locally by learning the current states of neighbour nodes. While this approach is scalable,
it may compromise stability and agility, because of the large convergence times involved, partial reliability of state of the art neighbour discovery mechanisms (see section on neighbour discovery) and the fact that the presumed neighbours might not have exactly the same environment as that of the existing node, hence intelligence learned can be misleading. A possible direction to address this challenge is to exploit doctive learning in distributed dynamic self configuration where learning should include an additional phase of tutor or supervisor selection as explained in subsection VII-A2.

B. Improving Neighbour List Discovery

As concluded in the subsection on neighbour list discovery, state of the art, neighbour list discovery mechanism do not yet provide a perfect solution but compromise on either agility, scalability or stability. A possible new approach could be to use a hybrid of position and SINR based approaches where the neighbour list is generated by weighting both the SINR measurements and positioning information. The weighting can be adapted dynamically through one of the learning or optimisation approaches discussed above so as to take into account the type of environment in which node is placed.

C. Designing Scope of Coordination among eNB

Most SO would require local coordination among the local nodes and the scope of this coordination is heavily dependent on the type of SO algorithm. 3GPP has agreed on an X2 interface among eNBs that can be used for such coordination. Different SO algorithms may require coordination among different number of neighbours i.e. they can have different scopes. As 3GPP currently does not specify the scope of the X2 interface, it is necessary to investigate the topology and scope of X2 interfaces that are flexible enough to yield diversity for SO algorithm implementation.

1) Coordination Mechanisms for Home eNB: 3GPP currently does not specify an X2 like interface for the control plane among Femtocells. With the number of indoor base stations expected to grow exponentially coupled with their mobility and opportunistic deployment, SO becomes more of an inevitable feature rather than just an added functionality [7]. SO requires at least local coordination, there is thus a dire need to investigate and design such an interface or mechanisms to enable this coordination not only among neighbouring Femtocells but also among Femtocells and neighbouring macro cells so as to avoid cross layer interference and other conflicts resulting in instability and poor performance. However a limitation of the X2 interface is latency. Delays associated with communication among eNB over this interface in live networks are significant and thus actions taken based on information/signalling with other nodes may no longer be necessary due to changing channel conditions. Another limitation is the the ambiguity in specification on X2 communication among eNBs of different operators.

D. SO of Antenna Parameters in Heterogeneous Network

Optimal solutions to configure the antenna tilt and power of neighbouring nodes when a new cell is inserted or when there is a system failure in any cell, is still lacking in the literature. The literature have focussed more on evaluation of the effect of centralised antenna parameter optimisation than on system performance. A truly SO solution for antenna parameter configurations and optimisation, particularly for heterogeneous future wireless cellular networks (containing macro, pico and femtocells and RS), is yet to be realised.

E. Issues in Self Optimisation of Multiple Objectives

Self optimisation still remains a broad and open area needing more research as many optimisation objectives have been identified and a large number of known parameters (up to $10^9$) that need to be adjusted dynamically for continuous self optimisation [153]. However, the coupling of one objective to another due to the commonality in the sets of their controlling parameters is still a problem to be addressed. Some authors [154] are of the opinion that a complex coordination plane is inevitable to overcome this challenge. This coordination plane may affect the scalability of the SO solutions. We have however identified how application of multi-objective optimisation techniques [100] can be used to resolve this issue. A specific extension could be to extend the framework currently presented to FFR and Femtocells in addition to IFR and outdoor relays. A potential approach to overcome the coupling problem is to exploit distributed optimisation either analytically or in an evolutionary heuristic way based on a target pareto optimal solution of local sub problems instead of targeting the overall optimal solution of the problem as explained in VII-B1.

1) Parallel operation of multiple time scale SO: In future cellular systems, multiple SO algorithms might be required in order to achieve different objectives (see Fig. 1 ). These algorithms would operate simultaneously but at different time scales as explained in Fig. 2. There are some common control parameters, and there is a need to investigate their mutual dependency and the effect of potential overlapping among these algorithms that might cause instability.

2) Parallel operation of multiple spatial scope SO: In addition to having different scopes in time, various SO algorithms might have different scopes 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 holistically smooth operation of SO in future cellular communication systems. In a given cellular system, one SO algorithm, say for energy efficiency, may need to operate over a large number of nodes cooperatively and another SO in the same network, say for capacity optimisation, might be operating locally. Since these two SO algorithms might have to use the same parameters of each node, say transmission power, proper coordination among them is required to avoid instability.

F. Open Research Issues on Self Healing Functionalities

While few research papers have focussed on self healing functionality, more research is still required in this area particularly to evaluate the best tradeoff between accuracy in making a quick detection whilst minimising probability of false alarm. It is desired to demonstrate how future systems
would autonomously detect a fault and accurately specify the nature of the fault. It is also desirable to produce simulation results to show that the compensation actions are halted when the cause of a fault has been externally cleared. The acceptable time frame for detection, analysis and compensation actions needs to be standardised by the 3GPP work group.

With the drive for energy efficient networks and enabling cells with low traffic to hibernate, it is also important to generate efficient monitoring schemes such that neighbouring cells do not erroneously trigger compensating actions for a hibernating cell rather than a cell in outage.

G. Challenges in Designing Self Healing Functionalities

The main research problems in self healing emanates from the difficulty of modelling and simulating the environmental and technical anomalies over large time scales. Alternatively, if an analytical approach is used, recovery probabilities can be predicted using Markov chain analysis or Bayesian analysis. Validating their probability via an appropriate system level simulator would be a subject worth exploring.

H. Designing Evaluation Methodologies for SO Algorithms

While SO requires a significant upgrade over the classic adaptive algorithms, tools to validate SO algorithms also require an upgrade. Classic system level simulators are usually designed to evaluate short term adaptive algorithms, e.g., scheduling, Transmission Control Protocol (TCP), power control etc. In order to assess the performance of SO functionalities like self configuration, multi objective self optimisation, or self healing, different forms of simulation and evaluation methodologies are required. The key feature of such simulation tools would be their capability to capture a much larger picture of the system, both in time and space.

I. Enabling Self Organisation

With much activity on SO solutions, attention also needs to be channelled in enabling SON algorithms/solutions. Enabling SO is key as we seek to implement SO in future networks through a gradual evolution from reduced human intervention to minimal human intervention and finally to zero human intervention in deploying optimising and managing future cellular networks. Key enablers for SO, would be seamless and scalable algorithms for autonomous coverage estimation, service estimation, autonomous cell boundary estimation, autonomous energy consumptions and QoS estimation. All of these estimation algorithms will essentially provide the necessary data to trigger appropriate SO, eliminating the need for human intervention and making future wireless systems truly SO. Determining the right Key Performance Indicators (KPIs) for SO systems is also an open area and would be one of the first steps toward enabling full SO.

X. Conclusion

Self organisation in wireless cellular communication networks has been reviewed. From presenting a deep understanding of what these new functionalities of future networks are, to the important standardisation work going on, we have been able to highlight major research projects in this area and have discussed important results achieved thus far. Three generations of research papers in this field have been identified. The first generation include papers (e.g. [139] and [133]) where the need and expected gains of SO in cellular networks were outlined, but no explicit designs on how it could be achieved in cellular systems was presented. A second generation of papers (e.g. [14] and [120]) provided us with principles and paradigms in designing SO systems, although no algorithms for cellular networks were demonstrated. The third generation, which includes contributions in [13], [113], [114] and [100], build on the previous work to present algorithms and solutions termed self organising. Although a careful analysis shows that some solutions in the literature are classic adaptive algorithms, others possess necessary features (scalability, stability and agility) required in any SO solution. To clarify any ambiguity of what is and what is not self organisation, we have presented a characterisation frame work for SO and used it to discuss the state of the art literature with simple classifications of self configuration, self optimisation and self healing. Table I also provides an insight of the analysis contained in some key selected papers in this area, highlighting how their proposed solutions address scalability, stability and agility. Furthermore, we have discussed approaches used in the literature for designing SO and also suggested others that have not so far been fully exploited. We also highlighted a number of open research issues on SO in wireless cellular systems with suggestions for potential approaches for solutions. We hope that this comprehensive survey will act as a basis for new research in wireless cellular communication systems and act as a pointer to the unsolved problems in future networks known broadly as self organising cellular networks.

ACKNOWLEDGMENT

The authors would like to thank reviewers for their time, comments and constructive criticism.

The research leading to this work was performed partly under the India-UK Advanced Technology Centre of excellence in next generation networks systems and services funded by EPSRC in UK and partly under the framework of the ICT project ICT-4-248523 BeFEMTO, which is partly funded by the European Union. The authors would like to acknowledge the contributions of their colleagues from the IU-ATC and BeFEMTO consortium.

REFERENCES


[125] Universe project http://www.univerself-project.eu/about-univerself.


Osianoh Glenn Aliu received his B.Sc. degree in Electrical Electronics Engineering from University of Lagos, Nigeria, in 2006, and M.Sc. in Mobile and Satellite Communications at the University of Surrey in 2009 where he finished with a distinction. He is currently studying towards a Ph.D. degree in the Centre for Communication Systems Research (CCSR) at the University of Surrey, UK. He has been actively involved in an EPSRC funded project: India UK Advanced Technology Centre and collaborates with researchers in UK and India. His main research interests include physical layer modelling wireless communication networks and self organisation in wireless cellular networks.

Ali Imran received his BSc in Elect Engineering, in 2005 from University of Engineering and Technology, Lahore, Pakistan. He received his MSc in Mobile and Satellite Communications with distinction, and PhD, both from university of Surrey, UK, in 2007 and 2011 respectively. From Nov-2005 to Jan-2007 he worked as cell planning and BS deployment team leader for major cellular operators of the country. Since, Aug-2007 till Nov-2011, he has been working in Center for Communication Systems Research (CCSR) at University of Surrey, UK, where he was involved in various EU projects, such as, DYZINE, Self Organization theme of ICT-India UK framework and BeFemto. Currently, he is working as Research Scientist, at Qatar University Wireless Innovation Center, (QUWIC), Doha, Qatar, where he is involved in and leading research projects focused on mobile broadband wireless access technologies. His research interests include Radio resource management, and self organization in cellular networks and heterogeneous networks.

Muhammad Ali Imran received his B.Sc. degree in Electrical Engineering from University of Engineering and Technology Lahore, Pakistan, in 1999, and the M.Sc. and Ph.D. degrees from Imperial College London, UK, in 2002 and 2007, respectively. He secured first rank in his BSc and a distinction in his MSc degree along with an award of excellence in recognition of his academic achievements, conferred by the President of Pakistan. He is currently a lecturer in the Centre for Communication Systems Research (CCSR) at the University of Surrey, UK. He has been actively involved in European Commission funded research projects ROCKET and EARTH, Mobile VCE funded project on fundamental capacity limits and EPSRC funded project India UK ATC. He is the principal investigator of EPSRC funded REDUCE project. His main research interests include the analysis and modelling of the physical layer, optimisation for the energy efficient wireless communication networks and the evaluation of the fundamental capacity limits of wireless networks.

Barry Evans (BSc, PhD,FREng, Ceng, FIET, SMIEEE, FRSA). Until 2009 was Pro Vice Chancellor for Research and Enterprise at the University of Surrey and Director of the Centre for Communication Systems Research which is a 150 strong post graduate Centre. Barry has a personal research background in satellite and mobile communications, researching in speech coding, radio propagation, advanced physical layers and cognitive radio with over 500 research publications and three text books to his name. He now runs a number of research projects in satellite communications with ESA and industry and leads the Surrey part of the UK-India network of excellence in next generation networks. He also leads the communications platform in Surrays Knowledge Transfer Network. He is Editor of the International Journal of Satellite Communications, an OFCOM spectrum advisory board member, chair of the steering committee of SatNEx-an EU/ESA network of excellence, a member of the steering council of the Integral Satellite Initiative and Director of a spin off company Mulsys Ltd. He is a Fellow of the Royal Academy of Engineering. Email: b.evans@surrey.ac.uk Telephone: ++44 1483 689131