

Some parts of this thesis may have been removed for copyright restrictions.

If you have discovered material in AURA which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown Policy](#) and [contact the service](#) immediately

Finding Structures in Information Networks using the Affinity Network

MICHEL F. RANDRIANANDRASANA

Doctor Of Philosophy



ASTON UNIVERSITY

April 2011

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

ASTON UNIVERSITY

Finding Structures in Information Networks using the Affinity Network

MICHEL F. RANDRIANANDRASANA

Doctor Of Philosophy, 2011

Thesis Summary

This thesis proposes a novel graphical model for inference called the Affinity Network, which displays the closeness between pairs of variables and is an alternative to Bayesian Networks and Dependency Networks. The Affinity Network shares some similarities with Bayesian Networks and Dependency Networks but avoids their heuristic and stochastic graph construction algorithms by using a message passing scheme.

A comparison with the above two instances of graphical models is given for sparse discrete and continuous medical data and data taken from the UCI machine learning repository. The experimental study reveals that the Affinity Network graphs tend to be more accurate on the basis of an exhaustive search with the small datasets. Moreover, the graph construction algorithm is faster than the other two methods with huge datasets.

The Affinity Network is also applied to data produced by a synchronised system. A detailed analysis and numerical investigation into this dynamical system is provided and it is shown that the Affinity Network can be used to characterise its emergent behaviour even in the presence of noise.

Keywords: Affinity Network, Bayesian Network, Dependency Network, Clinical Decision Support, Microelectromechanical Systems, Emergent Collective Behaviour.

Contents

1	Introduction	1
1.1	Graphical models	2
1.2	Graphical models and clinical decision-support	5
1.2.1	Clinical decision-support	5
1.2.2	Examples of CDSS	6
1.3	Emergence of collective behaviour in MEMS	9
1.3.1	Collective behaviour in MEMS	9
1.4	Thesis contribution	13
1.5	Notation	14
1.6	Thesis structure	14
2	Bayesian and Dependency Networks	16
2.1	Graphical and probabilistic backgrounds	16
2.2	Bayesian and Dependency Networks definitions	18
2.3	Learning Bayesian and Dependency Networks	19
2.3.1	Parameter learning	19
2.3.2	Structure learning	23
2.4	Probabilistic inference	29
2.4.1	Inference in BNs	29
2.4.2	Inference in DNs	31
2.5	Summary	33
3	Affinity Network	35
3.1	Affinity propagation	36
3.2	Soft-Constraint Affinity Propagation	39
3.3	Parameter and structure learning for ANs	42
3.3.1	Parameter learning for ANs	42
3.3.2	Structure learning for ANs	42

3.3.3	Quantification of the arcs' strength in an AN	44
3.4	Experimental results	45
3.4.1	Datasets	45
3.4.2	Comparison between AN and DN	48
3.5	Joint probability approximation	55
3.5.1	Synthetic model	55
3.5.2	Experimental methodology	55
3.5.3	Experimental results	56
3.6	Summary	58
4	A coupled dynamical system example	60
4.1	Single unit of a Duffing resonator coupled to a van der Pol oscillator	62
4.1.1	Multiple time scales analysis	62
4.1.2	Stability analysis	68
4.1.3	Asymptotic dynamics of the coupled system	71
4.2	Square lattice of driven resonators and oscillators	73
4.2.1	Experimental setup	73
4.2.2	Parameter effects	75
4.2.3	Architecture and size effects	85
4.3	Summary	88
5	Probabilistic square lattice	90
5.1	Synchronisation in random networks of oscillators	91
5.1.1	Characterisation of synchronisation in the probabilistic square lattice	92
5.2	Numerical results	93
5.2.1	Small noise added to the natural frequencies of the oscillators	94
5.2.2	Small noise added to the damping	94
5.2.3	Random links	95
5.2.4	Random links with various coupling strengths	97
5.2.5	Random driving sites	103
5.3	Summary	104
6	Cluster synchronisation	109
6.1	Overview	110
6.2	Cluster synchronisation and Affinity Network	111
6.2.1	Experimental results	112

6.3	Summary	120
7	Conclusion	122
7.1	Graphical model and existing instances	122
7.2	The AN graphical model	123
7.3	AN for medical data	124
7.4	AN of MEMS dynamical data	125
7.5	Limitations of the AN and future directions	126
7.5.1	AN parameters	126
7.5.2	Inference and belief propagation	126
7.5.3	Inference in MEMS	127
7.5.4	Pinning control	127
A	AN message-update equations	128
B	Dimensionless equations	133
B.1	Duffing resonator	133
B.2	Van der Pol oscillator	134
B.3	Coupled driven Duffing resonator and van der Pol oscillator	135
C	Multiple time scales analysis	136
	Bibliography	156

List of Figures

1.1	A medical example illustrating the differences between a BN, a DN and an AN.	4
1.2	Part of the Bayesian network modelling the evolution of a cancer of the nasopharynx (Galán et al., 2002).	7
1.3	Bayesian network structure of the PROMEDAS system.	8
1.4	The Affinity Network (AN) provides an interpretation layer on top of the physical layer composed of a square lattice of nonlinear dynamical systems.	10
1.5	An example of a complex behaviour that can occur in an array of 32 by 32 alternating Duffing resonators and van der Pol oscillators.	12
2.1	Markov equivalence class and DAG pattern.	18
2.2	An example of a Gaussian Bayesian Network.	21
2.3	Evidence nodes in a BN.	30
3.1	Factor graph for Affinity Propagation in the context of AN structure learning.	40
3.2	Comparison between the AN structure resulting from a combination of SCAP and the BDe score and a DN structure obtained from a classification tree and on a synthetic data with discrete variables.	50
3.3	Comparison between the AN structure learned using a combination of SCAP and the BDe score and the DN structure obtained from a classification tree and on a small real data with discrete variables.	52
3.4	Comparison between the AN structure from a combination of SCAP and the BDe score and the DN structure obtained from a classification tree on a small real data with continuous variables.	53
3.5	AN network structure of the ALARM data from a combination of SCAP and the BDe score and DN network structure from a classification tree.	54
3.6	True benchmark network structure with three random variables and a loop used in the synthetic data generation.	55

3.7	Approximation of a joint probability distribution by an AN, a DN and a BN.	57
4.1	Frequency-response curves of the driven Duffing resonator weakly coupled to a van der Pol oscillator.	65
4.2	Frequency-response curve of the driven Duffing resonator strongly coupled to a van der Pol oscillator.	67
4.3	Phase portraits and time series of a driven Duffing resonator strongly coupled to a van der Pol oscillator.	69
4.4	Bifurcation diagrams and maximum Lyapunov exponents of the driven Duffing resonator strongly coupled to a van der Pol oscillator with varying frequency Ω of the driving force.	72
4.5	Square lattice of alternating Duffing resonators and van der Pol oscillators.	74
4.6	Coupling strength effect on the global coherence of the square lattice. . . .	76
4.7	Phase and normalised polar plots of a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$	77
4.8	Phase portrait of selected elements in an array of 32x32 x -coupled alternating Duffing resonators and van der Pol oscillators. The coupling strength between the oscillators is $\kappa = 1$	78
4.9	Phase portrait of selected elements in an array of 32x32 x -coupled alternating Duffing resonators and van der Pol oscillators. The coupling strength between the oscillators is $\kappa = 100$	79
4.10	Effect of the amplitude of the driving force on the global coherence of the square lattice.	80
4.11	Plots of the instantaneous phase of components in a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators at a coupling strength of $\kappa = 10$	80
4.12	Selected phase portrait plots in an array of 32x32 x -coupled alternating Duffing resonators and van der Pol oscillators.	81
4.13	Relation between the quality factor Q of the elements of the square lattice and its global coherence.	82
4.14	Phase and normalised polar plots of a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$	83

4.15	Phase portrait of selected elements in an array of 32x32 x-coupled alternating Duffing resonators and van der Pol oscillators. The coupling strength between the oscillators is $\kappa = 10$	84
4.16	Impact of the network architecture on the square lattice global coherence.	85
4.17	Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators, and alternating van der Pol oscillators and Duffing resonators at a coupling strength of $\kappa = 40$	86
4.18	Lattice-size effect on the global coherence of the elements.	86
4.19	Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators for different sizes of the lattice.	87
4.20	Phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators. The array is driven from the top left corner of the lattice.	87
5.1	Natural frequency mismatch effect on the global coherence of the square lattice.	95
5.2	Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators with and without a small uniform random noise added to the natural frequencies ω_0 of the oscillators.	96
5.3	Damping noise effect on the global coherence of the square lattice.	97
5.4	Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators with and without a small Gaussian noise added to the damping term of the oscillators and resonators.	98
5.5	Effect of extra uniform random connections on the global coherence of the square lattice.	99
5.6	Phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators with different number of uniform random links.	100
5.7	Synchronisability and statistics in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	101

5.8	Intensity distributions in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	103
5.9	Effect of extra uniform random connections with various coupling strength on the global coherence of the square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	104
5.10	Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators with 11 extra uniform random links.	105
5.11	Effect of the number of random driving sites rs on the global coherence of the square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	106
5.12	Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators with different number rs of randomly selected sites as the source of the driving.	107
6.1	Characterisation of cluster synchronisation using AN and DN with varying coupling strength κ in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	115
6.2	Characterisation of cluster synchronisation using AN and DN with varying damping through the quality factor Q in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	116
6.3	Characterisation of cluster synchronisation using AN and DN with a different location of the external driving force in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	117
6.4	Characterisation of cluster synchronisation using AN and DN in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators with 11 extra uniform random links.	118
6.5	Characterisation of cluster synchronisation using AN and DN with different number rs of randomly selected sites as the source of the driving in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators.	119
6.6	Comparison between AN and k-means.	120

List of Tables

1.1	Comparison between a BN, a DN and an AN.	3
1.2	Notation.	14
3.1	Chest clinic dataset attribute information.	45
3.2	Distribution in 633 cases of primary invasive breast cancer of the clinical and pathological variables from the Ferrara breast cancer dataset.	47
3.3	Wisconsin breast cancer dataset attribute information.	48
3.4	Comparison between the network structures of small datasets.	49
3.5	Computational time for AN structure construction on big data using a classification tree and a combination of SCAP and the Bayesian score.	51
3.6	Joint probability table of the three random variables A , B and C	56
3.7	Consistency of the conditional distributions of an AN, a DN and a BN.	57

Acknowledgements

First of all I would like to thank my supervisor David Lowe for giving me the opportunity to work on this thesis. It has been a real pleasure and privilege to work with David and I have benefited greatly from his research experience. I am also grateful to Xueyong Wei for his valuable assistance in the MEMS project. Gabriele Migliorini, Rita Gupta and Chenchen Deng have also helped me in different ways when I started the latter project. My thanks also go to my thesis examiners and especially Ian Nabney for his thorough review, Elia Biganzoli for giving us access to the Ferrara breast cancer dataset, Diar Nasiev for providing the Matlab software to display graphs, David Saad for his useful comments on my first year report, Alex Brulo for his assistance with my computer, Vicky Bond and Kanchan Patel for their administrative help and my colleagues within the NCRG: Raje, Diar, Thomas, Ming, Erik, Jack and many others.

Last but not least, I would like to express my gratitude to my Father, family and friends for their love and support while working on this thesis. Special thanks go to my parents Etienne and Hanitra, my brothers Patrice, Jimmy and Joël, my aunties Olga, Lydia and Jocelyne for their support during my studies, the Unsworths (Graham, Annie, Ben and Jenny) for their love and warm welcome when I first arrived in the UK and for treating me like one member of the family during my stay with them, and my friends Nathanaël, P'Nong, Kyu, Bora, Sam, Han, Song, Sarah, Jane, Sue Ann, John, Nortin, Ming-Chin, Min, Simon and Shirley, Salem, Emmanuel, Wayne, Jason and many others for their love.

1 Introduction

This thesis concerns the development of a new type of graphical model focussed on providing a description of probabilistic relationships driven by similarity clustering. Graphical models are the fruits of a desire of researchers from many fields, including computer science, engineering, mathematics, physics, neuroscience, and cognitive science, to build systems that can adapt to their environments and learn from their experience. As the name suggests, graphical models are a combination of probability theory and graph theory. These two theories enable the existence of an important feature of a graphical model which is called *modularity*, that is a complex system made up of simpler parts. In particular, probability theory provides a principled way of combining the different parts and interfacing models to data. On the other hand, graph theory constitutes a user-friendly tool by which human operators can easily interpret relationships between sets of variables and a natural data structure for the design of efficient general-purpose algorithms. This thesis introduces a novel model called an Affinity Network (AN).

1.1 Graphical models

A graphical model is a graphical representation of a probability distribution. The nodes in the graph correspond to variables and the links between variables denote probabilistic relationships. This thesis only considers directed graphical models, that is, the links are represented by arrows. Each node in the graph is also associated to a conditional probability from which the joint probability of the variables in the dataset is computed.

Bayesian Networks (BNs) (Pearl, 1988) are well-known directed graphical models which usually display causal relationships between variables although they can also be used to represent acausal relationships (Heckerman et al., 1995). The BN graph is constructed using a heuristic hill-climbing approach in which nodes are added or deleted and the direction of the links reversed until an optimum graph is obtained. It should be noted that cycles are not allowed in a BN graph. The joint probability that is encoded by a BN factorises as the product of the conditional probabilities of the nodes.

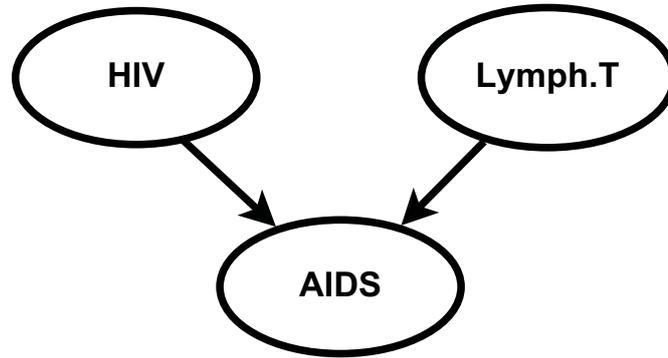
A Dependency Network (DN) (Heckerman et al., 2000) is another instance of a directed graphical model. DNs represent general dependency relationships and were developed in order to address criticism from BNs users who found it hard to interpret some causal and conditional independence relationships that are encoded in a BN graph. This can be illustrated by a medical example as shown in Figure 1.1 where if the variables “HIV” and “Lymph. T” (decrease in level of lymphocyte T) are predictive of the variable “AIDS”, it is also reasonable to say that “AIDS” is predictive of “HIV” and “Lymph. T” which induces a cycle, that is prohibited in a BN, between “AIDS” and “HIV” and also between “AIDS” and “Lymph. T”. Furthermore, the fact that “HIV” and “Lymph. T” are dependent given “AIDS” would not be meaningful given that the decrease in level of lymphocyte T can also be related to another major disease which attacks the immune system. Consequently, the insights that can be gained from the data are hidden by the task of interpreting those causal and conditional independence relationships (Heckerman et al., 2000). A classification or regression algorithm is used to build a DN graph and the joint probability distribution is computed through a stochastic method called pseudo-Gibbs sampling. Unlike a BN, a DN can have loops.

An Affinity Network (AN) is an alternative directed graphical model to a BN and a DN. The AN graph shows similarity relationships between variables in a dataset. In contrast to a BN and a DN, an AN graph is constructed using a message-passing algorithm known as Affinity Propagation (AP) (Frey and Dueck, 2007) and its extension called Soft-Constraint Affinity Propagation (SCAP) (Leone et al., 2007). The graph construction is

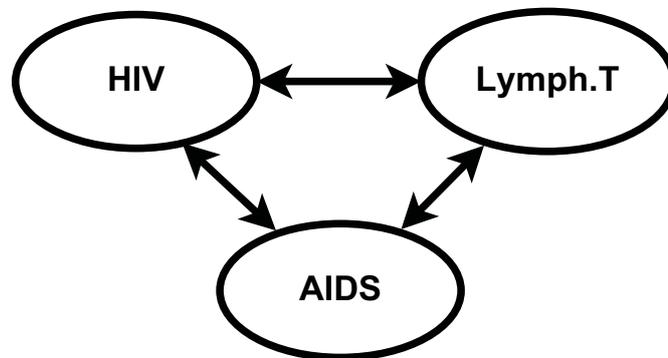
a crucial part in the specification of a graphical model since the conditional probabilities and consequently the joint probability distribution are derived from the graph structure. In contrast to a BN where a heuristic search procedure is used to construct the graph, the SCAP method, from which an AN graph is built, is based on a principled belief propagation algorithm. An AN graph can have loops and a similar approach as in a DN can be used to derive the joint probability distribution of an AN given that the joint distribution does not factorise as in a BN. The AN graphical model can serve as a clinical decision-support system and a tool to characterise the emergent collective behaviour of coupled dynamical systems respectively. A comparative summary of the differences between a BN, a DN and an AN is given in Table 1.1. The properties of the AN will be illustrated by application to sparse medical data and both regular and random dynamical time series data from microelectromechanical systems (MEMS).

	BN	DN	AN
Can approximate conditional and joint probabilities?	Yes	Yes	Yes
Easy-to-interpret network structure (graph)?	No	Yes	Yes
Can the network structure have loops?	No	Yes	Yes
Heuristic network structure search?	Yes	Yes	No
Network structure learning (Chapter 2 and 3)	Hill climbing + BDe score	Probabilistic decision tree + BDe score	SCAP + BDe score
Network parameter learning of conditional probabilities (Chapter 2 and 3)	Dirichlet/Bayesian model	Probabilistic decision tree	Dirichlet/Bayesian model
Joint probability	Product of conditional probabilities	Pseudo-Gibbs sampling	A pseudo-Gibbs sampling can be used

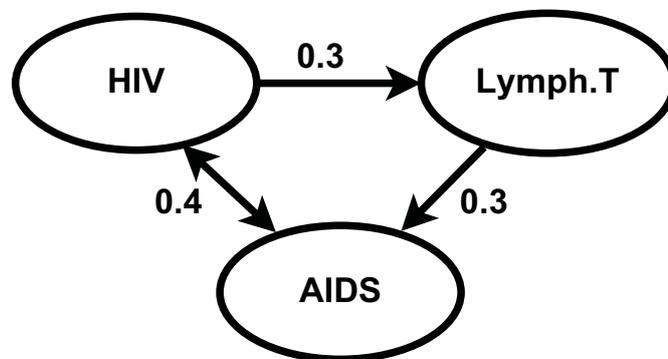
Table 1.1: Comparison between a BN, a DN and an AN.



(a) BN graph.



(b) DN graph.



(c) AN graph.

Figure 1.1: A medical example illustrating the differences between a BN, a DN and an AN. The node “Lymph. T” represents the decrease in level of lymphocyte T. The BN graph tells us that AIDS is caused by HIV and the decrease in level of lymphocyte T. On the other hand, the DN graph is more general in the sense that the three variables depend on each other without assuming any causal relationship as the decrease in level of lymphocyte T can also be dependent on HIV. Finally, the AN graph can give an alternative interpretation to the problem by providing affinity relationships between the variables based on the available dataset. The strengths of the arcs in the AN graph are given by the normalised mutual information between the connected nodes.

1.2 Graphical models and clinical decision-support

A key function in a medical decision-support tool is the ability to highlight to the clinician important relationships between variables. There may be many candidate variables influencing a given outcome making searching for relationships in a very high dimensional space NP-complete. Graphical models have been suggested and developed over the past decade as a possible technique to assist in the discovery of relationships between biomedical variables.

Probabilistic graphical models are the main tools in a clinical decision-support system for two main reasons. First of all, they provide a simple representation of the medical problem to clinicians thanks to their graphical nature. The probabilistic framework used is also a rigorous way of processing the data. A graphical model is then characterised by its graph structure and (conditional) probabilities which are also known as the parameters. Automatic structure and parameter learning and inference are the core functions in a graphical model-based decision-support system.

Most probabilistic clinical decision-support systems make use of a Bayesian Network to model a medical problem and answer probabilistic inference queries. The limitation of the existing systems such as the NasoNet (Galán et al., 2002) originates from the usage of medical expert knowledge to construct its Bayesian Network structure and parameters. As a result, only a few variables can be taken into account in the modelling process. A recent clinical decision-support system called PROMEDAS (Kappen and Neijt, 2002) is an improvement of the NasoNet in the sense that its network structure is automatically generated and hence the system can deal with a variable number of variables. However, the specification of the conditional probabilities requires medical expert knowledge which can still be tedious. It will be shown in Chapter 3 that both the graph and the conditional probabilities of an Affinity Network can be constructed automatically using only the data at hand which is an advantage over current clinical decision systems.

1.2.1 Clinical decision-support

Clinical Decision Support Systems (CDSS), also referred to as medical expert systems, are defined as “active knowledge systems which use two or more items of patient data to generate case-specific advice” (Bemmel and Musen, 1997). The first papers dealing with CDSS date back to the late 1950s and were followed by prototype systems. Then the field of Artificial Intelligence (AI) appeared at the beginning of the 1970s which made it possible for expert systems to deal with uncertain and incomplete medical knowledge.

From the above definition, a CDSS has three main components which are *medical data*, some *patient data* and *patient-specific advice*. The medical knowledge that is required to make decisions is encoded in a certain formalism. Two types of formalism are currently used:

- Well-defined statistical methods are used to capture knowledge from patient data. This type of formalism is also called a *model-based* approach or a *quantitative model*. Bayesian networks and neural networks are examples of model-based approach (Heckerman et al., 1992; Nicholson, 1996; Mani et al., 1997; Herskovits and Dagher, 1997; Chevrolat et al., 1998; Sakellropoulos and Nikiforidis, 1999; Onisko, 2001; Ogunyemi et al., 2002; Galán et al., 2002).
- The second formalism is less formal and uses directly the expert knowledge of the clinician as a reasoning tool. This formalism is also referred to as a *rule-based approach* or a *qualitative model*. Examples of rule-based approaches include decision trees and truth tables (Meditel, 1991; QMR, 1992; DXPLAIN, 1992; ILIAD, 1992).

When the medical knowledge is combined with the patient data, patient-specific advice is generated by the system.

In this work we consider quantitative models to aid decision-support by considering graphical models as the tools to achieve inference. Two instances of graphical models will be discussed in Chapter 2, namely Bayesian networks (BNs) (Neapolitan, 2004) and a more recent class called dependency networks (DNs) (Heckerman et al., 2000) which is very close to BNs as will be seen later. A new class of graphical models, called *Affinity Networks* (ANs), will be introduced in Chapter 3. ANs shares some interesting properties of BNs and DNs but with a more rigorous structure learning procedure.

One of the attractions of Bayesian networks resides in their practical use in real world applications. This is illustrated for instance by the NasoNet system which is a large-scale Bayesian network that performs diagnosis and prognosis of nasopharyngeal cancer (cancer involving the nasal passages) or PROMEDAS, which is a probabilistic medical diagnostic advisory system. These applications are briefly reviewed in the next section.

1.2.2 Examples of CDSS

Two examples of a medical decision-support system are described in this section. A comparison between them is also given along with a brief discussion on how the Affinity Network that will be introduced later can be regarded as an interesting option for a clinical advisory purpose.

NasoNet

The NasoNet system models the different dynamic causal interactions that take place during the spread of nasal-passage cancer (Galán et al., 2002). The degree of uncertainty that is present in the case of nasopharyngeal cancer is higher than other types of cancer because the nasopharynx is a hidden and difficult-to-enter cavity located in the highest part of the pharynx. Time is also a fundamental factor since it usually determines the stage of the disease and, consequently, the type of treatment to be applied. Figure 1.2 shows a part of the Bayesian network used in the NasoNet system.

A representation method that deals with both uncertainty and time is thus given by a temporal Bayesian network called a *network of probabilistic events in discrete time* (NPEDT). Under the NPEDT approach, time is discretised, nodes are associated with events, and each value of a node represents the occurrence of an event at a particular instant. Here an event is a change of state provoked by an anomaly. Diagnosis and prediction consist of fixing the values of the observed nodes and computing the posterior probabilities of some of the unobserved nodes.



Figure 1.2: Part of the Bayesian network modelling the evolution of a cancer of the nasopharynx (Galán et al., 2002).

Another CDSS that was developed in the same year as the NasoNet system and which also uses a Bayesian Network will be described next.

PROMEDAS

PROMEDAS stands for PRObabilistic MEical Diagnostic Advisory System (Kappen and Neijt, 2002). The authors aim to scale-up the system to thousands of variables by automating the generation of probabilistic networks, contrary to the standard procedure where the Bayesian network structure and the probabilities are defined by hand. Because of this, only medium sized networks (50-100 variables) can be constructed in the standard procedure.

The probabilities that are needed to define the local interactions between the nodes quantitatively are given as a conditional probability table whose entries are to be set on the basis of expert knowledge. The domain knowledge acquired from the literature is entered in a knowledge database which is “compiled” into a model represented by a causal probabilistic network. More specifically, a database structure is developed in which medical specialists can enter their knowledge in a way that is familiar for them. The database contains information from which the structure of the network can be derived, and the parameters of the network can be learned. Figure 1.3 shows the Bayesian Network used in the PROMEDAS system.



Figure 1.3: Bayesian network structure of the PROMEDAS system. Three kinds of nodes are handled: prior nodes H_i (risk factors such as occupation, drug use), diagnostic nodes D_j (diseases) and test nodes T_k (tests, symptoms, ...). The structure of the network is that arrows point from cause to effect: the prior nodes point to diagnostic nodes. Diagnostic nodes can point to other diagnostic nodes and test nodes. The arrows are read from a medical database (Kappen and Neijt, 2002).

Besides medical data where strong prior models are usually missing, it will be shown later how the Affinity Network can also be applied to physical data in order to provide a higher-level inference layer on top of a lower physical layer represented by coupled dynamical systems which are modelled by second order differential equations. In particular, it will be shown that the Affinity Network is useful in the study of the emergence of collective behaviour of MEMS. An overview of this area is given in the following section.

1.3 Emergence of collective behaviour in MEMS

Microelectromechanical systems (MEMS) are amongst the smallest functional machines that are currently engineered by humans. The field was considered highly speculative in its early days when Richard Feynman gave his insightful lecture, “There is plenty of room at the bottom” in 1959 (Feynman, 1959). He attempted to spur innovative miniature fabrication techniques for micromechanics, but failed to generate a fundamentally new fabrication technique. However, early successes in the 1970s with the first silicon pressure sensors and accelerometers contributed to the growth of the research activity.

MEMS are useful in building sensors and actuators, together with measurement, control, and signal conditioning circuitry, and equipped with power and communications, all in the tiniest space. MEMS-based sensor systems are particularly important in the automotive area where mechanical microsensors in modern cars are mainly pressure sensors, inertial sensors such as accelerometers and gyroscopes, and sensors for steering angle and torque. Biomedical applications are equally essential, where the combination of high-performance sensing and a high degree of miniaturisation through MEMS techniques offers unique opportunities to improve the quality of life for the ill and the disabled. For instance, microdosing systems combined with microsensors can control the supply of drugs such as insulin to the bloodstream, inertial sensors attached to the legs of people suffering from certain types of brain damage help improve their walking capabilities, to name a few.

The two idealised MEMS elements that will be used in this thesis are the Duffing resonator (Duffing, 1918) and the van der Pol oscillator (van der Pol, 1920). Usually MEMS elements are designed to be explicitly separate and non-interacting and any crosstalk due to capacitive or overlap effects is regarded as a problem to be resolved. However, in this thesis we are more interested in arrays of coupled systems where the interactions between devices may be exploited to positive effect. It is important to be able to characterise the emergent behaviour of such an array of coupled systems in order to design sensors. This is a rather difficult task since the elements of the array are composed of nonlinear dynamical systems as illustrated in Figure 1.4. It will be shown in Chapter 3 that the Affinity Network can provide an interpretation layer on top of the physical layer.

1.3.1 Collective behaviour in MEMS

Real biosensor networks, such as the mammalian olfactory bulb as an exemplar system, and neurally-integrated autonomous communication networks in many nervous systems, share common characteristics and behaviours which are still not understood. Such systems



Figure 1.4: The Affinity Network (AN) provides an interpretation layer on top of the physical layer composed of a square lattice of nonlinear dynamical systems. The red circles correspond to Duffing resonators and the blue ones are van der Pol oscillators. The state of the system can be determined by the clusters of synchronisation that are learned by the AN depending on the response of the sensor array to its external environment.

are composed of interconnected relatively simple dynamical systems, but which are nonlinear, or interact nonlinearly and permit dynamical interactions on multiple time scales. Such systems exhibit a huge diversity of behaviours and produce emergent and macroscopic properties which are responsive in ways they interact with the environment that are unfamiliar to traditional engineering sensor design.

For example Breaa et al. (2009) used biophysical models to explain the gamma-band oscillation in the olfactory bulb based on coupled sub-threshold oscillators which have been shown to exhibit synchronisation and coherence effects when driven with correlated inputs. Similarly, in recent models of temporal coding Burwick (2009), competition for coherence driven by increased phase velocity of neural units in cases of coherent input from the connected neurons leads to different perspectives on assembly formation in collections of coupled oscillator neurons. This assembly formation suggests novel mechanisms for pattern recognition. The ability of a network of coupled nonlinear oscillators to develop emergent behaviour including topological organisation, associative memory and gamma-band synchronisation, has already been used to illustrate a computational capability for abstract pattern recognition (Ursino et al., 2009).

However, a major query is whether we can learn from such biological sensor and communication networks to design more robust biomorphically engineered sensor, communication and computing arrays in recent technologies such as micro and nano-mechanical silicon-based systems.

Motivated by the examples presented above, we will investigate the effects of sensor arrays of coupled microelectromechanical oscillators and resonators. We are interested in the characterisation and detection of novel self-organising response patterns as prototype simultaneous sensing and computing primitives in a pattern recognition system. Whereas traditional design rules would attempt to create an array of isolated and linear-response sensors, the biomorphic design implies interaction and nonlinear oscillators. This allows the emergence of a much richer range of complex behaviours of the array beyond that expected in a traditional design strategy as shown in Figure 1.5.

As a preliminary stage to investigating whether MEMS arrays are capable of exhibiting and thus exploiting similar ranges of behaviour to the biological counterparts, this thesis focusses on the range of emergent properties of coupled oscillator arrays as expected in real microelectromechanical arrays where the simple ‘neuron’ oscillator elements are allowed to interact locally with neighbouring elements. Each neuron element is a simple nonlinear dynamical system characterised by self-sustaining and damped oscillator.

As already noted, in biological systems of coupled oscillatory devices, the interaction



Figure 1.5: An example of a complex behaviour that can occur in an array of 32 by 32 alternating Duffing resonators and van der Pol oscillators. The colours correspond to different phase values of the elements.

permits macroscopic temporal characteristics such as common locking of the phase or amplitude into synchronised states. In order to understand and predict the emergent phenomena of these systems, different mathematical models of chains and arrays of coupled resonators or oscillators have been constructed and studied (Acebron et al., 2005). In a strongly coupled system, the synchronisation effects may be observed globally in all elements (Wang and Slotine, 2004; Lohmiller and Slotine, 1998; Slotine and Lohmiller, 2001; Chiu et al., 1998, 2001; Afraimovich et al., 1997; Afraimovich and Lin, 1998; Hale, 1997) or locally in clusters of neighbouring elements (cluster synchronisation) (Kaneko, 1990; Belykh and Mosekilde, 1996; Xie and Hu, 1997; Hasler et al., 1990; Belykh et al., 2000; Belykh and Mosekilde, 2001).

Recent progress in micro/nano technology now enables scientists to design and manufacture coupled mechanical resonators to experimentally investigate the collective behaviour of such systems. For example, intrinsic localised modes were observed in a micro-mechanical oscillator array of about 220 coupled cantilevers (Sato et al., 2003). The formation and propagation range of the acoustic waves over a two-dimensional array of 400 coupled nano mechanical resonators were studied as a potential application in RF signal processing (Zalalutdinov et al., 2006b). It could be also expected that the collective behaviours of coupled oscillators can be utilised in novel sensing devices (Truitt et al., 2007), optomechanical signal processing devices (Buks and Roukes, 2002) or artificial intelligence decision-making (Rand and Wong, 2008a).

To extend these recent activities and also to link to recent mathematical work on coupled oscillator array systems, in this thesis the collective behaviour of alternating Duffing

resonators and van der Pol oscillators elastically coupled in a square lattice is numerically studied with typical MEMS parameters in Chapter 4, followed by an investigation of the behaviour of the lattice in the presence of random noise in Chapter 5. The stationary patterns are then analysed using the Affinity Network to provide a link between the microscopic many-element state description, to a macroscopic inference of response in terms of clusters.

1.4 Thesis contribution

This thesis proposes a new graphical model structure: the AN. Its novelty is in the construction of a network structure using a principled message-passing algorithm that uses similarities between variables in a dataset: this has not been used before as a search process for graphical models. This is a non-heuristic approach compared to existing methods to build a graphical model network structure where nodes are selected randomly then added or deleted from an initial network structure. Its diverse applicability is revealed in this thesis by illustrating its properties in two very different domains: graphical model representations of problems that have no strong prior models, and spatiotemporal patterns in coupled dynamical systems in which we regard the AN as providing a higher-level inference or logic layer above the lower physical ‘sensing’ layer. This thesis therefore provides a preliminary investigation into the use of a ‘similarity seeking’ graphical model (the AN), to achieve the pattern processing equivalent of ‘interpretation’.

The dynamical-system analysis in Chapter 4 and Appendix B are new results and have not appeared in the literature before and lead to the predictions that coupled van der Pol and Duffing systems will show multiple hysteresis effects, locking and synchronisation as discussed in Chapter 4. These results have now been accepted in the following journal publications:

- X. Wei, M. F. Randrianandrasana, M. Ward and D. Lowe. Nonlinear dynamics of a periodically driven Duffing resonator coupled to a van der Pol oscillator. *Mathematical Problems in Engineering*, 2010. In press.
- M. F. Randrianandrasana, X. Wei and D. Lowe. A preliminary study into emergent behaviours in a lattice of interacting nonlinear resonators and oscillators. *Communications in Nonlinear Science and Numerical Simulations*, 2010. In press.

Additionally, work in Chapter 3 and 4 has been published in the following peer-reviewed international conferences:

- M. F. Randrianandrasana and D. Lowe. Bayesian approach to learning Dependency Networks in biomedical domains. *Proceedings of the 3rd International Conference on Computational Intelligence in Medicine and Healthcare*, Plymouth, 2007.
- M. F. Randrianandrasana, X. Wei and D. Lowe. Collective Behaviour in a Square Lattice of Driven Duffing Resonators Coupled to van der Pol Oscillators”. *Proceedings of the 10th IEEE International Conference on Computer and Information Technology*, p785-790, 2010.

1.5 Notation

The following notation will be used throughout this thesis: a variable is denoted by an upper case letter and a state or value of that variable by the same lower case letter. A set of variables is denoted by a bold-face capital letter(s) and an assignment of state or value to each variable in a given set is denoted by the corresponding bold-face lower-case letter(s). And finally, calligraphic letters will be used to denote statistical models. The main symbols are summarised in Table 1.2.

Symbol	Meaning
D	A dataset
\mathcal{G}	A graph
M	Number of samples (rows) in the data D
$\mathbf{x}^{(h)}$	h th case of the data D
\mathbf{X}	A set of random variables
\mathbf{x}	The set of realisations of the variables in \mathbf{X}
n	Number of random variables
X_i	A random variable
x_i	A realisation of the variable X_i
r_i	Number of possible states of X_i
\mathbf{PA}_i	The set of parents of X_i
\mathbf{pa}_i	A realisation of the set of parents of X_i
q_i	Number of possible configurations of \mathbf{PA}_i
pa_{ij}	The j th possible configuration of \mathbf{PA}_i
$\boldsymbol{\mu}^T$	The transpose of the vector $\boldsymbol{\mu}$

Table 1.2: Notation.

1.6 Thesis structure

The structure of this thesis is as follows:

Chapter 2 describes state-of-the-art graphical model instances, namely the Bayesian Network and the Dependency Network. It shows how their network structures and parameters

are learned from a dataset.

Chapter 3 introduces the Affinity Network graphical model, compares it with the two previous models on various medical and general data with different sizes and shows that the Affinity Network is generally more accurate and faster to construct.

Chapter 4 provides analysis of the coupled driven Duffing resonator and van der Pol oscillator as a single unit to show that they exhibit interesting and complex behaviour. The unit is then used as a building block in a square lattice of such elements and the effects of system parameters and lattice topology are numerically studied. It is shown that, as expected, more complicated behaviour can occur in such a lattice.

Chapter 5 extends the regular square lattice of coupled dynamical systems by introducing random connections with various strengths and the effect on synchronisation in the lattice is investigated. The Affinity Network is then applied to this probabilistic network of dynamical systems and its satisfactory macroscopic inference capability is shown.

Chapter 6 uses the Affinity Network to analyse the spatiotemporal patterns in the square lattice of coupled dynamical systems and shows that the resulting clusters of synchronisation are sensible compared to the stationary patterns that are observed in the lattice.

Chapter 7 concludes the thesis and provides directions for future research.

Appendix A derives the dimensionless equation of the coupled Duffing resonator and van der Pol oscillator.

Appendix B provides a detailed explanation of the main results from the multiple time scale analysis of the coupled system.

Bayesian and Dependency 2 Networks

In this chapter we provide the background and framework for describing inference problems through graphical models. We specifically highlight the distinction between Bayesian and Dependency Network approaches. A brief description of the necessary concepts from graph theory and probability theory that are needed to understand this chapter is given in Section 2.1 followed by definitions of a Bayesian Network and a Dependency Network in Section 2.2. The remaining two sections will be concerned with how to construct a Bayesian network and a Dependency Network from a dataset and how to use them to answer probabilistic queries respectively.

2.1 Graphical and probabilistic backgrounds

A graph \mathcal{G} is composed of a set of nodes V (also called vertices) and a set of edges E (also called arcs) that connect the nodes, and the graph is usually represented using the notation $\mathcal{G} = (V, E)$. An edge can be directed, which means that it has an arrow on one of its ends, or undirected and consequently the graph will be called a *directed* or *undirected* graph. For example $X \rightarrow Y$ is a simple directed graph which has two nodes X and Y and a directed edge connecting them. In this example, there is an edge between X and Y ,

consequently X and Y are *adjacent*. When a directed graph contains no directed cycles, it is also called a directed acyclic graph (DAG). X is called the *parent* of Y , which is reciprocally called the *child* of X , since the direction of the edge goes from X to Y . If we have a set of nodes X_1, \dots, X_k with $k \geq 2$ such that $(X_{i-1}, X_i) \in E$ for $2 \leq i \leq k$, then the set of edges which connect the k nodes is called a *path* from X_1 to X_k . A node Z is called a *descendent* of a node X and X is called an *ancestor* of Z if there is a path from X to Z . When the directionality of every edge in a DAG is ignored, the resulting undirected graph is called the *skeleton* of the DAG. A *v-structure* in a DAG \mathcal{G} is an ordered triple of nodes (X, Y, Z) such that \mathcal{G} contains the edges $X \rightarrow Y$ and $Z \rightarrow Y$, and X and Z are not adjacent in \mathcal{G} .

The notions of independence and conditional independence are at the core of graphical models and hence it is important to define them clearly. Two random variables X and Y are independent if and only if $P(X, Y) = P(X)P(Y)$ and the statement that they are independent given another random variable Z is equivalent to $P(X, Y|Z) = P(X|Z)P(Y|Z)$ or $P(X|Y, Z) = P(X|Z)$. Another important property that follows from the previous definitions is the *Markov condition* which is a constraint that a variable is independent of its non-descendants given its parents (Pearl, 1988). This property constitutes the foundation of Bayesian networks (BNs) as the computation of the joint distribution requires this constraint which is encoded in the structure of the graph. The notion of *Markov equivalence*, which is related to the previous definition, is also useful to understand for the remaining of this chapter: two DAGs are Markov equivalent if and only if they have the same links, regardless of their directions, and the same set of uncoupled head-to-head (for example, $X \rightarrow Z \leftarrow Y$) meetings (Pearl et al., 1990). In particular, two DAGs are Markov equivalent if and only, based on the Markov condition, they entail the same conditional independencies (Neapolitan, 2004). A *DAG pattern* for a Markov equivalence class is a graph that has the same links as the DAGs in the equivalence class and has oriented all and only the edges common to all of the DAGs in the equivalence class as shown in Figure 2.1. The equivalence property has later been generalised by Pearl and Verma (1991b) as follows: two DAGs are equivalent if and only if they have the same skeleton and the same v-structures.

The relevance of Markov equivalence to BN structure learning is that DAG patterns should be learned instead of DAGs given that DAGs belonging to a Markov equivalence class represent the same statistical dependence. Therefore, the structure learning algorithm can be speeded up since time can be saved by avoiding rescoring DAGs in the same Markov equivalence class (Anderson et al., 1995). This will be discussed further in Section

2.3.2 after defining a BN and a DN in the following section.



(b)

Figure 2.1: Markov equivalence class and DAG pattern. The DAG pattern in (2.1(b)) represents the Markov equivalence class in (2.1(a)) (Neapolitan, 2004).

2.2 Bayesian and Dependency Networks definitions

Bayesian networks (BNs) (Pearl, 1988; Neapolitan, 2004) and Dependency Networks (DNs) (Heckerman et al., 2000) are both graphical representations of probability distributions. They are especially useful in the presence of uncertainty, which is a common feature in the field of artificial intelligence, due to their probabilistic nature. BNs usually display causal relationships between variables although they can also be used to represent acausal relationships (Heckerman et al., 1995). DNs were developed in order to address criticism from BNs users who found it hard to interpret some causal and conditional independence relationships that are encoded in a BN structure. This can be illustrated for example by the situation where if the variables “Age” and “Gender” are predictive of the variable “Income”, it is also reasonable to say that “Income” is predictive of “Age” and “Gender” which induces a loop, that is prohibited in a BN, between “Income” and “Age” and also between between “Income” and “Gender”. Furthermore, the fact that “Age” and

“Gender” are dependent given “Income” would not be meaningful. Consequently, the insights that can be gained from the data are hidden by the task of interpreting those causal and conditional independence relationships (Heckerman et al., 2000).

The two main components of a BN and a DN are a network *structure* and *parameters*. A DAG represents the structure of a BN whereas the graph structure of a DN is likely to have cycles. The edges of both BN and DN graphs are used to specify dependence properties between variables. Both graphs contain a node for each variable X_i . The parameters of a BN and a DN are the conditional probability distributions of the variables given their parents. The learning procedures of these two components will be described in the following section.

2.3 Learning Bayesian and Dependency Networks

Learning a Bayesian or a Dependency Network, and more generally a graphical model, consists of searching for the best network structure and estimating the model parameters. The learning procedure is usually performed from the data in hand, especially for large datasets where an expert knowledge of the domain is not always available. A comparative description between the parameter and structure learning in a BN and a DN will be presented in this section.

2.3.1 Parameter learning

As mentioned in Section 2.2, the parameters of a BN and a DN are represented by the conditional probabilities of each variable given their parents. The Bayesian approach to parameter learning in a BN puts a prior subjective probability distribution on the value of a random variable, updates that distribution based on the data and finally the parameters are computed using the properties of the corresponding probability distribution. In contrast, the conditional probabilities in a DN are estimated directly from the data using classification or regression algorithms.

Parameter learning for BNs

The Dirichlet and normal distributions are respectively used to learn the parameters of a BN with discrete and continuous variables due to their properties for multinomial (discrete) and normally (continuous) distributed data. These two different cases will be described next.

Discrete variables

Parameter learning in a BN with discrete variables is described as follows: the parameters are learned by modelling each variable with a Dirichlet distribution which is a *conjugate* distribution for the multinomial likelihood which means that the posterior distribution follows the same parametric form as the prior distribution (Gelman et al., 2004). As a result, the computation of the posterior distribution (model update) is simplified as well as the expression of the conditional distribution which is given by (Heckerman et al., 1995; Neapolitan, 2004)

$$P(X_i = k | \mathbf{pa}_{ij}) = \frac{a_{ijk}}{N_{ij}}, \quad (2.1)$$

where \mathbf{pa}_{ij} denotes the j th possible instantiation of the set of parents \mathbf{PA}_i of X_i , a_{ijk} are the Dirichlet parameters and $N_{ij} = \sum_k a_{ijk}$. The a_{ijk} are usually initialised to 1 to express the absence of knowledge concerning the value of the relative frequency or if one wants to be objective and let the data “speak” for itself. Consequently, the different states that a variable can have are equiprobable.

Continuous variables

In order to understand how the parameters are learned for continuous variables, a definition of a Gaussian Bayesian Network is first given, in which the value of each variable is a linear function of the values of its parents, as

$$x = w_X + \sum_{Z \in PA_X} b_{XZ} z,$$

where PA_X is the set of all parents of X , w_X is a value that the random variable W_X can have and the coefficient b_{XZ} in the linear combination represents the strength of the connection or edge between X and Z . The variable W_X represents the uncertainty in X 's value given the values of X 's parents; it has density function $\mathcal{N}(w; 0, \sigma_{W_X}^2)$ and is independent of each Z . $\sigma_{W_X}^2$ is the variance of X conditional on the values of its parents. For each root X , its unconditional density function $\mathcal{N}(x; \mu_X, \sigma_X^2)$ is specified from prior knowledge. For example, a density function equal to $\mathcal{N}(x; \mu_X, 0)$ means the root's value is known, while a density function equal to $\mathcal{N}(x; \mu_X, \infty)$ means complete uncertainty as to the root's value. Since $\sigma_{W_X}^2$ is the variance of X conditional on values of its parents,

the conditional density function of X is

$$\rho(x|pa_X) = \mathcal{N}(x, \sum_{Z \in PA_X} b_{XZ}z, \sigma_{W_X}^2). \quad (2.2)$$

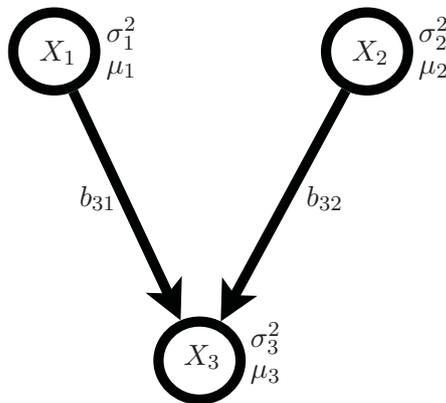


Figure 2.2: An example of a Gaussian Bayesian Network. Each continuous variable X_i , $i = 1, \dots, 3$, is characterised by its mean μ_i and its variance σ_i^2 . The strength of the link between two connected nodes is given by the scalar b_{ij} , where the index i denotes a child node X_i and j its parent X_j .

Figure 2.2 illustrates a Gaussian Bayesian network where each continuous variable X_i , $i = 1, \dots, 3$, is characterised by its mean μ_i and its variance σ_i^2 . The strength of the link between two connected nodes is given by the scalar b_{ij} , where the index i denotes a child node X_i and j its parent X_j .

The following variables will be used throughout the learning algorithm:

$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)^T$ and v are the mean vector in the hypothetical sample and the size of the hypothetical sample upon which the prior belief concerning the value of the unknown mean vector is based, respectively, $\boldsymbol{\beta}$ is the value of the sample covariance matrix $\boldsymbol{\psi}$ in the hypothetical sample, α represents the degrees of freedom and is given by $\alpha = v - 1$, n is the number of variables and M is the sample size of the data. The steps involved in the parameter learning algorithm are (Neapolitan, 2004; Groot, 1970):

1. the network structure is first constructed by initialising, for $1 \leq i \leq n$, μ_i , σ_i^2 , and b_{ij} as the coefficient in the linear combination,
2. the network structure is converted to the corresponding multivariate normal distribution $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{T}^{-1})$ using Algorithm 2.3.1 that creates the precision matrix \mathbf{T} , which is the inverse of the covariance matrix (Shachter and Kenley, 1989),
3. the prior values for α and v are assessed as $\alpha = v - 1$ and v is given by the size

of the hypothetical sample upon which the prior belief concerning the value of the unknown mean is based,

4. the prior value for $\boldsymbol{\beta}$ is assessed as

$$\boldsymbol{\beta} = \frac{v(\alpha - n + 1)}{(v + 1)} \mathbf{T}^{-1}, \quad (2.3)$$

5. the values of $\boldsymbol{\beta}$, α , $\boldsymbol{\mu}$ and v are updated as

$$\boldsymbol{\beta}^* = \boldsymbol{\beta} + \boldsymbol{\psi} + \frac{vM}{v + M} (\bar{\mathbf{x}} - \boldsymbol{\mu})(\bar{\mathbf{x}} - \boldsymbol{\mu})^T, \quad (2.4)$$

$$\alpha^* = \alpha + M, \quad (2.5)$$

$$\boldsymbol{\mu}^* = \frac{v\boldsymbol{\mu} + M\bar{\mathbf{x}}}{v + M}, \quad (2.6)$$

$$v^* = v + M, \quad (2.7)$$

where $\bar{\mathbf{x}}$ denotes the sample mean vector,

6. the precision matrix \mathbf{T} is updated as

$$(\mathbf{T}^*)^{-1} = \frac{(v^* + 1)}{v^*(\alpha^* - n + 1)} \boldsymbol{\beta}^*, \quad (2.8)$$

7. the variables are ordered as parent nodes coming first; then $\mathcal{N}(\mathbf{x}^{(M+1)}; \boldsymbol{\mu}^*, (\mathbf{T}^*)^{-1})$ is converted to a Gaussian network yielding updated values σ_i^{*2} and b_{ij}^* ,
8. and finally the distribution of $\mathbf{x}^{(M+1)}$ by the Gaussian network containing the graph and the parameter values μ_i^* , σ_i^{*2} , and b_{ij}^* is estimated.

Algorithm 2.3.1: Precision matrix \mathbf{T} construction.

```

 $\mathbf{T}_1 = \frac{1}{\sigma_1^2};$ 
for  $i = 2; i \leq n; i++$  do
     $\mathbf{b}_i = \begin{pmatrix} b_{i1} \\ \vdots \\ b_{i(i-1)} \end{pmatrix};$ 
     $t_i = \frac{1}{\sigma_i^2};$ 
     $\mathbf{T}_i = \begin{pmatrix} \mathbf{T}_{i-1} + t_i \mathbf{b}_i \mathbf{b}_i^T & -t_i \mathbf{b}_i \\ -t_i \mathbf{b}_i^T & t_i \end{pmatrix};$ 
end
 $\mathbf{T} = \mathbf{T}_n;$ 

```

Parameter learning for DNs

The parameters in a DN are represented by the local conditional probabilities given by:

$$p(x_i|\mathbf{pa}_i) = p(x_i|\mathbf{x}\setminus x_i), \quad (2.9)$$

where \mathbf{pa}_i represents an instance of the set of parents of the variable X_i . Classification or regression algorithms (for example, probabilistic decision trees (Buntine, 1991), generalised linear models (McCullagh and Nelder, 1989), neural networks (Bishop, 1995), probabilistic support vector machines (Platt, 1999)) are used in order to model those probabilities and a different model for each local distribution can be applied. The probabilistic decision tree approach will be described briefly as decision trees satisfy the near consistency assumptions discussed in Section 2.3.2. It is recalled that near consistency means strong inconsistencies are rare when the dataset contains many samples because each local distribution is learned from the same dataset, which is assumed to be generated from a single underlying joint distribution. In the probabilistic decision tree approach, a child variable X_i is the target variable and its parents $\mathbf{X}\setminus X_i$ are the input variables in the probabilistic decision tree method and each leaf (node that has no child) of the tree is modelled as a multinomial distribution. A hill-climbing approach combined with a BD score is used to learn the decision tree structure as follows: for a variable X_i , the search algorithm is initialised with a singleton root node that has no children; then each leaf node is replaced with a binary split on some variable X_j in the set of parents $\mathbf{X}\setminus X_i$ until the score of the tree ceases to increase. The binary split on X_j is a decision tree node having two children: one child corresponds to a particular value of X_j and the other child corresponds to all other values of X_j . A uniform prior distribution for the parameters of all multinomial distributions is used by the BD score function. Its structure prior is proportional to κ^f , where $\kappa > 0$ is a tunable parameter and f is the number of free parameters in the decision tree. A value of $\kappa = 0.1$ is used in Heckerman et al. (2000) because it has been found to give accurate predictions over a wide variety of data.

2.3.2 Structure learning

The methods for learning the structure of BNs can be grouped into two main categories: algorithms based on conditional independence tests (Cheng et al., 2002; de Campos, 1998; de Campos and Huete, 2000; Meek, 1995; Pearl and Verma, 1991a; Spirtes et al., 1993; Verma and Pearl, 1990; Wermuth and Lauritzen, 1983), which are also called *constraint-based* algorithms, and algorithms that use a scoring function and a search method (Chow

and Liu, 1968; Cooper and Herskovits, 1990; Bouckaert, 1993, 1995; Friedman and Goldszmidt, 1998; Lam and Bacchus, 1994; Suzuki, 1993; Cooper and Herskovits, 1992; Heckerman et al., 1995; Kayaalp and Cooper, 2002). A combination of independence-based and scoring-based methods can also be found in the literature (Acid and de Campos, 2000, 2001; Dash and Druzdzel, 1999; de Campos et al., 2003; Singh and Valtorta, 1995; Spirtes and Meek, 1995). On the other hand, the structure of a DN is governed by a classification/regression process.

Structure learning for BNs

In this section, the focus will be put on the techniques that use a scoring metric and a search procedure where the task of learning a network structure can be summarised as follows: given a finite set $\mathbf{U} = \{X_1, \dots, X_n\}$ of n random variables and a dataset $D = \{\mathbf{u}_1, \dots, \mathbf{u}_M\}$ of instances of \mathbf{U} (M samples), find a DAG \mathcal{G} such that

$$\tilde{\mathcal{G}} = \operatorname{argmax}_{\mathcal{G} \in \mathcal{G}_S} s(\mathcal{G}, D),$$

where $s(\mathcal{G}, D)$ is a scoring function that measures the goodness of fit of the DAG \mathcal{G} with respect to the data D and \mathcal{G}_S is the search space which is composed of all the DAGs that contain all the variables in the data.

Scoring functions

Most scoring functions can also be grouped into two categories: scoring functions based on a Bayesian approach, such as K2 (Cooper and Herskovits, 1992), Bayesian Dirichlet (BD) (Heckerman et al., 1995), and those that are based on information theory such as the Minimal Description Length (MDL) (Rissanen, 1986; Lam and Bacchus, 1994) and the Akaike Information Criterion (AIC) (Akaike, 1974). Attention will be focused on the BD scoring function in this chapter. The likelihood equivalence property needs to be specified in order to derive the BD score. Under the likelihood equivalence assumption, two network structures \mathcal{G}_1 and \mathcal{G}_2 are equivalent if $p(D|\mathcal{G}_1) = p(D|\mathcal{G}_2)$, in other words, the probability of the data D is the same given hypotheses corresponding to any two equivalent network structures. This also implies that the data cannot distinguish between equivalent network structures.

The likelihood of a network structure \mathcal{G} given the data D can be obtained using Bayes'

rule as follows:

$$p(\mathcal{G}|D) \propto p(D|\mathcal{G}) \cdot p(\mathcal{G}). \quad (2.10)$$

The possible network structures are generally equiprobable so that the prior $p(\mathcal{G})$ can be neglected and hence the computation of the data likelihood $p(D|\mathcal{G})$ is enough to select the best network. Its expression is described next depending on the type of the variables.

Discrete variables

The Bayesian score is a scoring function which represents the goodness of fit of a network with respect to the data and has been introduced by Heckerman et al. (1995). It is proportional to the posterior probability of a network structure \mathcal{G} given the data D in Equation (2.10). Learning consists then in searching for the network that maximises this probability.

The BD score is built under different assumptions:

1. the variables must be discrete with a finite number of states,
2. the parameter sets $\theta_{ij} = x_i | \mathbf{pa}_{ij}$, where \mathbf{pa}_{ij} denotes the j th possible instantiation of the parents \mathbf{PA}_i of X_i , are mutually independent:

$$p(\theta_G|G) = \prod_{i=1}^n \prod_{j=1}^{q_i} p(\theta_{ij}|G), \quad (2.11)$$

where n denotes the number of variables and q_i represents the number of different instantiations of the parents \mathbf{PA}_i of the variable X_i ;

3. each parameter set θ_{ij} has a Dirichlet distribution:

$$p(\theta_{ij}|G) = \text{Dir}(\theta_{ij} | a_{ij1}, \dots, a_{ijr_i}), \quad (2.12)$$

where r_i is the number of possible states of the variable X_i ;

4. the data must be complete, that is every variable is observed in every case of the data D .

Under the previous assumptions the BD score reduces to the marginal likelihood (the proof can be found in Heckerman et al. (1994)):

$$p(D|G) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(N_{ij})}{\Gamma(N_{ij} + M_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(a_{ijk} + z_{ijk})}{\Gamma(a_{ijk})}, \quad (2.13)$$

where a_{ijk} are the Dirichlet parameters for the variable X_i being in the state k and having the parents \mathbf{pa}_{ij} , $N_{ij} = \sum_k a_{ijk}$, z_{ijk} is the number of times in the data where the variable X_i is in the state k and its set of parents is \mathbf{pa}_{ij} , $M_{ij} = \sum_k z_{ijk}$. Γ is the Gamma function. The parameters a_{ijk} specify one's current knowledge about the domain for purposes of learning network structures. But in practice, the specification of a_{ijk} for all possible variable-parent configurations and for all values of i , j and k is not an easy task. The K2 score (Cooper and Herskovits, 1992) is obtained when $a_{ijk} = 1$. When the likelihood equivalence assumption is considered, then the hyperparameters can be computed as follows (Heckerman et al., 1995):

$$a_{ijk} = N_e \times p(\cdot | \mathcal{G}_0), \quad (2.14)$$

where $p(\cdot | \mathcal{G}_0)$ represents a probability distribution associated with a prior BN \mathcal{G}_0 and $N_e = \sum_{j=1}^{q_i} N_{ij}$ is called the *equivalent sample size* of the network. This constitutes the BDe score where “e” denotes likelihood equivalence. A particular case of the BDe metric, suggested by Buntine (91), is obtained using another uninformative assignment: $p(\cdot | \mathcal{G}_0) = \frac{1}{r_i \cdot q_i}$, so that the probability distribution in the resultant BN is uniform. The metric is therefore abbreviated as BDeu (“u” for uniform).

Continuous variables

The conditional density function $\rho(D|G)$ of the data given a graph is maximised in order to learn the network structure with continuous variables and is given by (Neapolitan, 2004):

$$\rho(D|G) = \left(\frac{1}{2\pi}\right)^{\frac{Mn}{2}} \left(\frac{v}{v+M}\right)^{\frac{n}{2}} \left(\frac{c(n, \alpha)}{c(n, \alpha+M)}\right) \left(\frac{|\beta|^{\frac{\alpha}{2}}}{|\beta^*|^{\frac{\alpha+M}{2}}}\right), \quad (2.15)$$

where

$$c(n, \alpha) = \left[2^{\alpha n/2} \pi^{n(n-1)/4} \prod_{i=1}^n \Gamma\left(\frac{\alpha+1-i}{2}\right) \right]^{-1}.$$

The steps 1-5 in Section 2.3.1 are used to determine the different parameter values in Equation (2.15).

Once a scoring function has been specified, it is then combined with a search procedure in order to learn a BN structure.

Network structure search methods

The search for a network structure (DAG or DAG pattern) that best fits some data is called model selection where a model in this case means a candidate DAG or DAG pattern. This task is usually done using a heuristic search algorithm which is characterised by a search space and a set of operations. The search space is usually composed of all the DAGs that contain all the variables in the data which is the simplest search space. The operations typically consist of adding or removing edges or reversing the direction of an edge.

One of the first popular heuristic search algorithms was called K2 and was developed by Cooper and Herskovits (1992) as an evolution of an expert system named Kutató (Cooper and Herskovits, 1990) from which the name of the algorithm is derived. The search space of the K2 algorithm is given by the set of all DAGs which contain all the variables and the addition of a parent to a node is its unique operation. In addition, an ordering of the variables is required as a first step of the algorithm. This ordering can be obtained from domain knowledge where a time ordering can be used in which the variables that are suspected to be the cause of a certain situation precede those that are suspected to be the effects of the latter variables. In practice, this prior ordering phase can be very tedious and hence undesired especially when the number of variables in a dataset is huge. Greedy algorithms that do not need prior ordering have therefore been devised where the search space is still the set of all DAGs containing the variables but where removal and reversal operations are combined with edge additions. Among those algorithms are the standard *greedy hill-climbing* (Friedman and Goldszmidt, 1998) in which the search is initialised with an empty network, for example, then edges addition and removal that lead to the largest improvement in the score are repeatedly applied to the “current” candidate network. This “upward” step is repeated until a local maxima is reached, meaning that no modification of the current candidate improves the score.

A common problem with heuristic search algorithms is the fact of being stuck in local minima. Different strategies have been proposed to avoid this problem among which are iterated hill-climbing, in which a random perturbation is applied to the current structure then the search procedure is repeated and in the end the maximum over the local maxima is taken, simulated annealing (Metropolis et al., 1953), best-first search (Korf, 1993) and Gibbs sampling (Geman and Geman, 1983; Neal, 1993). However those techniques that avoid local minima only improve the search but do not guarantee that an optimal network structure will be found. It should be noted also that those methods belong to the class of algorithms that search over DAGs. In an attempt to reach an optimal solution, algorithms that search DAG patterns have been developed (Chickering, 2002b) and have

been improved recently (Chickering, 2002a). These algorithms are based on the greedy equivalent search property, also called Meek’s conjecture (Meek, 1997). This conjecture says that if a perfect map of the generative distribution P exists, that is, there is a DAG that encodes the distribution P , the limit of the probability of finding a perfect map of P as the size of the dataset approaches infinity is equal to 1. As mentioned earlier in Section 2.1, algorithms that search for DAGs are less efficient than algorithms that search for DAG patterns since time can be wasted when they encounter and rescore DAGs in the same Markov equivalence class (Anderson et al., 1995). Nevertheless, algorithms that search for DAG patterns will also fail to find an optimal solution if the sample size is not large enough or it may be large but an unlikely event occurs or if the generative distribution of the data does not have a faithful DAG representation.

Another class of search algorithms is called approximate model averaging (Heckerman et al., 1999) where an average over all structures is computed. When the number of possible structures is large, high-probability structures are heuristically searched for instead. This is usually done using the Markov Chain Monte Carlo method. Approximate model averaging has been reported to work well in situations where the number of variables is small and the amount of data is big or when the amount of data is small relative to the number of variables (Friedman et al., 2000) or when learning relationships among some of the variables, also called partial structure learning, is more of interest than using the network structure for inference and decision making.

Structure learning for DNs

In a general DN, the network structure is such that each variable has the remaining variables as parents, which can be denoted formally as $\mathbf{PA}_i = \mathbf{X}/X_i$. If the DN is constructed from data using a classification/regression method, the resulting DN is highly likely to be *inconsistent*. The inconsistency can be both structural and numerical. For instance, if the data is $\mathbf{X} = X_1, X_2$, it can happen that X_1 helps to predict X_2 but X_2 does not help to predict X_1 when the estimator of $p(x_1|x_2)$ discards X_2 as an input (conditioning variable) whereas the estimator of $p(x_2|x_1)$ keeps X_1 as an input, which is inconsistent in terms of the structure. On the other hand, numerical inconsistency stems from the absence of a joint distribution $p(\mathbf{x})$ from which each of the local distributions $p(x_1)$ and $p(x_2|x_1)$ can be obtained from the rules of probability. However, the authors argue that when the dataset contains many samples, strong inconsistencies will be rare because each local distribution is learned from the same dataset, which is assumed to be generated from a single underlying joint distribution. The resulting DN will be “almost”

consistent. Another necessary assumption to get an “almost” consistent DN is that the model classes (classification or regression algorithms) which are used to estimate the local distributions can closely approximate the conditional distributions consistent with the underlying joint distribution. The authors use probabilistic decision trees (Buntine, 1991) since they satisfy the previous assumption for variables with finite domains.

Once the structure and parameters of a graphical model have been estimated from a dataset containing the set of variables \mathbf{X} , it can then be used to answer probabilistic queries of the form $p(\mathbf{y}|\mathbf{z})$, where the *target* variables \mathbf{Y} and the *input* variables \mathbf{Z} are disjoint subsets of \mathbf{X} . This task is called *probabilistic inference* and its important special case is the determination of the joint probability distribution $p(\mathbf{x})$. The estimation of these two types of distributions in a BN and a DN will be described next.

2.4 Probabilistic inference

In a BN, the expression of the joint distribution relies on an assumption which allows its factorisation using the conditional probabilities of the variables. The computation of the marginal probabilities can then be derived from this factorisation through a message-passing algorithm. Conversely, the probabilistic inference task in a DN is based upon the independencies encoded in the network structure for a consistent DN and on a stochastic sampling algorithm for a general DN.

2.4.1 Inference in BNs

The key property in a Bayesian Network is that a variable is independent of its non-descendants given the values of its parents in the graph, which is referred to as the *Markov condition* (Pearl, 1988). This implies that any distribution p that satisfies the independence assumption encoded in the graph can be factored as

$$p(X_1, \dots, X_n) = \prod_{i=1}^n p(X_i | \mathbf{PA}_i), \quad (2.16)$$

where \mathbf{PA}_i is the set of parents of the variable X_i . Inference queries using marginal probabilities, which is the most typical case in practice, can be answered through Pearl’s belief propagation algorithm which is a message-passing-based method.

Pearl's belief propagation algorithm

Belief propagation (Pearl, 1988) is an inference algorithm where random variables are represented by nodes in a graph and the relationship between these variables are symbolised by directed edges. Belief propagation is valid for directed acyclic network structures where the algorithm gives exact inference results. However, for networks having cycles, the algorithm, which is then referred to as loopy belief propagation (Weiss, 1997), can also be applied. In that case, when the algorithm converges, good approximate inference results can be obtained (Murphy et al., 1999).

In this section, we shall refer to the following network: $U \rightarrow X \rightarrow Y$ where the node X is the node under consideration and whose sets of parents and children are denoted by U and Y respectively. Each directed link $X \rightarrow Y$ is associated with a *fixed* conditional probability matrix in which the (x, y) entry is given by $P(Y = y|X = x)$. Belief propagation computes the *belief*, which is a *dynamic* probability value, in variable X being in the state x by fusing two types of support information: messages λ , also called evidential or diagnostic support, from X 's children, and π messages, also called causal or predictive support, from X 's parents.

The belief update equation is derived by partitioning the total evidence e , that is, the set of all instantiated variables, into disjoint subsets: e_X^- and e_X^+ which represent evidence connected to X through its set of children Y and evidence connected to X through its set of parents U respectively, as shown in Figure 2.3.

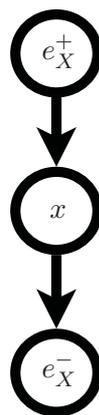


Figure 2.3: Evidence nodes in a BN. e_X^- and e_X^+ represent evidence connected to X through its set of children and parents respectively.

We then have:

$$\begin{aligned} \text{BEL}(x) &= P(x|e), \\ &= \alpha P(e|x)P(x), \text{ using Bayes' rule, where } \alpha = \frac{1}{P(e)} \end{aligned} \quad (2.17)$$

$$= \alpha P(e_X^-|x)P(x|e_X^+), \text{ from Figure 2.3} \quad (2.18)$$

$$= \alpha \lambda(x)\pi(x), \quad (2.19)$$

where $\lambda(x)$ and $\pi(x)$ are defined as $\lambda(x) = P(e_X^-|x)$ and $\pi(x) = P(x|e_X^+)$, respectively.

By using probability rules, independence relationships and graph separation, it can be shown that

$$\text{BEL}(x) = \alpha \left[\prod_{j=1}^m \lambda_{Y_j}(x) \right] \left[\sum_u P(x|u) \prod_{i=1}^n \pi_X(u_i) \right], \quad (2.20)$$

where $\lambda_{Y_j}(x) = P(e_{XY_j}^-|x)$ are messages received from X 's children Y , $e_{XY_j}^-$ stands for evidence contained in the subnetwork on the head side of the link $X \rightarrow Y_j$. $\pi_X(u_i) = P(u_i|e_{U_iX}^+)$ are messages received from X 's parents U , $e_{U_iX}^+$ stands for evidence contained in the subnetwork on the tail side of the link $U_i \rightarrow X$, and $P(x|u)$ is an entry of the conditional probability matrix that quantifies the link between X and U .

The message going to a parent can be calculated from the messages received from the children and the matrix stored on the link from the parent:

$$\lambda_X(u_i) = \beta \sum_x \lambda(x) \sum_{u_k: k \neq i} P(x|u) \prod_{k \neq i} \pi_X(u_k), \quad (2.21)$$

where $\lambda(x) = \prod_{j=1}^m \lambda_{Y_j}(x)$, $\lambda_{Y_j}(x) = P(e_{XY_j}^-|x)$, and $\pi_X(u_k)$ are messages received from X 's parents U .

The message going to a child is given by

$$\pi_{Y_j}(x) = \alpha \prod_{k \neq j} \lambda_{Y_k}(x) \pi(x), \quad (2.22)$$

where $\pi(x) = \sum_u P(x|u) \prod_{i=1}^n \pi_X(u_i)$ and $\pi_X(u_i) = P(u_i|e_{XY_j}^+)$. Therefore, outgoing messages, $\lambda_X(u_i)$ and $\pi_{Y_j}(x)$ are determined from the incoming messages $\lambda_{Y_j}(x)$ and $\pi_X(u_i)$.

2.4.2 Inference in DNs

The computation of joint and marginal distributions in a DN is not as straightforward as in a BN mainly because of inconsistency issues. Consequently, methods based on stochastic

sampling are used in order to approximate the joint distribution. This also makes sense since the conditional distribution of each variable is known as a result of the parameter learning step in which the classification or regression algorithms compute the necessary conditional probabilities as discussed in Section 2.3.1. However, in the ideal case where a DN is consistent, the independencies encoded in the network structure can be used to lighten the computation involved in the inference task. These two different scenarios will be discussed next.

Consistent DN

Probabilistic inference in a consistent DN \mathcal{D} for a set of random variables \mathbf{X} can be done by converting \mathcal{D} into a Markov network \mathcal{M} , triangulating \mathcal{M} , giving \mathcal{T} , and then applying one of the standard algorithms for probabilistic inference, such as the junction tree algorithm of Jensen et al. (1990), in the triangulated representation \mathcal{T} . A triangulated graph, also called a *chordal* graph, is a graph whose cycles of four or more nodes have a *chord*, which is an edge joining two nodes that are not adjacent in the cycle.

An alternative solution to the previous procedure is to use Gibbs sampling in order to recover the joint distribution $p(\mathbf{x})$ of a consistent DN for \mathbf{X} . Gibbs sampling consists of cycling through each variable X_1, \dots, X_n , in that order, and resampling each X_i according to $p(x_i | \mathbf{x} \setminus x_i) = p(x_i | \mathbf{pa}_i)$, after having initialised each variable to some arbitrary value. This procedure is also called an *ordered Gibbs sampler* and the proof showing that it recovers the joint distribution for \mathbf{X} , when applied to a consistent DN, is given in Heckerman et al. (2000). The proof is based on the irreducibility of the Markov chain, that comes from the sequential samples of x , which is due to the positivity of the local distributions. Those samples are then used to estimate $p(\mathbf{X})$.

The Gibbs sampler is also used to compute a conditional probability but with a slight modification: the instance of the set of conditioning variables is fixed during the Gibbs sampling algorithm in order to speed up its convergence. The reason is a possible small value of the probability of the set of conditioning variables, due to the fact that this set has many variables, which requires many iterations to get an accurate estimate of the conditional probability of interest. For example, if $p(\mathbf{y} | \mathbf{z})$ needs to be computed, then $\mathbf{Z} = \mathbf{z}$ is fixed during the ordered Gibbs sampling. This procedure is also called a *modified ordered Gibbs sampler*.

A decomposition of the inference task into a set of inference tasks on single variables is also possible by using the independencies encoded in the corresponding DN along with the law of total probability in order to compute the conditional probability of interest.

For example, if a DN \mathcal{D} is given by $[X_1 \quad X_2 \leftrightarrow X_3]$, then the independencies in \mathcal{D} imply that $p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3|x_2)$, and $p(x_1, x_2, x_3)$ can be obtained by computing each term separately as follows: the distribution of $p(x_1)$ can be read directly from the local distribution for X_1 in \mathcal{D} and thus no Gibbs sampling is required. A modified ordered Gibbs sampler can be used to determine $p(x_2)$ and $p(x_3|x_2)$ where the target variable in the sampler is x_2 and x_3 , respectively.

General DN

In most situations, a DN will be inconsistent due to the way its network structure and parameters are learned as explained in Section 2.3.2. The joint distribution of an inconsistent DN is approximated using the ordered Gibbs sampling described earlier. Since the joint distribution is likely to be inconsistent with the conditional distributions that are used to produce it, the procedure is called an *ordered pseudo-Gibbs sampler*. Another drawback of this procedure is that the joint distribution obtained will depend on the order in which the pseudo-Gibbs sampler visits the variables.

A comparison between BNs and DNs regarding the accuracy of the joint probabilities produced by each of them on different datasets has been carried out in Heckerman et al. (2000) and a better performance of BNs has also been reported although the results from the DNs are close to those of the BNs. This outcome is attributed to the number of parameters which are fewer in a BN than in a DN and also because the joint probabilities are computed via multiplication in a BN whereas they are obtained via pseudo-Gibbs sampling in a DN where non-convergence of sampling can occur due to the inconsistency of the joint distribution.

2.5 Summary

BNs and DNs are both graphical representations of probability distributions. They are characterised by a network structure, which is a graph where each node corresponds to a random variable from a dataset, and some parameters represented by the conditional probabilities of the variables. BNs usually display causal relationships between random variables whereas DNs were devised to relax this hard constraint associated with the causal relationships in BNs. Consequently, unlike a BN, which is represented by a directed acyclic graph, a DN structure is likely to have cycles.

The structure of a BN is usually learned by combining the BDe scoring function with a hill-climbing heuristic search although a conditional independence test can also be used to

construct a BN structure. On the other hand, a classification/regression algorithm governs the structure learning of a DN. The conjugacy property of the Dirichlet distribution for discrete variables, and the Normal distribution properties for continuous variables enable the computation of the conditional probabilities in a BN. In contrast, these conditional probabilities are directly estimated from the data using a classification or regression techniques in a DN. The structure and parameters of a BN and a DN are both learned from data although expert knowledge can also be used to specify the conditional probabilities in a BN.

The computation of the joint distribution in a BN is performed through the Markov property. The joint probability is obtained by taking the product of all the conditional probabilities as a result of the Markov property. The Markov property simplifies the computation of the joint distribution in a BN whereas in a DN, a stochastic sampling called an ordered pseudo-Gibbs sampling is needed due to the fact that conditional probabilities are not consistent with the joint distribution.

The structure learning of a graphical model is especially crucial since both the parameter learning and the inference task depend on it. In BNs, the structure is learned first then the parameters whereas in DNs the structure and parameters are learned simultaneously. However the structure learning in both BNs and DNs is based on heuristic optimisation techniques. This issue can be addressed using a more principled structure search algorithm and the resulting graphical model instance will be discussed in the next chapter.

3 Affinity Network

In this chapter a novel graphical model structure that we call an Affinity Network (AN) and which has preferable characteristics to the graphical models discussed in the previous chapter is introduced. An AN is a directed, possibly cyclic, graphical model in which the nodes represent random variables. The links between those random variables denote probabilistic relationships based on similarities between two connected nodes. The AN shares some properties of a BN and a DN in the sense that the conditional probabilities of an AN are learned using a similar approach to learn the conditional probabilities of a BN. In addition, the joint probability of an AN can be computed the same way as in a DN. The AN differs from the BN and the DN in its graph learning, also called network structure learning.

An important aspect of a graphical model in practice is its interpretability. Dependency Networks (DNs) are easier to interpret than Bayesian Networks (BNs) due to their more flexible dependency relationships compared to the strong causal meaning of the arcs in BNs. An affinity network (AN) shares this user-friendly property of DNs since it displays variables that are close to each other given the data. Since BNs do not have cycles, they cannot discriminate spurious causal relations between confounding variables (Pearl,

1998). In contrast, AN and DN graphs can be cyclic and can therefore capture more general dependency relationships. The way an AN structure is constructed is superior to BNs and DNs structure learning as it is based on a data-driven message update of simple formulae that search for minima of a scoring function, whereas BNs and DNs structure learning relies on heuristic search procedures.

An AN can then be viewed as a DN but with a better network structure search method. This will be illustrated later in this chapter using small and large real-world and synthetic datasets. It will also be shown in Chapter 6 that an AN can outperform a DN when characterising cluster synchronisation of dynamical systems from time series data. Moreover, given that the AN uses similarity measures between variables, in the absence of raw data, when only those similarity values are available, it would still be possible to construct an AN whereas a DN needs the raw data to learn the network structure through a classification/regression procedure.

In this chapter the new AN structure learning, using a combination of the soft-constraint affinity propagation (SCAP) algorithm (Leone et al., 2007) and a scoring function, will be introduced after describing the affinity propagation (AP) algorithm, upon which SCAP is based, and how the parameters are learned.

3.1 Affinity propagation

AP is a general optimisation algorithm that searches for a global maximum of an energy function. It can be derived as the max-sum algorithm in a factor graph, which is another graphical representation for the factorisation of a joint probability distribution (Frey et al., 1997; Kschischang et al., 2001). It has been used for data clustering in Frey and Dueck (2007) where AP looks at pairs of variables, also called data points, take a similarity matrix s as an input from which a responsibility matrix r and an availability matrix a are constructed, and return a vector containing the cluster center for each data point. In the following, the discussion will be put in the context of learning an AN structure, as an introduction to the AN structure, where the goal is to find the set of parents \mathbf{PA} of each variable in a dataset. In other words, a random variable X_i and its potential parent variable pa_i , in the AN structure learning, can be viewed as a data point at index i and its cluster center c_i respectively, in the AP clustering algorithm.

The problem of learning a graphical network structure can then be summarised as finding the set of parents \mathbf{PA}_i for each variable X_i , $i = 1, \dots, n$ in a dataset. To facilitate the discussion, we suppose that a variable has only one parent and the general case will be

described later. The identification of the parents can also be viewed as searching over the space of valid configurations of $\mathbf{PA} = \{pa_1, \dots, pa_n\}$, where the scalar $pa_i, i = 1, \dots, n$ is the parent of X_i , so as to minimise the energy

$$E(\mathbf{PA}) = - \sum_{i=1}^n s(X_i, pa_i), \quad (3.1)$$

where $s(X_i, pa_i)$ indicates how well the variable pa_i is suited to be the parent of the variable X_i . Usually the maximisation of the network similarity, S_{net} , which is the negative energy plus a constraint function that enforces valid parent configurations, is performed:

$$S_{net}(\mathbf{PA}) = -E(\mathbf{PA}) + \sum_{k=1}^n \delta_k(\mathbf{PA}), \quad (3.2)$$

$$= \sum_{i=1}^n s(X_i, pa_i) + \sum_{k=1}^n \delta_k(\mathbf{PA}), \quad (3.3)$$

where $\delta_k(\mathbf{PA})$ is a penalty term that equals $-\infty$ if some variable X_k , whose parent is different from itself, has been chosen to be the parent of variable X_i and 0 otherwise. This means that whenever a variable is selected as a parent by another variable, it is not allowed to have a parent other than itself. Equation (3.3) can be represented using a factor graph \mathcal{F} in which each term in (\mathbf{PA}) is represented by a function node and each parent pa_i is represented by a variable node. Function nodes and variable nodes are connected by edges, and a variable node is connected to a function node if and only if its corresponding term depends on the variable. For example, the term $s(X_i, pa_i)$ in the above expression has a corresponding function node that is connected to the single variable pa_i and the term $\delta_k(\mathbf{PA})$ has a corresponding function node that is connected to all variables pa_1, \dots, pa_n as shown in Figure 3.1. The global function $S(\mathbf{PA})$ in the factor graph \mathcal{F} is given by the sum of all the functions represented by function nodes. The max-sum algorithm, which is the log-domain version of the sum-product algorithm (and by replacing the “sum” operator with the “max”), can be used to search over configurations of the parents \mathbf{PA} in the factor graph that maximise $S(\mathbf{PA})$. The max-sum algorithm for the factor graph \mathcal{F} consists of recursively sending messages from variables to functions and from functions to variables as follows:

- the message sent from pa_i to $\delta_k(\mathbf{PA})$ consists of n real numbers $\begin{pmatrix} \rho_{i \rightarrow k}(X_1) \\ \vdots \\ \rho_{i \rightarrow k}(X_n) \end{pmatrix}$ (Figure 3.1(b)).

- The message sent from $\delta_k(\mathbf{PA})$ to pa_i also consists of n numbers $\begin{pmatrix} \alpha_{i \leftarrow k}(X_1) \\ \vdots \\ \alpha_{i \leftarrow k}(X_n) \end{pmatrix}$

(Figure 3.1(c)).

The value of pa_i can be estimated at any time by summing all incoming availability α and similarity S messages (Figure 3.1(d)). The responsibility ρ messages are computed as the element-wise sum of all incoming messages because they are outgoing from variables:

$$\rho_{i \rightarrow k}(pa_i) = s(X_i, pa_i) + \sum_{k': k' \neq k} \alpha_{i \leftarrow k'}(pa_i). \quad (3.4)$$

The sum of the incoming messages and the maximum over all variables except the variable the message is being sent to, give the messages sent from functions to variables. For instance, the message sent from the penalty function δ_k to the parent variable pa_i is:

$$\alpha_{i \leftarrow k}(pa_i) = \overbrace{\max_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n} \left[\delta_k(X_1, \dots, X_{i-1}, pa_i, X_{i+1}, \dots, X_n) + \sum_{i': i' \neq i} \rho_{i' \rightarrow k}(X_{i'}) \right]}^{\text{best possible configuration satisfying } \delta_k \text{ given } pa_i}. \quad (3.5)$$

After some simplifications that are described in Appendix A, messages can be considered to be scalar. This allows to define the responsibilities $r(X_i, X_k)$ and availabilities $a(X_i, X_k)$ as:

$$r(X_i, X_k) = s(X_i, X_k) - \max_{X_j: X_j \neq X_k} \left[s(X_i, X_j) + a(X_i, X_j) \right] \quad (3.6)$$

$$a(X_i, X_k) = \begin{cases} \sum_{i': i' \neq k} \max(0, r(X_{i'}, X_k)), & \text{for } X_k = X_i \\ \min \left[0, r(X_k, X_k) + \sum_{i': i' \notin \{i, k\}} \max(0, r(X_{i'}, X_k)) \right], & \text{for } X_k \neq X_i \end{cases} \quad (3.7)$$

The responsibility $r(X_i, X_k)$ reflects the accumulated evidence for how well-suited the variable X_k is to serve as the parent of variable X_i . The availability $a(X_i, X_k)$ reflects the accumulated evidence for how appropriate it would be for variable X_i to choose variable X_k as its parent, taking into account the support from other variables that variable X_k should be a parent.

The $\min[0, \cdot]$ in the availability update comes from the fact that $x - \max(0, x) = \min(0, x)$.

The value of a variable pa_i can be estimated after any iteration by summing together all

incoming messages to pa_i and taking the value, $p\hat{a}_i$, that maximises it:

$$p\hat{a}_i = \operatorname{argmax}_{X_j} \left[\sum_k \alpha_{i \leftarrow k}(X_j) + s(X_i, X_j) \right], \quad (3.8)$$

$$= \operatorname{argmax}_{X_j} \left[\sum_k \tilde{\alpha}_{i \leftarrow k}(X_j) + \sum_k \bar{\alpha}_{i \leftarrow k} + s(X_i, X_j) \right], \quad (3.9)$$

$$= \operatorname{argmax}_{X_j} \left[a(X_i, X_j) + s(X_i, X_j) \right], \quad (3.10)$$

where $\bar{\alpha}_{i \leftarrow k} = \alpha_{i \leftarrow k}(pa_i : pa_i \neq X_k)$ which will make $\tilde{\alpha}_{i \leftarrow k}(pa_i)$ zero for all $pa_i \neq X_k$. This is also equivalent to $\sum_{k': k' \neq k} \tilde{\alpha}_{i \leftarrow k'}(pa_i) = \tilde{\alpha}_{i \leftarrow pa_i}(pa_i)$ for $pa_i \neq X_k$ and the summation is zero for $pa_i = X_k$. When the energy function in Equation (3.1) has multiple minima with corresponding multiple fixed points of the update rules, then the max-sum algorithm that is used to maximise the network similarity in Equation (3.3) represented by the factor graph in Figure 3.1 may not converge. This can be tackled by damping the messages or by adding a small noise to the similarities (Frey and Dueck, 2007). The messages are damped by setting each message to λ times its value from the previous iteration plus $1 - \lambda$ times its prescribed updated value, where λ is a damping factor between 0 and 1.

In the Affinity Propagation (AP) algorithm, the fact that a parent variable has to be its own parent is considered as a hard constraint. This hard constraint can be relaxed in the Soft-Constraint Affinity Propagation (SCAP) and will be discussed next.

3.2 Soft-Constraint Affinity Propagation

In this section, we provide a generalisation of the SCAP equations (Leone et al., 2007) which are similar to the AP equations except that a constraint \tilde{p} is introduced in the expressions of the self-responsibility $r(X_i, X_i)$ and the self-availability $a(X_i, X_i)$ as follows:

$$r(X_i, X_j) = s(X_i, X_j) - \max_{k \neq j} [s(X_i, X_k) + a(X_i, X_k)] \quad (3.11)$$

$$r(X_i, X_i) = \max \left[-\tilde{p}, s(X_i, X_i) - \max_{k \neq i} \{s(X_i, X_k) + a(X_i, X_k)\} \right] \quad (3.12)$$

$$a(X_i, X_j) = \min \left[0, r(X_j, X_j) + \sum_{k \neq j} \max(0, r(X_k, X_j)) \right] \quad (3.13)$$

$$a(X_i, X_i) = \min \left[\tilde{p}, \sum_{k \neq i} \max\{0, r(X_k, X_i)\} \right] \quad (3.14)$$

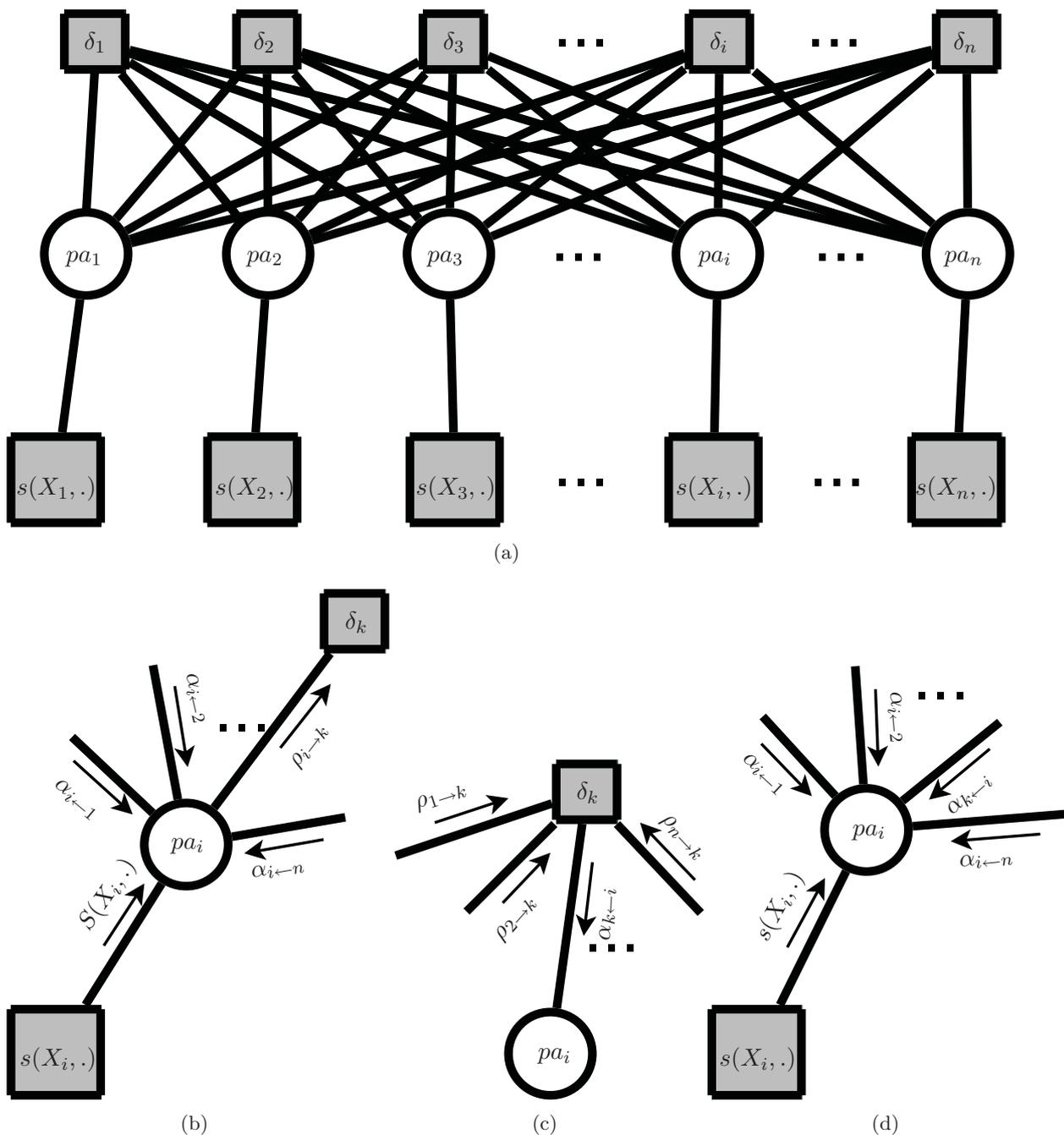


Figure 3.1: Factor graph for Affinity Propagation in the context of AN structure learning. (a): $s(X_i, \cdot)$ indicates how well the scalar variable pa_i is suited to be the parent of the variable X_i . δ_k is a penalty term that equals $-\infty$ if some variable X_k has been chosen to be the parent of variable X_i without X_k being a parent (i.e., $X_k \notin \mathbf{PA}$) and 0 otherwise. (b): the message sent from pa_i to $\delta_k(\mathbf{PA})$ consists of n real numbers (one for each possible value, X_j , of pa_i) and can be denoted as $\rho_{i \rightarrow k}(X_j)$. $\alpha_{i \rightarrow k}$ represents the message sent from the penalty function δ_k to the parent variable pa_i and $\rho_{i \rightarrow k}$ denotes the message sent from the parent variable pa_i to the penalty function δ_k . (c): the message sent from $\delta_k(\mathbf{PA})$ to pa_i also consists of n numbers and can be written as $\alpha_{i \rightarrow k}(X_j)$. (d): the value of pa_i can be estimated at any time by summing all incoming availability α and similarity S messages.

Using the same reasoning as for Equation (3.10), the parent $p\hat{a}_i$ of any variable X_i can be computed by maximising the marginal a posteriori probability:

$$p\hat{a}_i = \operatorname{argmax}_{X_j} \left[a(X_i, X_j) + r(X_i, X_j) \right]. \quad (3.15)$$

When $\tilde{p} = +\infty$ (in practice \tilde{p} is set to a large positive number), from (3.12) and (3.14), we have

$$r(X_i, X_i) = s(X_i, X_i) - \max_{k \neq i} \{s(X_i, X_k) + a(X_i, X_k)\}$$

and

$$a(X_i, X_i) = \sum_{k \neq j} \max(0, r(X_k, X_i))$$

which is equivalent to the AP equations. On the other hand, if $\tilde{p} = -\infty$ (that is, a large negative number in practice), then $a(X_i, X_i) = -\infty$, $r(X_i, X_i) = +\infty$ and their sum is equal to zero. From Equation (3.15), self-loops are discouraged and the hard constraint in AP is relaxed when $a(X_i, X_i) + r(X_i, X_i) > 0$. From (3.13) and (3.11), this occurs when

$$\min \left[0, r(X_j, X_j) + \sum_{k \neq j} \max(0, r(X_k, X_j)) \right] + s(X_i, X_j) - \max_{k \neq j} [s(X_i, X_k) + a(X_i, X_k)] > 0. \quad (3.16)$$

Inequality (3.16) holds if

$$\min \left[0, r(X_j, X_j) + \sum_{k \neq j} \max(0, r(X_k, X_j)) \right] \geq 0, \quad (3.17)$$

and

$$s(X_i, X_j) - \max_{k \neq j} [s(X_i, X_k) + a(X_i, X_k)] > 0. \quad (3.18)$$

Inequality (3.17) is valid if $r(X_i, X_i) \geq 0$, that is, using (3.12),

$$\max \left[-\tilde{p}, s(X_j, X_j) - \max_{k \neq j} \{s(X_j, X_k) + a(X_j, X_k)\} \right] \geq 0. \quad (3.19)$$

This is equivalent to

$$\tilde{p} \leq 0, \quad (3.20)$$

and

$$\tilde{p} < s(X_j, X_j) - \max_{k \neq j} \{s(X_j, X_k) + a(X_j, X_k)\}. \quad (3.21)$$

In summary, the inequalities in (3.18), (3.20) and (3.21) need to be satisfied to avoid self-loops. In Leone et al. (2007), \tilde{p} was set to be in the interval $[0, +\infty]$ but here we have

expanded the interval to $[-\infty, +\infty]$ in order to remove the possibility of having self-loops.

Since the parent of a variable is selected through the maximum M of the sum of the availability a and the responsibility r this implies that a variable can only have one parent. However, by using a threshold on the value of M multiple parents could be obtained but we do not pursue this further.

3.3 Parameter and structure learning for ANs

In this section, algorithms that learn the parameters and structure in an AN will be discussed along with a simple method to quantify the strengths of the arcs between connected nodes.

3.3.1 Parameter learning for ANs

Parameter learning in an AN is similar to learning the parameters of a BN and is summarised as follows: the parameters are learned by modelling each variable with a Dirichlet distribution for discrete variables or a Gaussian distribution for continuous variables. As a result, the computation of the posterior distribution is simplified as well as the expression of the conditional distribution which is given by Equation (2.1) and Equation (2.2) for discrete and continuous variables respectively.

3.3.2 Structure learning for ANs

The SCAP method of constructing an AN combines a scoring function and soft-constraint affinity propagation (SCAP) (Leone et al., 2007) as a search algorithm. SCAP is used as a local network structure search algorithm in our case since it can be viewed as a method that searches for minima of an energy function where the energy function here is represented by a scoring function. The search is local in the sense that each node looks at its most probable parent where a parent is equivalent to a cluster center in the AP clustering algorithm. The $\mathcal{O}(n^2)$ complexity of the SCAP algorithm, where n is the number of data points, makes it suitable for big datasets and for network structures having loops where this approach of recursively propagating messages has been used to approximate NP-hard problems (Frey and Dueck, 2007).

A general AN structure might have loops and hence the structure search algorithm should be able to find them. Unfortunately, standard AP cannot perform this task as empirically shown in Leone et al. (2007) which is why a soft-constraint version of AP, called SCAP, has been devised in Leone et al. (2007) to cluster gene-expression data. Like AP,

SCAP looks at pairs of variables, also called data points, takes a similarity matrix s as an input from which a responsibility matrix r and an availability matrix a are constructed, and return a vector containing the cluster center for each data point. In our case, the entry $s(i, j)$ of the similarity matrix s is given by the BDe score of the two-nodes network structure where the node j is the parent of the node i . This means that the similarity $s(i, j)$ indicates how well the variable with index j is suited to be the parent of the variable i . The responsibility $r(i, j)$ reflects the accumulated evidence for how well-suited the variable j is to serve as the parent of variable i and the availability $a(i, j)$ reflects the accumulated evidence for how appropriate it would be for the variable i to choose the variable j as its parent.

The values in the diagonal of the similarity matrix s are also called *preferences* and variables with larger values of $s(k, k)$ are more likely to be chosen as parents. If all variables are equally suitable as parents then the same value is assigned to the preference of each variable. This value is usually set to the median of the input similarities (Frey and Dueck, 2007). It should be noted that setting the preference value in this way might not always yield the best result in terms of clustering and implicitly in terms of model selection especially for high-dimensional data. This observation has been verified on a range of clustering problems. Heuristic solutions such as the Adaptive AP (Wang et al., 2007) have been proposed in order to tackle this problem but no efficient method has been found so far to the best of our knowledge. However, in our case, the value of $s(k, k)$ is set to the BDe score of the single-node network X_k .

The affinity between two data points is given by the sum of their availability and responsibility. The corresponding SCAP equations are given by:

$$\begin{aligned} r(i, j) &= s(i, j) - \max_{k \neq j} [s(i, k) + a(i, k)], \\ r(i, i) &= \max \left[-\tilde{p}, s(i, i) - \max_{k \neq i} \{s(i, k) + a(i, k)\} \right], \\ a(i, j) &= \min \left[0, r(j, j) + \sum_{k \neq j} \max(0, r(k, j)) \right], \\ a(i, i) &= \min \left[\tilde{p}, \sum_{k \neq i} \max\{0, r(k, i)\} \right], \end{aligned}$$

where i , j and k are variable indices (the variable at index j being the parent of the variable at index i as defined previously). The responsibility and availability matrices are then added together and the index of the maximum value for each row is considered as the best parent index for the variable in the corresponding row. The search algorithm is stopped when a predefined maximum number of iterations has been reached or when

there is no change in the output. It should be noted that, like in the AP algorithm, the messages are damped to avoid numerical oscillations during their update. A default value of $\lambda = 0.5$ has been used as in Frey and Dueck (2007). The AN structure learning method is summarised in Algorithm 3.3.1. In our experiments, we have set \tilde{p} to $-\infty$ in order to relax the hard constraint in AP and to remove the possibility of having self-loops.

Algorithm 3.3.1: AN structure learning. The entry `vector_parent_indices(k)` corresponds to the parent index of the variable at index k , where $k = 1, \dots, n$.

Input: Data having n variables

Output: Vector of parent index for each variable

for $i = 1; i \leq n; i = i + 1$ **do**

for $j = 1; j \leq n; j = j + 1$ **do**

 // similarity matrix

`sim_mat(i,j)` = BDe score of the network structure where the variable at index j is the parent of the variable at index i ;

end

end

`vector_parent_indices` = SCAP(`sim_mat`);

3.3.3 Quantification of the arcs' strength in an AN

In a DN, the significance of connections between variables was originally ranked by determining the order in which arcs would be added during a greedy structure search (Heckerman et al., 2000). The magnitude of the arcs' significance is quantified by computing the mutual information in the case of discrete variables. The mutual information I which measures the mutual dependence of random variables X and Y is given by

$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \quad (3.22)$$

In our case, the joint probability $p(x, y)$ is computed using the conditional probability and the product rule of probability which gives $p(x, y) = p(x|y)p(y)$. A transformation of Equation (3.22) can be used so that the value of the mutual dependence between two nodes is normalised and hence can be interpreted as a generalisation of the correlation as discussed in Joe (1989); Gentleman et al. (2005). This transformation will be denoted by δ^* (Gentleman et al., 2005) which is given by

$$\delta^* = \sqrt{1 - \exp(-2I)}, \quad (3.23)$$

where I is the mutual information given by Equation (3.22).

For continuous variables, the arc strength value is given by the covariance between two

connected variables. The covariance matrix is computed using Equation (2.4).

3.4 Experimental results

In this section, we compare AN and DN network structures after giving a description of the main datasets that have been used to carry out the comparison.

3.4.1 Datasets

Synthetic and real-world datasets of different sizes will be presented in the following section. The description will focus only on the data for which graphical AN and DN network structures have been generated and a reference to the remaining data will be given.

Table 3.1: Chest clinic dataset attribute information. There are 16 instances that contain a single missing value and there are 458 (65.5%) benign cases and 241 (34.5%) malignant cases. The sample code number is not used in the analysis.

α :	$p(a)$	=	0.01	ϵ :	$p(e l, t)$	=	1
					$p(e \bar{l}, \bar{t})$	=	1
τ :	$p(t a)$	=	0.05		$p(e \bar{l}, t)$	=	1
	$p(t \bar{a})$	=	0.01		$p(e \bar{l}, \bar{t})$	=	0
σ :	$p(s)$	=	0.50	ζ :	$p(x e)$	=	0.98
					$p(x \bar{e})$	=	0.05
λ :	$p(l s)$	=	0.10	δ :	$p(d e, b)$	=	0.90
	$p(l \bar{s})$	=	0.01		$p(d e, \bar{b})$	=	0.70
β :	$p(b s)$	=	0.60		$p(d \bar{e}, b)$	=	0.80
	$p(b \bar{s})$	=	0.30		$p(d \bar{e}, \bar{b})$	=	0.10

Synthetic data

Chest clinic dataset:

The Chest clinic data is a synthetic binary dataset that has been generated using probabilistic logic sampling (PLS) (Henrion, 1988; Lin and Druzdzel, 1989), which is a stochastic sampling algorithm, by randomly instantiating each node to one of its possible states given the instantiated states of its parents. The instantiations are performed in a topological order, in other words, parent nodes are sampled before their children nodes.

The network, which comprises eight nodes and eight arcs, models the following hypothetical situation (Lauritzen and Spiegelhalter, 1988) in which the variable associated to each node is put between brackets: a patient presents at a chest clinic with dyspnoea (δ),

and has recently visited Asia (α). Smoking history (σ) and chest X-ray (ζ) are not yet available. The doctor would like to know the chance that each of the diseases is present, and if tuberculosis (θ) were ruled out by another test, how would that change the belief in lung cancer (λ)? Also, would knowing smoking history or getting an X-ray contribute most information about cancer, given that smoking may ‘explain away’ the dyspnoea since bronchitis is considered a possibility? Finally, when all information is incorporated, can we identify which was the most influential in forming our judgement?

The probabilities of each state conditional on all possible parent states were assessed from medical quantitative knowledge and are given in Table 3.1 where, for example, a is used to indicate a positive response on the node α ‘VisitAsia’, \bar{a} to indicate a negative response, and $p(a)$ to stand for $P(\alpha = a)$ and similarly for the other nodes.

ALARM dataset:

ALARM (A Logical Alarm Reduction Mechanism) is a patient monitoring system that uses a Bayesian network to represent the causal relationships between diagnosis, measurements and other non-measurable variables for diagnosis purposes (Beinlich et al., 1989). The system contains prior probabilities and conditional probabilities computed from equations relating the variables and a number of subjective assessments. These probabilities are used to generate the data through the PLS algorithm.

Real-world data

Ferrara breast cancer dataset:

This is a breast cancer dataset which was used in Ambrogi et al. (2006) where the authors tried to identify tumour profiles with possible clinical relevance based on clusters of immunohistochemical molecular markers measured on a large, single institution, case series.

Tumour biological profiles were explored on 633 archival tissue samples taken from patients who underwent surgery for primary infiltrating breast cancer between 1983 and 1992 at the University of Ferrara, Italy. Five validated markers were considered namely estrogen receptors (ER), progesterone receptors (PR), Ki-67/MIB1 proliferation index (PROLIND), HER2/NEU (NEU), and p53 in their original scale of measurement.

The percentage of expression values of ER, PR, and NEU tended to distribute around the following values: 0%, 10%, 25%, 50%, 75% and 100%, and were consequently discretised on these values. Percentages of Ki-67 and p53-expressing cells were analysed without discretisation in the original paper but have been discretised to the previous values in this

work because they are continuous. The dataset is summarised in Table 3.2. We would like to thank Elia Biganzoli from the University of Milan for giving us access to this data.

Wisconsin breast cancer dataset:

Table 3.2: Distribution in 633 cases of primary invasive breast cancer of the clinical and pathological variables, of the discretised biologic variables ER, PR, NEU, and of the categorised variables, KI-67 and p53 from the Ferrara breast cancer dataset (Ambrogi et al., 2006).

Variable	Frequency (%)
ERs	
0	116 (18.3)
10	36 (5.7)
25	80 (12.6)
50	131 (20.7)
75	184 (29.1)
100	86 (13.6)
PR	
0	182 (28.8)
10	79 (12.5)
25	65 (10.3)
50	76 (12.0)
75	108 (17.0)
100	123 (19.4)
HER2/NEU	
0	328 (51.8)
10	132 (20.9)
25	45 (7.1)
50	26 (4.1)
75	78 (12.3)
100	24 (3.8)
Ki-67/MIB-1 proliferation index (PROLIND)	
q1 (0-2.5)	129 (20.4)
q2 (2.5-5.75)	125 (19.7)
q3 (5.75-13)	143 (22.6)
q4 (13-30)	114 (18.0)
q5 (30-90.8)	122 (19.3)
p53	
0 (0-0.9)	293 (46.3)
10 (0.9-10)	175 (27.7)
75 (10-75)	78 (12.3)
100 (75-100)	87 (13.7)

This incomplete dataset was taken from the Machine Learning Repository¹ which is available publicly. It is also a breast cancer dataset provided by Dr. William H. Wolberg from the University of Wisconsin Hospitals and has been used by Wolberg and Mangasarian

¹<http://www.ics.uci.edu/~mllearn/MLSummary.html>

(1990) and Zhang (1992) for medical diagnosis and learning respectively.

The dataset reflects a chronological grouping of the data since samples arrive periodically as Dr. Wolberg reports his clinical cases. This grouping information appears below, having been removed from the data itself:

Group 1:	367	instances	(January 1989)
Group 2:	70	instances	(October 1989)
Group 3:	31	instances	(February 1990)
Group 4:	17	instances	(April 1990)
Group 5:	48	instances	(August 1990)
Group 6:	49	instances	(Updated January 1991)
Group 7:	31	instances	(June 1991)
Group 8:	86	instances	(November 1991)
Total:	699	points	(as of the donated database on 15 July 1992)

Table 3.3: Wisconsin breast cancer dataset attribute information. There are 16 instances that contain a single missing value and there are 458 (65.5%) benign cases and 241 (34.5%) malignant cases. The sample code number is not used in the analysis.

#	Attribute	Domain
1.	Sample code number	id number
2.	Clump Thickness	1-10
3.	Uniformity of Cell Size	1-10
4.	Uniformity of Cell Shape	1-10
5.	Marginal Adhesion	1-10
6.	Single Epithelial Cell Size	1-10
7.	Bare Nuclei	1-10
8.	Bland Chromatin	1-10
9.	Normal Nucleoli	1-10
10.	Mitoses	1-10
11.	Class	(2 for benign, 4 for malignant)

3.4.2 Comparison between AN and DN

In the following, AN and DN structures obtained from a classification tree and from a combination of SCAP and the BDe score are compared. The comparison first starts from small discrete synthetic and real datasets which are taken from the Machine Learning Repository except the Ferrara breast cancer data and the Chest data. In the following comparisons the variable which has the maximum BDe score is selected as the parent of the corresponding child variable in the method for constructing the AN structure. Likewise, only the root of the classification tree is used as the parent of the corresponding child

variable in the DN structure learning approach.

Since the small datasets have few variables, an exhaustive search can be used to find their network structures as benchmarks. More specifically, for each variable in the data, the parent combination that maximises the BDe score among the set of all possible parent combinations is chosen as the parent set. The gold standard network is computed as such given that a local structure search is performed for both the AN and the DN. In other words, for each variable in the data, the selected parent set is the one that optimises the scoring criterion or the regression/classification function. The synthetic data are generated using probabilistic logic sampling (Henrion, 1988) from the chest-clinic model (Lauritzen and Spiegelhalter, 1988).

As can be seen in the main results of Table 3.4, the combination of SCAP and BDe is able to reconstruct the essential graphs (undirected graphs) of the small datasets as shown in Table 3.4 while the classification tree usually adds extra parents or fewer parents but with more deleted arcs when only the root of the tree is taken into account. The Nursery data is the only case where the root of the classification tree gives the true network structure. Figure 3.2 shows the AN and DN structures of the Chest data. In Figure 3.2(b), the AN obtained from the SCAP and BDe method provides extra information through the quantification of the strength of each arc which is useful especially in the medical domain.

	AN (SCAP+BDe)	DN (Tree)	DN (Tree root)
Ferrara	0,0,0	0,15,0	0,3,3
Wisconsin	8,0,0	0,78,0	1,9,9
Chest	0,0,0	2,32,2	0,1,1
Lenses	3,0,0	0,8,1	0,1,1
Nursery	7,0,0	0,63,0	0,0,0
Balance scale	3,0,0	0,15,0	1,4,4
Car evaluation	5,0,0	0,35,0	1,6,6
MONK'S	5,0,0	0,8,3	0,2,5
Endgame	8,0,0	0,80,0	1,7,7

Table 3.4: Comparison between the network structures of small datasets given by an exhaustive search in conjunction with the BDe score, the AN structure learned from a combination of SCAP and the BDe score, the DN structures built using a classification tree, and a tree where only the root is taken as the parent of a child variable. The data were taken from the Machine Learning Repository except the Ferrara and Chest data which are private real and synthetic data respectively. The numbers in each entry of the table correspond to the numbers of reversed, added and deleted arcs respectively. Therefore a good network structure has a minimal number of reversed, added and deleted nodes, in other words, has a maximal number of zero entries in the table. The best candidate structure for each dataset is highlighted in bold face.

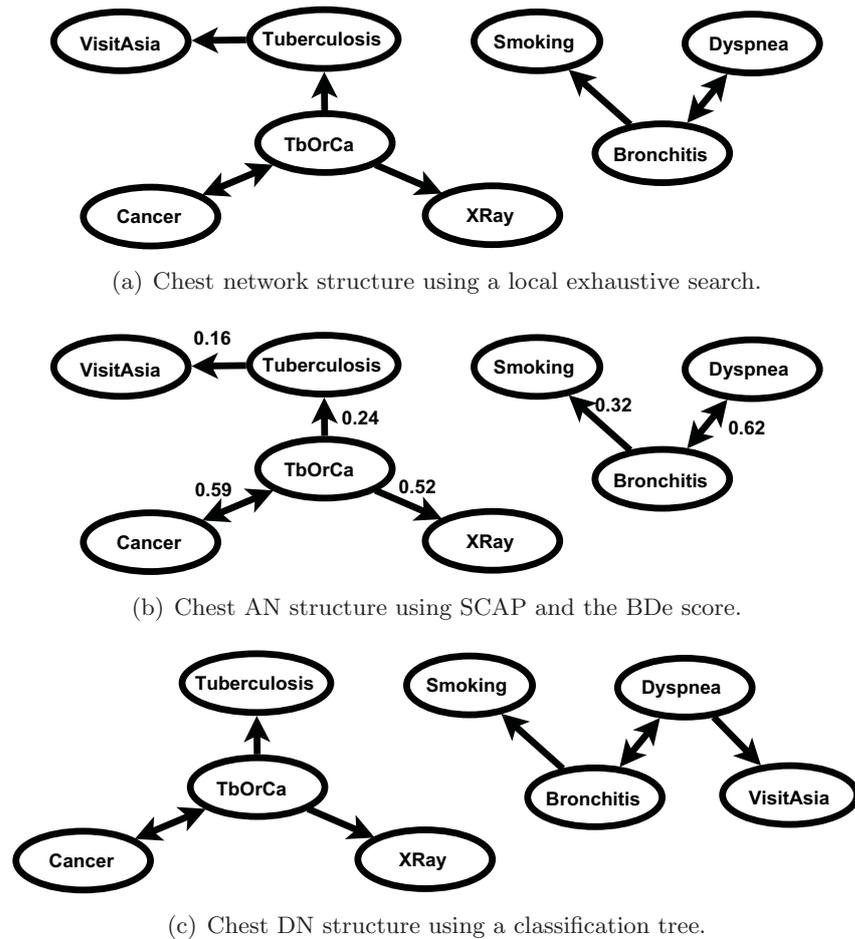


Figure 3.2: Comparison between the AN structure resulting from a combination of SCAP and the BDe score and a DN structure obtained from a classification tree and on a synthetic data with discrete variables. The Chest data is generated by probabilistic logic sampling. Based on a comparison with an exhaustive structure search (see Table 3.4), the AN structure, which is the same as the benchmark network structure, is better than the DN structure, which has one added and one deleted arcs.

Figure 3.3 shows the AN and DN structures for the Ferrara breast cancer data. These two networks are particularly interesting because they respectively show that the progesterone (PR) and estrogen (ER) receptors are the main variables in this dataset. This fact is sensible since breast cancer is considered to be a hormone-dependent cancer and that the balance between these two hormones is vital to the health of every woman.

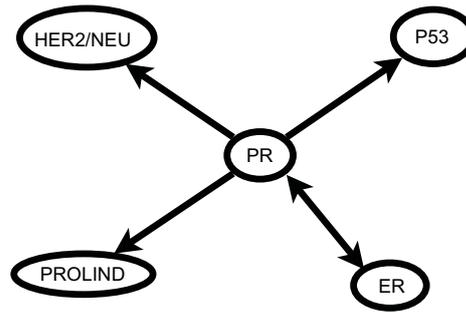
Two arcs are missing when AP and the Bayesian score are used, while with the SCAP and the Bayesian score, only one arc is missing. The correct network structure, when the directions of the arcs are ignored, is recovered with the SCAP when self-loops are avoided. Likewise, when the AP algorithm and the Bayesian score are applied to the Wisconsin breast cancer data, three arcs are missing. When SCAP is used instead of AP the loop is recovered but two arcs are still missing. Finally, by not allowing self-loops, the exact essential graph structure is obtained as in Figure 3.4. Note that each of the three networks contain a loop.

The method has also been tested on bigger artificial (ALARM and Madelon) and real (Insurance company) data. The Madelon and Insurance data are taken from the Machine Learning Repository while the ALARM data is generated using probabilistic logic sampling (Beinlich et al., 1989). The AN structure of the ALARM data is shown in Figure 3.5 along with the DN structure obtained from a classification tree where only the root of the tree is displayed as the parent of a variable.

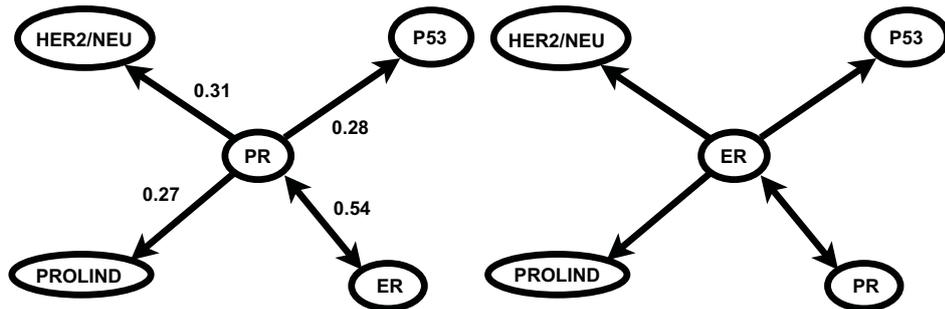
For information, the corresponding BN was constructed in 998 seconds on a Macintosh II running LightSpeed Pascal version 2.0 with one missing and one extra arc using the K2 algorithm and an ordering of the variables (Cooper and Herskovits, 1991) compared to the AN approach where the network structure was constructed in 20 seconds on a machine with two 2.40 GHz Intel Xeon processors with 512 KB memory each.

Database	#Attributes	#Instances	Search time (seconds)	
			Tree	SCAP+BDe
ALARM	37	10000	135	20
Insurance Company	86	5822	230	52
Madelon	500	4400	32415	6421

Table 3.5: Computational time for AN structure construction on big data using a classification tree and a combination of SCAP and the Bayesian score. The experiments were performed on a machine with two 2.40 GHz Intel Xeon processors with 512 KB memory each.

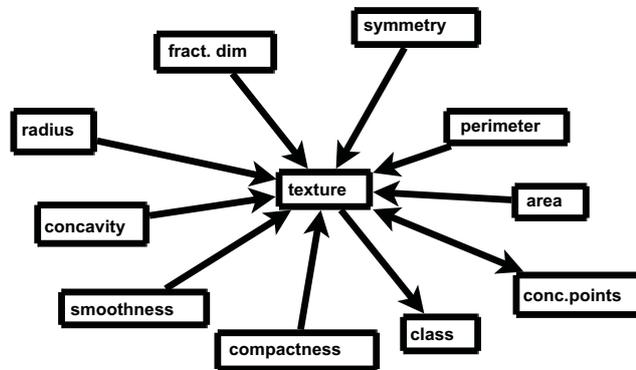


(a) Ferrara network structure using a local exhaustive search.

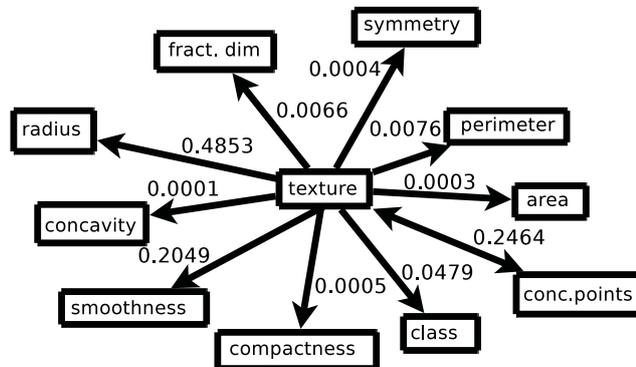


(b) Ferrara AN structure using SCAP and (c) Ferrara DN structure using a classification tree.

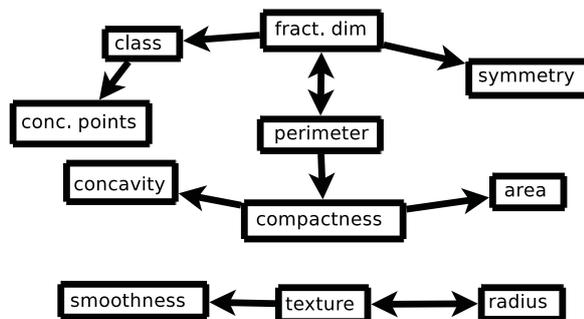
Figure 3.3: Comparison between the AN structure learned using a combination of SCAP and the BDe score and the DN structure obtained from a classification tree and on a small real data with discrete variables. The network structure is learned from a breast cancer data provided by the University of Ferrara. Based on a comparison with an exhaustive structure search (see Table 3.4), the AN structure, which is the same as the benchmark structure, is better than the DN structure, which has three added and three deleted arcs.



(a) Wisconsin structure using a local exhaustive search



(b) Wisconsin AN structure using SCAP and the BDe score



(c) Wisconsin DN structure using a classification tree

Figure 3.4: Comparison between the AN structure from a combination of SCAP and the BDe score and the DN structure obtained from a classification tree on a small real data with continuous variables. The network structures are learned from a breast cancer data from the University of Wisconsin. Based on a comparison with an exhaustive structure search (see Table 3.4), the AN structure, which has eight reversed arcs, is better than the DN structure, which has one reversed, nine added and nine deleted arcs.



(a) ALARM AN structure using SCAP and the BDe score



(b) ALARM DN structure using a classification tree



(c) ALARM network structure shared by both AN and DN structures above

Figure 3.5: AN network structure of the ALARM data from a combination of SCAP and the BDe score and DN network structure from a classification tree. The network structure was constructed in 20 seconds with SCAP and the BDe score, and in 135 seconds with a classification tree on a machine with two 2.40 GHz Intel Xeon processors with 512 KB memory each. The corresponding BN (structure not shown) was constructed in 998 seconds on a Macintosh II running LightSpeed Pascal version 2.0 with one missing and one extra arc using the K2 algorithm and an ordering of the variables (Cooper and Herskovits, 1991).

3.5 Joint probability approximation

In this section we would like to test the ability of the AN to mimic an underlying joint distribution. To this end, a synthetic data problem is generated from a known joint distribution that is represented by a network structure having three binary variables and one loop. The generated data is then used to learn an AN, a DN and a BN. Besides the ability to give a good approximation of a joint distribution, it is also important to check whether the local conditional distributions are consistent with the set of data sampled from a network structure when the samples are large enough.

3.5.1 Synthetic model

We would like to generate data from a joint distribution that is specified by the network structure in Figure 3.6. By using the product rule of probability, we can write that

$$p(ABC) = p(A|BC)p(BC), \quad (3.24)$$

$$= p(A|BC)p(B|C)p(C). \quad (3.25)$$

From the independencies in the network structure of Figure 3.6, we obtain $p(A|BC) = p(A|B)$ and using Bayes' rule we have $p(B|C) = (p(C|B)p(B))/p(C)$ which implies that

$$p(ABC) = p(A|B)p(C|B)p(B). \quad (3.26)$$

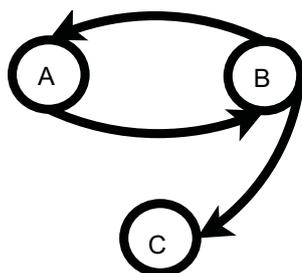


Figure 3.6: True benchmark network structure with three random variables and a loop used in the synthetic data generation.

3.5.2 Experimental methodology

For convenience, we are going to adopt the following notation: the probability of a single random variable will be represented by a lower case letter with a subindex indicating the

value of the random variable. A conditional probability (given one random variable only) will be represented by two lower case letter with subindices indicating the values of the random variables respectively. For example, $p(A = 0) = a_0$ and $p(A = 0|B = 0) = ab_{00}$. We have the following set of constraints:

$$a_0 + a_1 = 1, \quad (3.27)$$

$$b_0 + b_1 = 1, \quad (3.28)$$

$$c_0 + c_1 = 1, \quad (3.29)$$

$$ab_{00} + ab_{10} = 1 \quad \text{and} \quad ab_{01} + ab_{11} = 1, \quad (3.30)$$

$$ba_{00} + ba_{10} = 1 \quad \text{and} \quad ba_{01} + ba_{11} = 1, \quad (3.31)$$

$$cb_{00} + cb_{10} = 1 \quad \text{and} \quad cb_{01} + cb_{11} = 1. \quad (3.32)$$

Given the expression of the joint probability of the three random variables in Equation (3.26), the constraints (3.28), (3.29) and (3.30) will be needed in order to compute the joint distribution. We then have the following joint probability table:

<i>A</i>	<i>B</i>	<i>C</i>	$p(ABC)$
0	0	0	$ab_{00} \times cb_{00} \times b_0$
0	0	1	$ab_{00} \times cb_{10} \times b_0$
0	1	0	$ab_{01} \times cb_{01} \times b_1$
0	1	1	$ab_{01} \times cb_{11} \times b_1$
1	0	0	$ab_{10} \times cb_{00} \times b_0$
1	0	1	$ab_{10} \times cb_{10} \times b_0$
1	1	0	$ab_{11} \times cb_{01} \times b_1$
1	1	1	$ab_{11} \times cb_{11} \times b_1$

Table 3.6: Joint probability table of the three random variables. The following probability values are used to generate 1000 samples from the joint probability given in this table: $b_0 = 0.7$, $b_1 = 0.3$, $ab_{00} = 0.6$, $ab_{10} = 0.4$, $ab_{01} = 0.5$, $ab_{11} = 0.5$, $cb_{00} = 0.2$, $cb_{10} = 0.8$, $cb_{01} = 0.4$, $cb_{11} = 0.6$.

The following probability values are used to generate 1000 samples from the joint probability given in Table 3.6: $b_0 = 0.7$, $b_1 = 0.3$, $ab_{00} = 0.6$, $ab_{10} = 0.4$, $ab_{01} = 0.5$, $ab_{11} = 0.5$, $cb_{00} = 0.2$, $cb_{10} = 0.8$, $cb_{01} = 0.4$, $cb_{11} = 0.6$.

3.5.3 Experimental results

The different network structures are given in Figure 3.7 where the AN and the DN are able to learn the true network structure whereas the BN has one missing and one extra arcs. The BN graph is constructed using the greedy hill-climbing approach that was discussed

in Section 2.3.2 of Chapter 2. The hill-climbing algorithm was implemented using the BNT² (Bayes Net Toolbox) and SLP³ (Structure Learning Package) Matlab toolboxes.

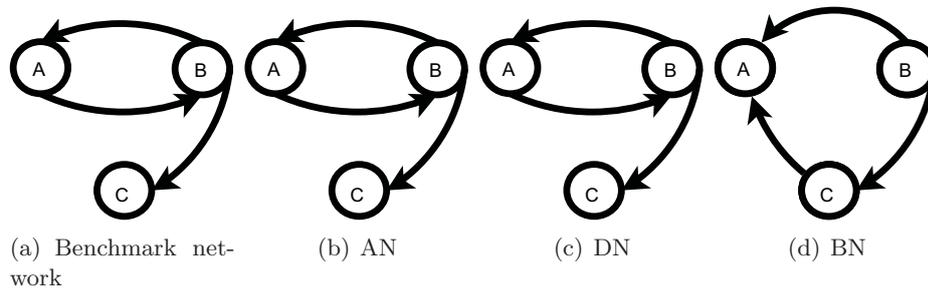


Figure 3.7: Approximation of a joint probability distribution by an AN, a DN and a BN. The benchmark network structure is generated from a known joint probability distribution specified in Table 3.6. The AN and the DN are able to learn the true network structure whereas the BN has one missing and one extra arcs. The BN graph is constructed using a hill-climbing approach that was discussed in Section 2.3.2 of Chapter 2. The AN was constructed in 0.23 seconds, the DN in 0.66 seconds and the BN in 1.10 seconds on a machine with two 2.40 GHz Intel Xeon processors with 512 KB memory each.

The conditional probabilities given in Table 3.7 confirm the fact that in the limit of a large dataset, the AN, the DN and the BN are consistent with the true joint probability.

	Benchmark model	AN	DN	BN
$p(A = 0 B = 0)$	0.60	0.73	0.60	0.73
$p(A = 1 B = 0)$	0.40	0.27	0.40	0.27
$p(A = 0 B = 1)$	0.50	0.50	0.50	0.50
$p(A = 1 B = 1)$	0.50	0.50	0.50	0.50
$p(B = 0 A = 0)$	0.74	0.74	0.58	-
$p(B = 1 A = 0)$	0.26	0.26	0.42	-
$p(B = 0 A = 1)$	0.65	0.50	0.52	-
$p(B = 1 A = 1)$	0.35	0.50	0.48	-
$p(C = 0 B = 0)$	0.20	0.25	0.20	0.25
$p(C = 1 B = 0)$	0.80	0.75	0.80	0.75
$p(C = 0 B = 1)$	0.40	0.40	0.40	0.40
$p(C = 1 B = 1)$	0.60	0.60	0.60	0.60

Table 3.7: Consistency of the conditional distributions of an AN, a DN and a BN. The true conditional probabilities are computed directly from the joint probability specified in Table 3.6. The conditional probabilities of the AN, the DN and the BN are obtained from the learned network structures in Figure 3.7. The values of $p(B|A)$ and $p(C|B)$ are not available for the BN due to its graph. The results confirm the fact that in the limit of a large dataset, the conditional probabilities of an AN, a DN and a BN are consistent with the true joint probability.

²<http://code.google.com/p/bnt/>

³<http://ofrancois.tuxfamily.org/slp.html>

3.6 Summary

A new class of graphical model called an Affinity Network (AN) and its structure learning algorithm using the soft-constraint affinity propagation (SCAP) algorithm combined with the BDe scoring function have been described. The SCAP algorithm is derived from AP which is an optimisation algorithm that is used to drive the AN structure search. SCAP is able to reconstruct the possible loops in the network structure contrary to AP because of its soft constraints. The parameter learning algorithm for an AN is similar to BN parameter learning.

Contrary to a BN, an AN and a DN do not necessarily represent a factorisation of the joint distribution of the data. In a DN, inconsistencies are likely to occur when there is no joint distribution from which the conditional probabilities may be obtained via the rules of probability. However when the sample size is large, inconsistencies will be rare given that the conditional distribution is learned from the same dataset, which is assumed to be generated from a single underlying distribution (Heckerman et al., 2000). Hence, when the data has many samples, an AN structure, similarly to a DN, can also be seen as a representation of the joint distribution of the data since the conditionals are directly learned from the data. A DN displays the dependencies between variables whereas an AN shows the closeness between pairs of variables with respect to the chosen scoring function.

The proposed approach has also been compared with the Dependency Network (DN) where a classification tree is used to construct the DN structure. The combination of SCAP and the BDe score to construct an AN always recovers the essential graph, that is the undirected graph, of the benchmark networks obtained from an exhaustive search whereas the classification tree adds more parents than required. However, when only the root of the tree is taken as the parent of the corresponding child variable, the classification tree performs better than the normal DN.

SCAP is a good alternative to heuristic or stochastic search algorithms since the SCAP algorithm is based on a message passing scheme. Furthermore, the algorithm has been proved to be efficient, fast and especially suited for huge data such as the alarm network where the proposed method is faster than a classification tree. Moreover, this alternative method to learn an AN structure provides extra information by quantifying the strength of each arc in the network.

The limitations of the AN reside in the choice of the SCAP parameters such as the preference values in the similarity matrix and the parameter \tilde{p} in the message-update equations. Different values of these parameters will produce different network structures

and so far an automatic procedure for setting their values is not available. Moreover, an AN structure only has one parent node for each variable node. The number of parents can be increased by taking the first m maximum values of the affinity function instead of the maximum value. However, the choice of m also remains an open problem that needs to be solved.

In this chapter the AN algorithm has been applied to sparse and static data. It will be seen in the following chapter that it can also be useful in the analysis of dense and dynamic time series data.

4 A coupled dynamical system example

The previous chapter introduced the Affinity Network (AN) graphical model to discover dependency relationships between sparse data taken from the medical domain. The AN has more preferable characteristics to the Bayesian Network (BN) and Dependency Network (DN) graphical models. In particular, ANs, like DNs, are easier to interpret than BNs due to their more flexible dependency relationships compared to the strong causal meaning of the arcs in BNs. Furthermore, the AN graph construction algorithm is superior to that of BNs and DNs since it is based on a data-driven message update of simple formulae, whereas BNs and DNs structure learning relies on heuristic search procedures. Besides the ability to display important dependency relationships, we also would like to exploit the capability of the AN for clustering.

In addition to the sparse and static medical data of the previous chapter, we now look at high-dimensional dynamical data from large many-body systems. Although the AN is not intrinsically designed for dynamical data due to the absence of a temporal factor, the quasi-stationary state of the system can be used as input to the AN algorithm. We shall consider exemplar systems that represent sensor arrays for real biosensor networks. The biomorphic design implies interaction and nonlinear oscillators as opposed to the

traditional design rules where the focus is on arrays of isolated and linear-response sensors. The interaction of nonlinear elements permits the emergence of a much richer range of complex behaviours of the array beyond that expected in a traditional design strategy. We are interested in the characterisation and detection of emergent collective behaviours in such arrays of interacting nonlinear systems. This aim can be achieved through the clustering of a priori unknown self-organising macrostates that are present in the array.

This chapter therefore discusses the behaviour of coupled nonlinear dynamical systems in a square lattice representing a microelectromechanical system (MEMS) device. Typically, the number of elements in such a device is very large and hence the analysis of their collective behaviour is complex. It will be seen later that two different types of elements will be coupled in this work which will increase the challenge of trying to predict their behaviour. To the best of our knowledge, only a few analytical results on the study of such systems are available and those studies only consider identical coupled elements. The logical state of the lattice can be viewed as a graphical model where nodes represent collective states and links display weak nonlinear interactions resulting in different macrostates. Therefore, the AN graphical model will be used as an inference model in Chapter 6 in order to characterise the emergent collective behaviour of the elements in the lattice. Chapter 5 will extend the discussion provided in the current chapter to a square lattice that is subject to noise. Before we apply the AN to the dense regular time series data produced by a synchronised system, the current chapter provides an analysis and numerical investigation into this dynamical system. We have chosen this system as it has relevance for future design of sensor arrays. We provide a detailed analysis since there has been very little work on this type of system in the literature.

In studies of the collective behaviour of oscillatory systems, when some elements share similar temporal characteristics such as the phase or amplitude, they are generally said to be synchronised. The synchronisation effects may be observed globally in all elements or locally in clusters of neighbouring elements. This chapter focuses on a population of elements composed of Duffing resonators coupled to van der Pol oscillators in a square lattice. The fundamental study of coupled nonlinear oscillators is important in understanding the emergent behaviour of complex dynamical systems and developing novel M/NEMS devices (Zalalutdinov et al., 2006a; Aubin et al., 2004; Mendelowitz et al., 2009). The analysis of a simpler case as the building block can help to gain insight into larger complicated systems (Bernardo et al., 1998; Rand and Wong, 2008b). Among the building blocks studied in the literature, the essential elements are either self-sustained oscillators (van der Pol oscillators) or dissipative oscillators (Duffing-type resonators) and the most intensively

studied cases are the coupled van der Pol and the coupled Duffing oscillators (Rand and Holmes, 1980; Barron and Sen, 2009; Kapitaniak, 1993). To the best of our knowledge, less has been done in a dynamic system consisting of a Duffing-type resonator coupled to a self-excited oscillator. Therefore, the current chapter starts with a discussion of one driven Duffing resonator coupled to a van der Pol oscillator, followed by the case where they are coupled together in a square lattice.

4.1 Single unit of a weakly driven Duffing resonator coupled to a van der Pol oscillator

In this section, the behaviour of a periodically driven Duffing resonator coupled to a van der Pol oscillator is discussed. A good understanding of the behaviour of those two coupled elements is necessary as they constitute the building block of a more complicated system such as a square lattice of Duffing resonators coupled to van der Pol oscillators which will be investigated in the next section. An analytical treatment of the two coupled elements is possible due to the small size of the system. This can be done through multiple time scales analysis (Nayfeh and Mook, 1995) which is derived from perturbation analysis. The linear stability of the system can also be analysed through the eigenvalues of the Jacobian matrix of the system. The study of the presence of chaos in the system can be conducted through bifurcation analysis or through the computation of its maximum Lyapunov exponent. These different aspects of the analysis will be discussed in the following.

4.1.1 Multiple time scales analysis of the driven Duffing resonator coupled to a van der Pol oscillator

The motion of the coupled dynamical system in dimensionless form is described by the following set of equations:

$$\ddot{x} + \gamma_x \dot{x} + x + \delta x^3 = \kappa(y - x) + f_d \cos\left(\frac{\omega_d}{\omega_x} \tau\right), \quad (4.1a)$$

$$\ddot{y} + \gamma_y(y^2 - 1)\dot{y} + \left(\frac{\omega_y}{\omega_x}\right)^2 y = \kappa(x - y), \quad (4.1b)$$

where, x_1, γ_x, ω_x and x_2, γ_y, ω_y are the displacement, damping coefficient and fundamental frequency of the Duffing resonator and the van der Pol oscillator respectively. f_d and ω_d are the amplitude and frequency of the external driving force. δ is the nonlinearity of the Duffing resonator and κ is the coupling stiffness between the two coupled elements. Usually, MEMS resonators and oscillators are working in a high Q scenario (Appendix B)

while they are driven into their resonances with a small force. However, if the Duffing resonator and van der Pol oscillator are working at different fundamental frequencies (the nonresonant case $\omega_x \neq \omega_y$), it can be proved that the dynamic systems are uncoupled and the time evolution of the amplitudes is the same as the classical driven Duffing resonator and van der Pol oscillator (Woafu et al., 1996). In this section, we consider the case where both the internal and external resonances coincide ($\omega_x = \omega_y$). As the coupling stiffness varies across a large range, the system of weak coupling ($\kappa \ll 1$) and strong coupling ($\kappa \gg 1$) are dealt with separately.

The resonant case of weak coupling

When the van der Pol oscillator is weakly connected to the Duffing resonator, Equation (4.1) is rewritten as follows:

$$\ddot{x} + x = \epsilon \left[\beta(y - x) + F \cos(\Omega\tau) - \mu_x \dot{x} - \alpha x^3 \right], \quad (4.2a)$$

$$\ddot{y} + y = \epsilon \left[\beta(x - y) - \mu_y (y^2 - 1) \dot{y} \right], \quad (4.2b)$$

where the coefficients μ_x , μ_y , α , β and F are such that $\gamma_x = \epsilon\mu_x$, $\gamma_y = \epsilon\mu_y$, $\delta = \epsilon\alpha$, $\kappa = \epsilon\beta$ and $f_d = \epsilon F$. $\epsilon \ll 1$ is a positive parameter and $\Omega = \frac{\omega_d}{\omega_x}$ is the ratio of the frequency of the excitation to the linear natural frequency of the system.

The response of the system for small dissipation and driving can be calculated from Equation (4.2) using the standard multiple time scales method (Nayfeh and Mook, 1995). In Nayfeh and Mook (1995), the basic method has been described to characterise the behaviour of a single driven Duffing resonator and a single van der Pol oscillator. In this work, the derivations are related to a driven Duffing resonator coupled to a van der Pol oscillator, which is novel and has appeared in the following two journal papers: Wei et al. (2010); Randrianandrasana et al. (2010). Since the calculation involved in the analysis is intricate, the following results are an abridged version of the main derivations that are given in Appendix C. In general, we consider $x(\tau)$ and $y(\tau)$ in the form

$$x(\tau, \epsilon) = x_0(\tau_0, \tau_1) + \epsilon x_1(\tau_0, \tau_1), \quad (4.3a)$$

$$y(\tau, \epsilon) = y_0(\tau_0, \tau_1) + \epsilon y_1(\tau_0, \tau_1), \quad (4.3b)$$

where $\tau_0 = \tau$ is a fast scale and $\tau_1 = \epsilon\tau$ is a slow scale characterising the modulation in the amplitudes and phases caused by nonlinearity, coupling and resonances. The time

derivatives, in which the single overdot is $d/d\tau$ and the double overdot is $d^2/d\tau^2$, become

$$\begin{aligned}\frac{d}{d\tau} &= D_0 + \epsilon D_1 + \dots, \\ \frac{d^2}{d\tau^2} &= D_0^2 + 2\epsilon D_0 D_1 + \dots,\end{aligned}$$

where $D_n = \partial/\partial\tau_n$.

After eliminating the secular terms, that is, the slow varying terms, in the multiple time scales analysis of Equation (4.2), we obtain the following set of equations:

$$2\frac{dA_x}{d\tau_1} + \mu_x A_x - 3i\alpha|A_x|^2 A_x + \beta i(A_y - A_x) - \frac{i}{2} F e^{i\sigma\tau_1} = 0, \quad (4.5a)$$

$$\frac{dA_y}{d\tau_1} + i\frac{\beta}{2}(A_x - A_y) + \frac{\mu_y}{2}(|A_y|^2 - 1)A_y = 0, \quad (4.5b)$$

where the quantities $A_x(\tau_1)$ and $A_y(\tau_1)$ are arbitrary, complex functions which represent the amplitudes of the Duffing resonator and the van der Pol oscillator respectively. A detuning parameter σ has been introduced since the case $\Omega \approx 1$ is considered. σ is defined by

$$\Omega = 1 + \epsilon\sigma. \quad (4.6)$$

Then

$$\Omega\tau_0 = (1 + \epsilon\sigma)\tau_0 = \tau_0 + \epsilon\sigma\tau_0 = \tau_0 + \sigma\tau_1. \quad (4.7)$$

Expressing $A_x(\tau_1, \epsilon)$ and $A_y(\tau_1, \epsilon)$ in the polar forms $A_x(\tau_1, \epsilon) = a_x(\tau_1, \epsilon)e^{i\theta_x(\tau_1, \epsilon)}$ and $A_y(\tau_1, \epsilon) = a_y(\tau_1, \epsilon)e^{i\theta_y(\tau_1, \epsilon)}$, where $a_x(\tau_1, \epsilon)$ and $\theta_x(\tau_1, \epsilon)$, respectively, $a_y(\tau_1, \epsilon)$ and $\theta_y(\tau_1, \epsilon)$ are the amplitudes and the phases of the fundamental solutions. Separating real and imaginary parts in (4.5) yields

$$2\frac{da_x}{d\tau_1} = -\mu_x a_x + \beta a_y \sin \Gamma_y - \frac{F}{2} \sin \Gamma_x, \quad (4.8a)$$

$$2\frac{da_y}{d\tau_1} = -\mu_y (a_y^2 - 1)a_y - \beta a_x \sin \Gamma_y, \quad (4.8b)$$

$$2a_x \frac{d\Gamma_x}{d\tau_1} = -(\beta - 2\sigma)a_x + 3\alpha a_x^3 - \beta a_y \cos \Gamma_y - \frac{F}{2} \cos \Gamma_x, \quad (4.8c)$$

$$2a_y \frac{d\Gamma_y}{d\tau_1} = (\beta - 2\sigma)a_y - \beta a_x \cos \Gamma_y - 2a_y \frac{d\Gamma_x}{d\tau_1}, \quad (4.8d)$$

where the following transformations have been introduced:

$$\Gamma_x = \theta_x - \sigma\tau_1, \quad (4.9)$$

and

$$\Gamma_y = \theta_y - \theta_x. \quad (4.10)$$

The amplitude response of the steady-state motions of Equation (4.5) are given by the following set of algebraic equations,

$$\left(a_x(\beta - 2\sigma) + 3a_x^3\alpha - \frac{a_y^2(\beta - 2\sigma)}{a_x} \right)^2 + \left(a_x\mu_x - \frac{a_y^2\mu_y}{a_x}(1 - a_y^2) \right)^2 = \frac{1}{4}F^2 \quad (4.11a)$$

$$(\beta - 2\sigma)^2 a_y^2 + \mu_y^2 a_y^2 (1 - a_y^2)^2 = a_x^2 \beta^2 \quad (4.11b)$$

The corresponding frequency-response curves, for the variation of the amplitude A with the frequency Ω , of the Duffing resonator and van der Pol oscillator are shown in Figure 4.1, where the coupling strength $\kappa = 0.002$.

If the coupling stiffness κ is not the same order of magnitude as the other parameters, the above analysis does not hold, so a modified analysis considering the strong coupling case is given in the next section.

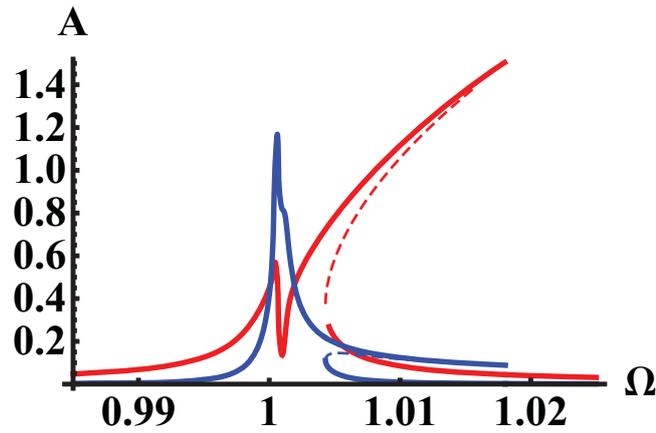


Figure 4.1: Frequency-response curves of the driven Duffing resonator (red curve) weakly coupled to a van der Pol oscillator (blue curve). The unstable branches are drawn in dashed lines. The system parameters are $\gamma_x = \gamma_y = 0.001$, $\delta = 0.005$, $\kappa = 0.002$, $f_d = 0.003$.

The resonant case of strong coupling

The case of strong coupling can be dealt with by using the transformation $\epsilon = \frac{1}{2\kappa+1}$. When the coupling strength κ is bigger than 4.5 then ϵ is less than 0.1 and it can be assumed that the other system parameters are all small as defined in the previous section. By adding

and subtracting (4.2a) and (4.2b), we obtain:

$$\ddot{x} + \ddot{y} + \epsilon\mu_x\dot{x} + \epsilon\mu_y(y^2 - 1)\dot{y} + x + y + \epsilon\alpha x^3 = \epsilon F \cos(\Omega\tau), \quad (4.12a)$$

$$\ddot{x} - \ddot{y} + \epsilon\mu_x\dot{x} - \epsilon\mu_y(y^2 - 1)\dot{y} + x - y + \epsilon\alpha x^3 = 2\epsilon\beta(y - x) + \epsilon F \cos(\Omega\tau), \quad (4.12b)$$

or,

$$\ddot{x} + \ddot{y} + \epsilon\mu_x\dot{x} + \epsilon\mu_y(y^2 - 1)\dot{y} + x + y + \epsilon\alpha x^3 = \epsilon F \cos(\Omega\tau), \quad (4.13a)$$

$$\ddot{x} - \ddot{y} + \epsilon\mu_x\dot{x} - \epsilon\mu_y(y^2 - 1)\dot{y} + (x - y)(1 + 2\epsilon\beta) + \epsilon\alpha x^3 = \epsilon F \cos(\Omega\tau). \quad (4.13b)$$

Since $(1 + 2\epsilon\beta) = (1 + 2\kappa) = \frac{1}{\epsilon}$, multiplying (4.13b) by ϵ , Equation (4.1) can then be written as

$$\ddot{x} + \ddot{y} + \epsilon\mu_x\dot{x} + \epsilon\mu_y(y^2 - 1)\dot{y} + x + y + \epsilon\alpha x^3 = \epsilon F \cos(\Omega\tau), \quad (4.14a)$$

$$\epsilon(\ddot{x} - \ddot{y}) + \epsilon^2\mu_x\dot{x} - \epsilon^2\mu_y(y^2 - 1)\dot{y} + x - y + \epsilon^2\alpha x^3 = \epsilon^2 F \cos(\Omega\tau). \quad (4.14b)$$

The approximate solutions of Equation (4.14) can be obtained by following the same steps as in the previous section and are given by

$$D_0^2(x_0 + y_0) + x_0 + y_0 = 0, \quad (4.15a)$$

$$x_0 - y_0 = 0, \quad (4.15b)$$

for the zero-order approximation. Equation (4.15) means that the Duffing resonator and the van der Pol oscillator are completely synchronised. The solutions of Equation (4.15) can be written as

$$x_0(\tau_0, \epsilon) = y_0(\tau_0, \epsilon) = A(\tau_1)e^{i\tau_0} + \bar{A}(\tau_1)e^{-i\tau_0}. \quad (4.16)$$

The first-order approximation gives

$$D_0^2(x_1 + y_1) + x_1 + y_1 = -2D_1D_0(x_0 + y_0) - \mu_x D_0 x_0 - \mu_y (y_0^2 - 1)D_0 y_0 - \alpha x_0^3 + F \cos(\Omega\tau), \quad (4.17a)$$

$$x_1 - y_1 = 0. \quad (4.17b)$$

This also means that the Duffing resonator and the van der Pol oscillator are still completely synchronised in the first-order approximation when they are strongly coupled. By substituting the expressions of x_0 and y_0 from Equation (4.16) into the right hand side of

Equation (4.17a) and by imposing secular conditions, the following equation is obtained:

$$4\frac{dA}{d\tau_1} + (\mu_x - \mu_y)A + \mu_y A|A|^2 - 3i\alpha A|A|^2 + i\frac{1}{2}F e^{i\sigma\tau_1} = 0. \quad (4.18)$$

Expressing A in polar form and substituting the resulting expression into Equation (4.18), the following set of first-order differential equations can be obtained for the amplitude and phase of the system:

$$4\frac{da}{d\tau_1} = (\mu_y - \mu_x)a - \mu_y a^3 - \frac{1}{2}F \sin \Gamma_x, \quad (4.19a)$$

$$4a\frac{d\Gamma_x}{d\tau_1} = -4a\sigma + 3\alpha a^3 - \frac{1}{2}F \cos \Gamma_x, \quad (4.19b)$$

where $a = a_x = a_y$. For the steady state solutions, the following algebraic equation describing the frequency and the amplitude response of the system is obtained:

$$a^2(-4\sigma + 3\alpha a^2)^2 + a^2(\mu_x - \mu_y + \mu_y a^2)^2 = \frac{1}{4}F^2. \quad (4.20)$$

The corresponding frequency-response curve of the system is shown in Figure 4.2, where the coupling strength $\kappa = 8$. The shark's fin shape of the curve is typical of a single driven Duffing resonator.

After eliminating the secular terms in the two time scales analysis of Equations (4.2a)

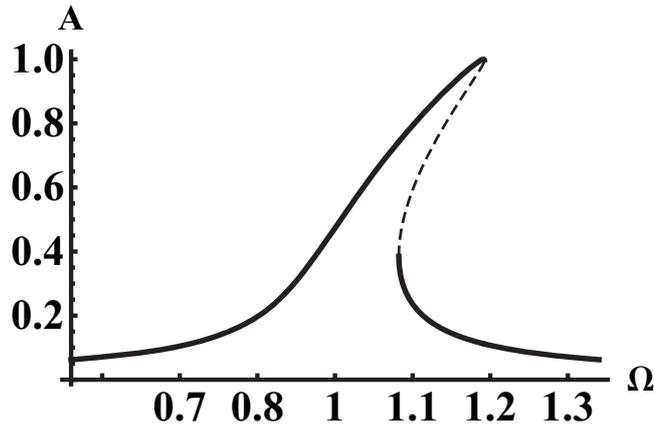


Figure 4.2: Frequency-response curve of the driven Duffing resonator strongly coupled to a van der Pol oscillator. The unstable branch is drawn in dashed line. The system parameters are $\gamma_x = 0.1$, $\gamma_y = 0.05$, $\delta = 0.3$, $\kappa = 8$, $f_d = 0.2$.

and (4.2b), the second-order differential equation can be solved and x_1 and y_1 are obtained as follows:

$$x_1 = y_1 = \frac{1}{16}(\alpha - \mu_x i)A^3 e^{3i\tau_0} + cc. \quad (4.21)$$

Therefore, the general solution of the strongly coupled Duffing resonator and van der Pol oscillator in the first order approximation is given by

$$x = y = \Re(x_0 + \epsilon x_1) = \Re\left(ae^{i(\Omega\tau_0 + \phi)} + \frac{1}{16}\epsilon(\alpha - \mu_y i)a^3 e^{3i(\Omega\tau_0 + \phi)} + cc\right), \quad (4.22)$$

which shows that in the case of a strong coupling, the van der Pol oscillator is completely synchronised to the Duffing resonator and their oscillations are also locked to the external driving force by a phase lag ϕ . This is confirmed numerically, for a coupling strength $\kappa = 8$, by the plots in Figure 4.3: Figures 4.3(a) and 4.3(b) show the early complicated behaviour, also called *initial transient*, of the Duffing resonator and van der Pol oscillator during the first hundred periods of oscillations. In these phase portraits, the trajectories of the Duffing resonator and the van der Pol oscillator are plotted in *phase space*, that is the abscissa corresponds to the displacement and the ordinate represents the velocity of each element. After an initial transient of 2000 periods, the Duffing resonator and the van der Pol oscillator become synchronised as shown in Figure 4.3(c).

4.1.2 Stability analysis

The frequency-response curves obtained in the previous section showed the steady-state approximate solutions of the coupled system given by the multiple time scales analysis. In this section the stability of those *equilibrium points*, also called *fixed points*, is investigated. The stability of a fixed point characterises whether the equilibrium point will stay close to its initial condition or will be repelled away from it when it is subject to a small perturbation. The linear stability of fixed points of dynamical systems is determined by the sign of the real part of the eigenvalues of the Jacobian matrix, also called *eigenvalues of equilibrium*. The Jacobian matrix J of the coupled system of Duffing resonator and van der Pol oscillator is the matrix of partial derivatives of the right-hand side of Equations (4.8) with respect to the state variables a_x , a_y , γ_x and γ_y :

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial a_x} & \frac{\partial f_1}{\partial a_y} & \frac{\partial f_1}{\partial \Gamma_x} & \frac{\partial f_1}{\partial \Gamma_y} \\ \frac{\partial f_2}{\partial a_x} & \frac{\partial f_2}{\partial a_y} & \frac{\partial f_2}{\partial \Gamma_x} & \frac{\partial f_2}{\partial \Gamma_y} \\ \frac{\partial f_3}{\partial a_x} & \frac{\partial f_3}{\partial a_y} & \frac{\partial f_3}{\partial \Gamma_x} & \frac{\partial f_3}{\partial \Gamma_y} \\ \frac{\partial f_4}{\partial a_x} & \frac{\partial f_4}{\partial a_y} & \frac{\partial f_4}{\partial \Gamma_x} & \frac{\partial f_4}{\partial \Gamma_y} \end{pmatrix},$$

where the functions f_1 to f_4 correspond to the right-hand side of Equations (4.8). Equations (4.8c) and (4.8d) are divided by a_x and a_y respectively in order to get the expressions

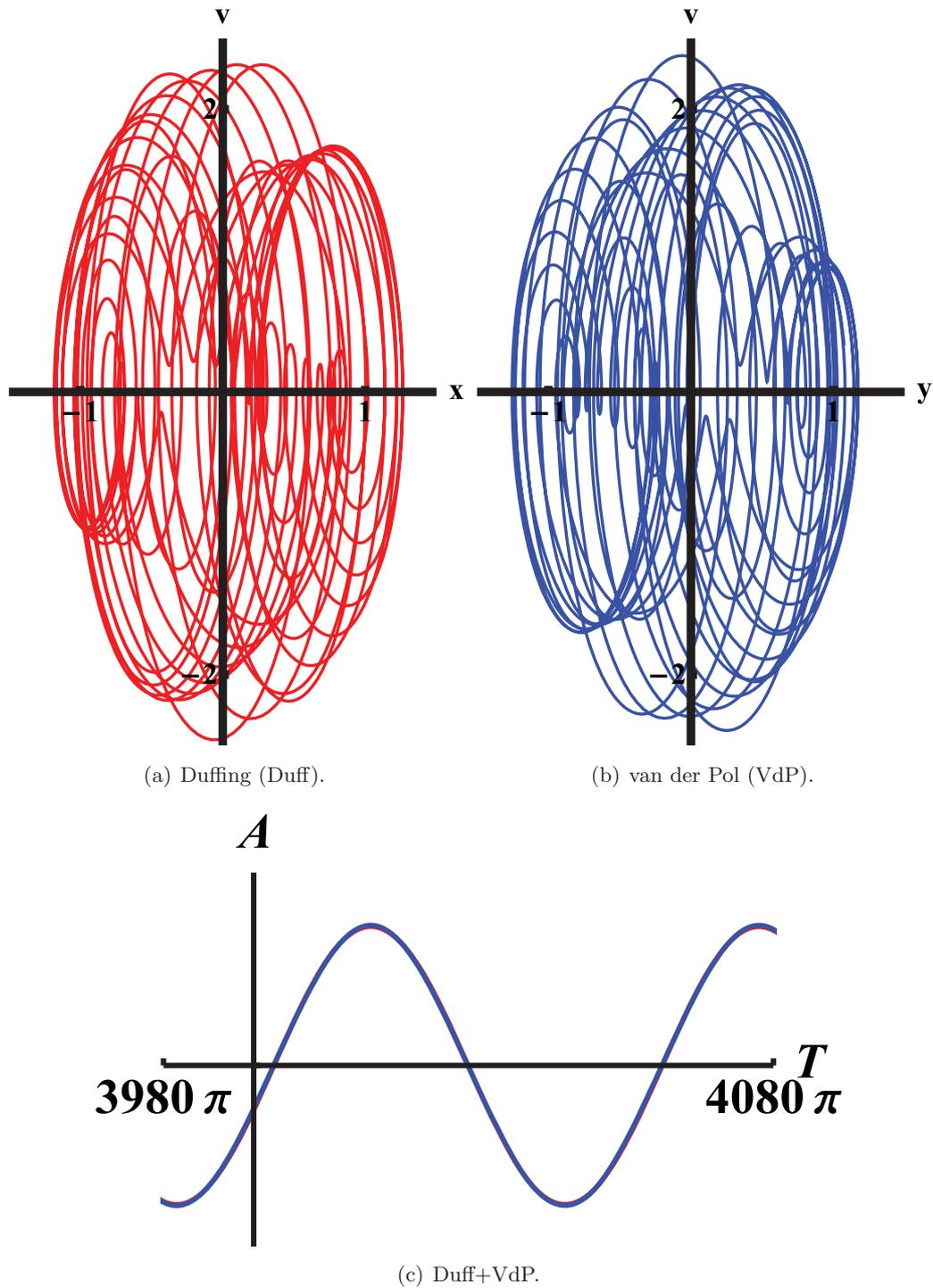


Figure 4.3: Phase portraits and time series of a driven Duffing resonator (red) strongly coupled to a van der Pol oscillator (blue). The system parameters are $\gamma_x = 0.1$, $\gamma_y = 0.05$, $\delta = 0.3$, $\kappa = 8$, $f_d = 0.2$. Both elements show an early complicated behaviour, also called *initial transient*, during the first hundred periods of oscillations (4.3(a) and 4.3(b)). After an initial transient of 2000 periods, the Duffing resonator and the van der Pol oscillator become synchronised as shown by the time series in 4.3(c) where the red and blue curves are superimposed.

for $\frac{d\Gamma_x}{d\tau_1}$ and $\frac{d\Gamma_y}{d\tau_1}$. The expression of the resulting Jacobian matrix is then given by

$$J = \frac{1}{2} \begin{pmatrix} -\mu_x & \beta \sin \Gamma_y & -\frac{1}{2}F \cos \Gamma_x & \beta a_y \cos \Gamma_y \\ -\beta \sin \Gamma_y & \mu_y (1 - 3a_y^2) & 0 & -a_x \beta \cos \Gamma_y \\ J_{31} & -\frac{\beta}{a_x} \cos \Gamma_y & \frac{F}{2a_x} \sin \Gamma_x & \frac{a_y \beta}{a_x} \sin \Gamma_y \\ J_{41} & \frac{\beta}{a_x} \left(1 + \frac{a_y^2}{a_x^2}\right) \cos \Gamma_y & -\frac{F}{2a_x} \sin \Gamma_x & \beta \frac{a_x^2 - a_y^2}{a_x a_y} \sin \Gamma_y \end{pmatrix},$$

where $J_{31} = 6\alpha a_x + \frac{a_y \beta}{a_x^2} \cos \Gamma_y + \frac{F}{2a_x^2} \cos \Gamma_x$ and $J_{41} = -\frac{\beta}{a_y} \left(1 + \frac{a_y^2}{a_x^2}\right) \cos \Gamma_y - 6\alpha a_x - \frac{F}{2a_x^2} \cos \Gamma_x$. The Equations (4.8a) to (4.8d) are equal to zero for fixed points of the coupled system. This enables the computation of the trigonometric quantities in the Jacobian matrix which are given by

$$\begin{aligned} \sin \Gamma_x &= \frac{2}{F} \left(-a_x \mu_x + \frac{a_y^2 \mu_y (1 - a_y^2)}{a_x} \right), \\ \cos \Gamma_x &= \frac{2}{F} \left(a_x (\beta - 2\sigma) + 3\alpha a_x^3 - \frac{a_y^2 (\beta - 2\sigma)}{a_x} \right), \\ \sin \Gamma_y &= \frac{\mu_y a_y (1 - a_y^2)}{a_x \beta}, \\ \cos \Gamma_y &= \frac{a_y (\beta - 2\sigma)}{a_x \beta}. \end{aligned}$$

The derivatives in J are evaluated at a pair of equilibrium points (a_{e_1}, a_{e_2}) which corresponds to the pair of points in the frequency-response curves of Figure 4.1 for the Duffing resonator and van der Pol oscillator respectively. The eigenvalues of J determine linear stability properties of an equilibrium. An equilibrium is asymptotically stable if all eigenvalues have negative real parts whereas it is unstable if at least one eigenvalue has positive real part. Figure 4.1 shows the stable (solid lines) and unstable (dashed lines) branches of the coupled system. As can be expected, the lower branches, in red and blue colours, of the hysteresis for the Duffing resonator and the van der Pol oscillator respectively are unstable for the system parameters that have been chosen. Similarly, in the strong coupling case, Figure 4.2 shows that the middle branch of the frequency-response curve of the coupled system is unstable. The stability of the system is further investigated numerically in the next section where an analysis of the time series is carried out using some tools to study chaos in nonlinear dynamical systems.

4.1.3 Asymptotic dynamics of the coupled system

The Poincaré section is a useful way of analysing the dynamics of a chaotic motion. As the name suggests, it consists in taking a section of the trajectory at a fixed time t rather than considering the whole trajectory for all times, which gives a continuous curve. For a driven oscillator, for example, the time t corresponds to the period of the driving force, namely $t = 2\pi/\omega$, $t = 4\pi/\omega$, $t = 6\pi/\omega$, etc, where ω is the frequency of the driving force. Consequently, for a periodic motion the Poincaré section is a single point, when the period has doubled it consists of two points, and so on. A cloud of points would then be expected for a chaotic motion in the sense that no periodic orbit is present.

It is also interesting to see the effect of changing one system parameter on its trajectory. To this end, the parameter of interest can be varied in small steps and the Poincaré sections can be merged together. The resulting plot is called a *bifurcation diagram*. It should be noted that after each step, the last location and velocity of the trajectory are used as new initial conditions for the next step.

Besides the Poincaré section, which provides a reduced visual representation of a trajectory, chaotic motions of dynamical systems can also be quantified by a measure of the divergence of nearby trajectories called *Lyapunov exponents*. The idea is that a set of initial conditions, which characterise the nearby trajectories, will be distorted by the effect of the dynamics for sufficiently short time scales. The largest Lyapunov exponent measures the asymptotic rate of expansion of the largest axis upon which the set of initial conditions in the phase space is distorted. The largest axis corresponds to the most unstable direction of the flow. Formally, the largest Lyapunov exponent for an initial point x_0 that is under a small perturbation u_0 is given by

$$\lambda(x_0, u_0) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{\|u_t\|}{\|u_0\|} = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \|D_{x_0} f^t(x_0) \cdot u_0\|, \quad (4.23)$$

where $u_t \equiv f^t(x_0 + u_0) - f^t(x_0) \cdot u_0$ denotes the perturbation of the images of the nearby points x_0 and $x_0 + u_0$ under the flow f^t after a time t . and $\|u\|$ represents the length of a vector u . The last term in (4.23) is obtained by linearising f^t . If $\lambda(x, u) > 0$, then an exponential divergence of nearby orbits is observed which means that chaos is present, otherwise the system is stable. The limit in (4.23) exists and is finite for almost all points x_0 and for almost all tangent vectors u_0 (Oseledets, 1968).

Figure 4.4 shows the bifurcation diagrams and the maximum Lyapunov exponents of the strongly coupled system ($\kappa = 5$) with varying frequency Ω of the external driving field where an abrupt transition to chaos can be observed. 4.4(a) and 4.4(b)

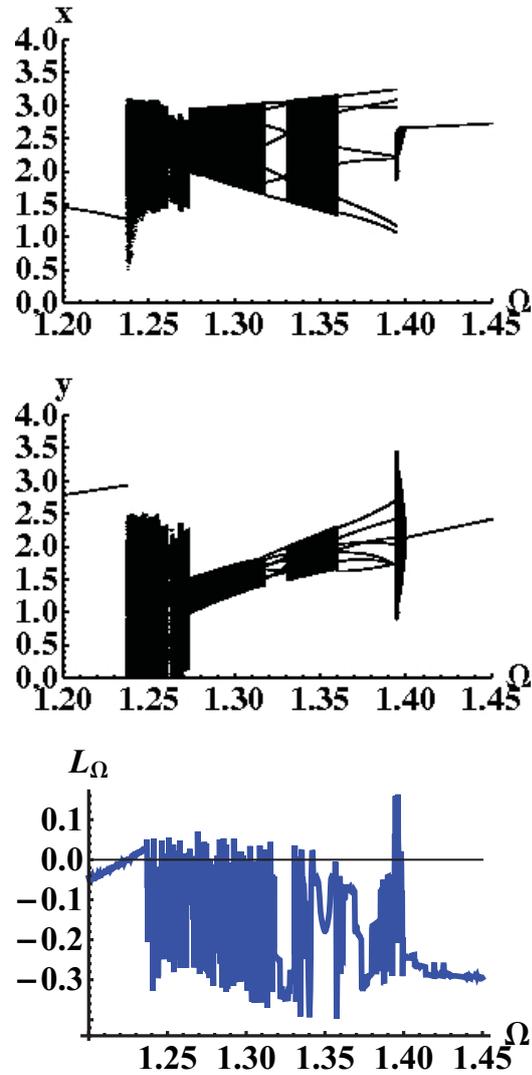


Figure 4.4: Bifurcation diagrams and maximum Lyapunov exponents of the driven Duffing resonator strongly coupled to a van der Pol oscillator with varying frequency Ω of the driving force. An abrupt transition to chaos can be observed. The system parameters are $\gamma_x = \gamma_y = 0.01$, $\delta = 1$, $\kappa = 5$, $f_d = 4.5$.

The frequency-response curves obtained previously showed the steady-state approximate solutions of the coupled systems given by the multiple time scales analysis. For the weak coupling case, the system can be driven into multiple states as shown in Figure 4.1. One can then expect that in a lattice of such units, more complicated behaviour can appear and it is therefore interesting to investigate their collective behaviour. The system also displays interesting phenomena such as an anti-resonance at $\Omega \approx 1$ and a shoulder in the curve which is due to a stagnant response of the van der Pol oscillator. For the strong coupling case, the system behaves like a driven Duffing resonator as shown in Figure 4.2. Now that some insights about the behaviour of the two coupled elements have been gained, a new topology in which the Duffing resonators and van der Pol oscillators are coupled alternatively in a square lattice will be considered in the next section.

4.2 Collective behaviour in a square lattice of driven Duffing resonators coupled to van der Pol oscillators

In this section, the collective behaviour of alternating Duffing resonators and van der Pol oscillators elastically coupled in a square lattice is numerically studied with typical MEMS parameters. The architecture of the lattice of resonators and oscillators along with some characterisations of synchronisation will be described first. Then the effect of some parameters on the behaviour of the system will be discussed followed by different configurations and sizes of the lattice.

4.2.1 Experimental setup

The architecture of the network of oscillators consists of a square lattice of alternating Duffing resonators and van der Pol oscillators in a checkerboard configuration as shown in Figure 4.5. The dimensionless equations governing the coupled systems are:

$$\ddot{x}_{i,j} + \gamma \dot{x}_{i,j} + x_{i,j} + \delta x_{i,j}^3 = \kappa(x_{i,j+1} + x_{i,j-1} + x_{i-1,j} + x_{i+1,j} - 4x_{i,j}), \quad (4.24)$$

$$\ddot{x}_{i,j} + \gamma(x_{i,j}^2 - 1)\dot{x}_{i,j} + x_{i,j} = \kappa(x_{i,j+1} + x_{i,j-1} + x_{i-1,j} + x_{i+1,j} - 4x_{i,j}), \quad (4.25)$$

where γ is the damping of the Duffing resonator and the van der Pol oscillator, δ is the cubic nonlinear spring constant of the Duffing resonator and κ is the coupling strength between the resonators and oscillators. The element $x_{i,j}$ represents the displacement of the resonator or oscillator located at the i -th row and j -th column in the square lattice.

Since the array of oscillators will act as a sensing device, it is usually subject to some

external signal that can be represented as a driving force or energy for the array. The equations for the driven Duffing resonator and van der Pol oscillator respectively are therefore given below:

$$\ddot{x} + \gamma\dot{x} + x + \delta x^3 = F_d \cos(\Omega\tau), \quad (4.26)$$

and

$$\ddot{x} + \gamma(x^2 - 1)\dot{x} + x = F_d \cos(\Omega\tau), \quad (4.27)$$

where F_d is the normalised driving force.

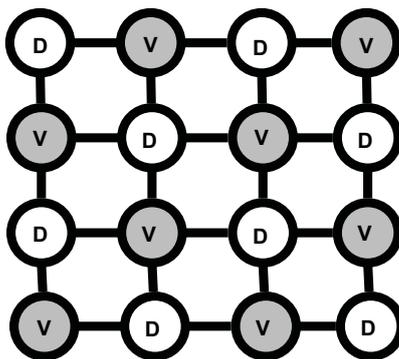


Figure 4.5: Square lattice of alternating Duffing resonators and van der Pol oscillators.

In this section, the synchronous behaviour of the resonators and oscillators in the square lattice will be characterised in terms of the global and local synchronisations of their amplitude and instantaneous phase. The instantaneous phase $\phi(t)$ of a real-valued signal $s(t)$ is given by

$$\phi(t) = \arg(Z(t)), \quad (4.28)$$

where $Z(t)$ is a complex-valued function which is called the analytic representation of $s(t)$ and is obtained as follows:

$$Z(t) = s(t) + jH[s(t)]. \quad (4.29)$$

H represents the Hilbert transform which is defined as

$$H[s(t)] = s(t) * \frac{1}{\pi t}, \quad (4.30)$$

where $*$ denotes the convolution operation.

In the first case, the resonators and oscillators will mainly be considered as one group whereas the elements in the array can be gathered in different subgroups in the second case. The global coherence of the array will be measured by a synchronisation index based

on the normalised entropy (Tass et al., 1998):

$$\tilde{\rho} = \frac{S_{\max} - S}{S_{\max}}, \quad (4.31)$$

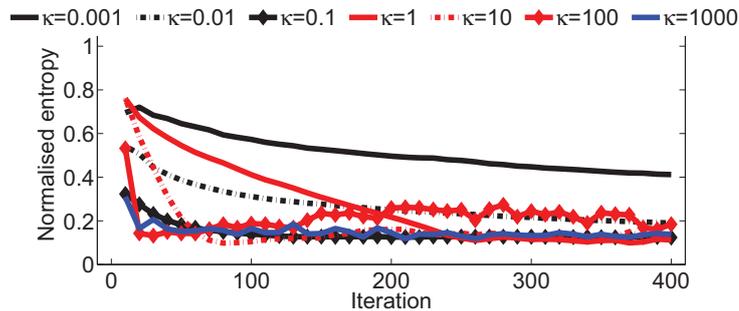
where $S = -\sum_{k=1}^N \rho_k \ln \rho_k$ is the entropy of the distribution of the instantaneous phase or the amplitude of the oscillators, which is given by their respective histograms from which the probabilities ρ_k are computed, and $S_{\max} = \ln N_h$, where N_h is the number of bins of the histogram. $0 \leq \tilde{\rho} \leq 1$ and $\tilde{\rho} = 0$ means that the array is not coherent or in total disorder, whereas $\tilde{\rho} = 1$ indicates that all the oscillators in the array are perfectly synchronised. Besides full synchronisation it is also interesting to consider another aspect of synchronisation namely the cluster synchronisation where the oscillators involved in the same cluster have identical temporal dynamics and completely synchronise between each other. Unlike the global synchronisation case, no cluster synchronisation index is available yet. Therefore, only a visual inspection of the displacement and phases of the elements in the lattice will be carried out. To this end, phase images and normalised polar plots of the elements of the square lattice will be produced. The phase and normalised polar plots will be complementary as the phase images will help to locate the resonators and oscillators within the array while the normalised polar plots will display both the amplitudes and phases of the elements. In the following, the colour scale in the phase images lies between $-\pi$ and π and the polar plots are normalised in amplitude which means that the maximum amplitude corresponds to a distance of 1 in the unit circle.

4.2.2 Parameter effects

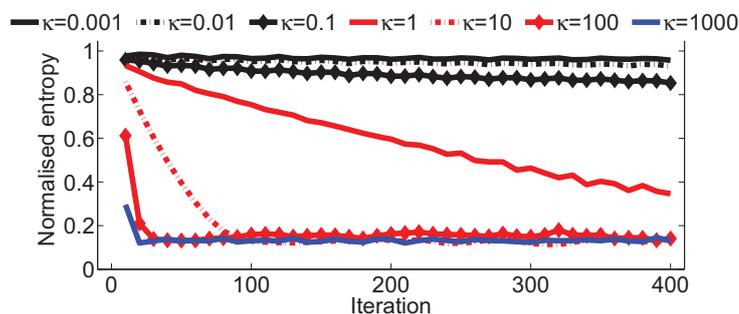
In this section we investigate the effect of different parameters which have an impact on the behaviour of the lattice of oscillators and resonators. These include the normalised coupling strength κ between the oscillators, the amplitude F_d of the driving force that is injected from one side of the lattice and finally the quality factor Q of each element of the lattice. It should be noted that the quality factor Q of the oscillators and resonators is directly related to their damping γ through the expression $\gamma = 1/Q$. It is also important to stress the fact that only the Duffing resonators are driven since they absorb energy whereas the van der Pol oscillators are self-sustained and that the system is not sensitive to initial conditions (plots not shown).

In the following, the parameters are given in their dimensionless values. In the context of MEMS devices, the coupling strength κ ranges from 0.001 to 1000. The amplitude F_d of the driving force is assumed to be in the range 0.3 to 30. The quality factor Q lies

within the interval $[10, 10000]$, where the lowest and biggest Q values correspond to the devices being in air and in vacuum respectively.



(a) Phase.



(b) Amplitude.

Figure 4.6: Coupling strength effect on the global coherence of the square lattice. The phase coherence does not seem to have a direct relation with the coupling strength, although for a very high coupling ($\kappa = 1000$) the phases of the elements of the lattice tend to converge faster. On the other hand, the amplitude coherence clearly depends on the coupling strength value: the higher the coupling strength, the quicker the convergence into a stable state. It should be noted that the normalised entropy is low when the coupling strength is high because the Duffing resonators and the van der Pol oscillators evolve into two different clusters as shown in Figure 4.7. The normalised entropy is maximum only when all the elements have the same value or in other words when there is only one cluster.

The coupling strength between the elements of the array has a counter-intuitive effect on the global coherence of the amplitudes of the resonators and oscillators. The lower the coupling the more coherent is the system as shown in Figure 4.6. This is due to the fact that the value of the normalised entropy is maximum only when all the elements have the same value, or in other words, when there is only one cluster. However, when the coupling strength is high a synchronisation across rows is observed as shown in Figure 4.7(b). The value of the coupling strength has no clear relation with the global phase coherence of the elements although the global phase coherence of the system tends to converge quicker with higher coupling values.

More detailed information regarding the cluster synchronisation of the elements in the lattice are given by the phase and normalised polar plots in Figure 4.7. It can clearly be

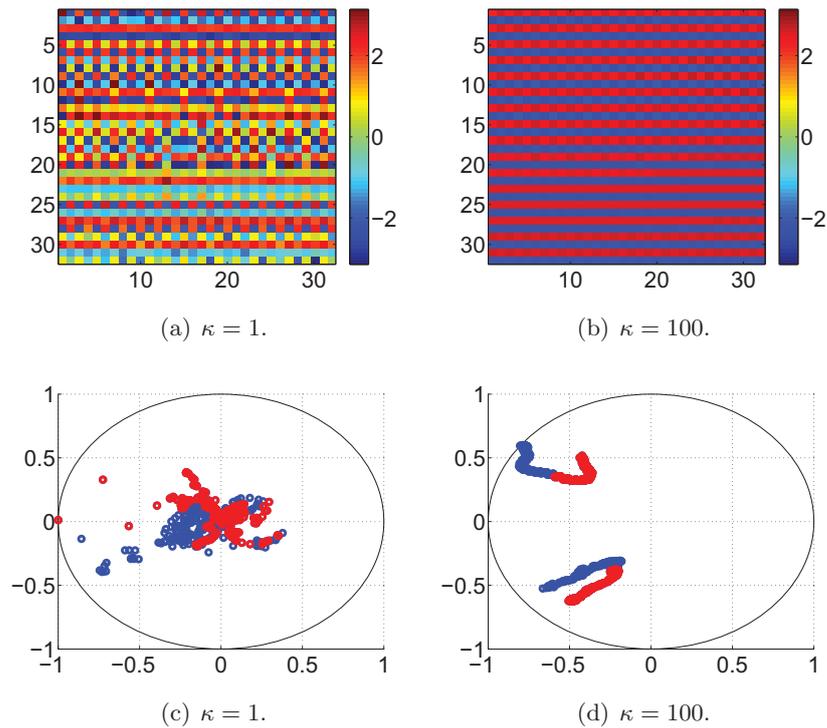


Figure 4.7: Phase and normalised polar plots of a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The figures are plotted at $t = 3600$ with different coupling strength values. At low coupling ($\kappa = 1$) both the phases and amplitudes of the resonators and oscillators seem to be random. For high coupling ($\kappa = 100$) a synchronisation across rows is observed as shown in Figure 4.7(b). This is an example of distributed cluster synchronisation where there is a division into quasi-independent sublattices.

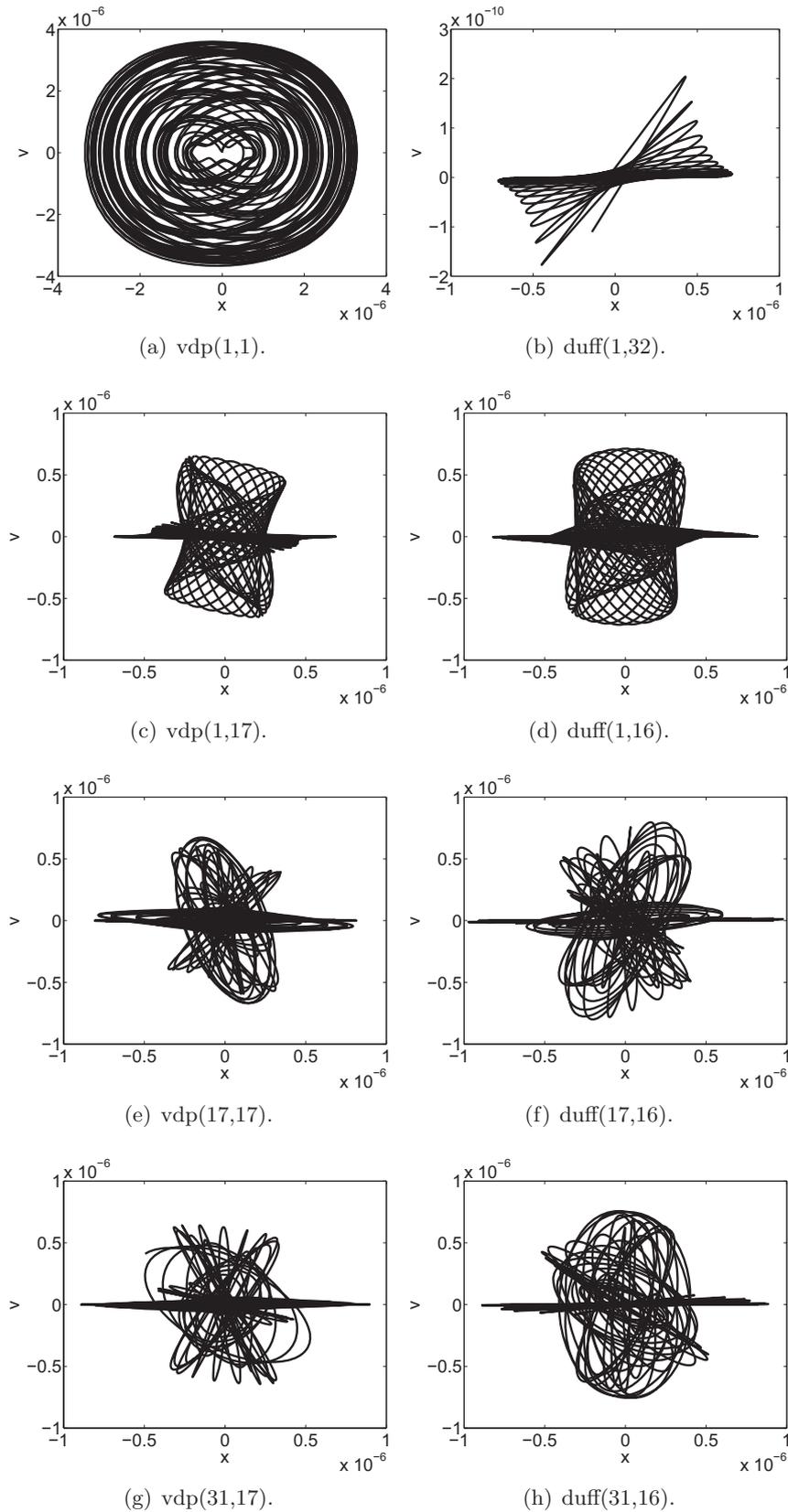


Figure 4.8: Phase portrait of selected elements in an array of 32×32 x -coupled alternating Duffing resonators and van der Pol oscillators. The notation $\text{vdp}(x, y)$ and $\text{duff}(x, y)$ indicates the selected locations (x, y) of a van der Pol or Duffing device. An energy, whose amplitude is $F_d = 0.3$, is injected from the top edge of the lattice. The coupling strength between the oscillators is $\kappa = 1$. The figure shows that small coupling induces different types of complicated behaviours.

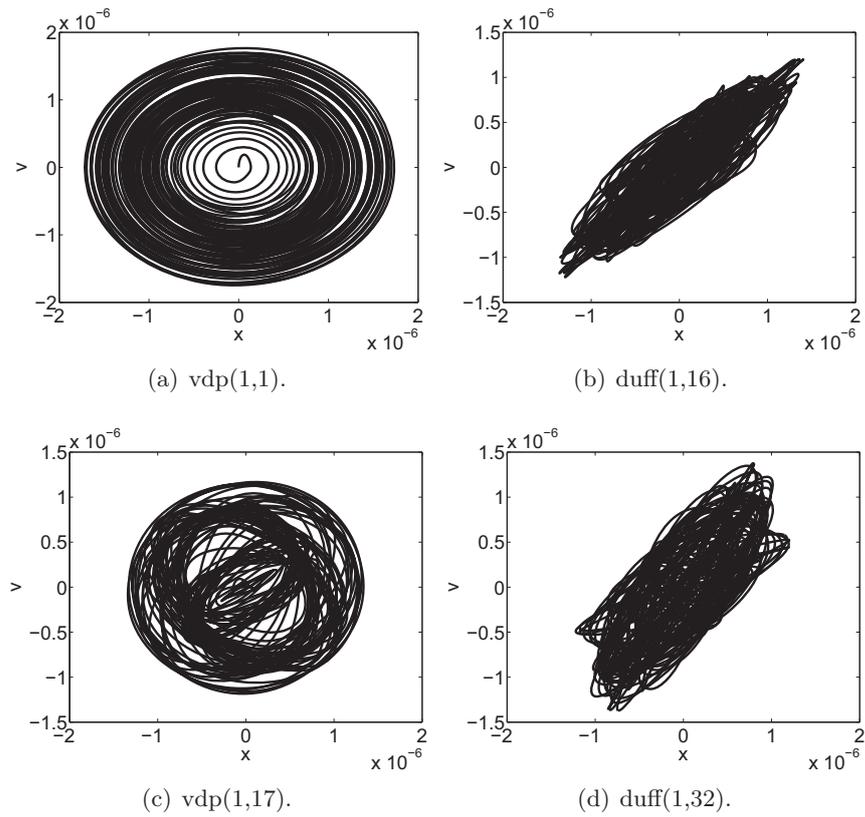


Figure 4.9: Phase portrait of selected elements in an array of 32×32 x -coupled alternating Duffing resonators and van der Pol oscillators. An energy, whose amplitude is $F_d = 0.3$, is injected from the top edge of the lattice. The coupling strength between the oscillators is $\kappa = 100$. The figure shows that more homogeneous behaviours between each type of element are observed with strong coupling.

seen that with low coupling ($\kappa = 1$) the behaviour of the elements is more like random. However, with high coupling ($\kappa = 100$) the phases of the resonators and oscillators located on odd rows, and respectively on even rows, are mutually synchronised. These two distinct clusters are shown in both the phase and amplitude spaces by the normalised polar plot of Figure 4.7(d).

Regarding the dynamical properties of each single element, it can be noticed from the phase portrait plots in Figure 4.8 and 4.9, where the displacement x is plotted against the velocity v , that different types of complicated behaviours of the elements appear with small coupling ($\kappa = 1$) whereas more homogeneous behaviours between each type of element are observed with strong coupling ($\kappa = 100$).

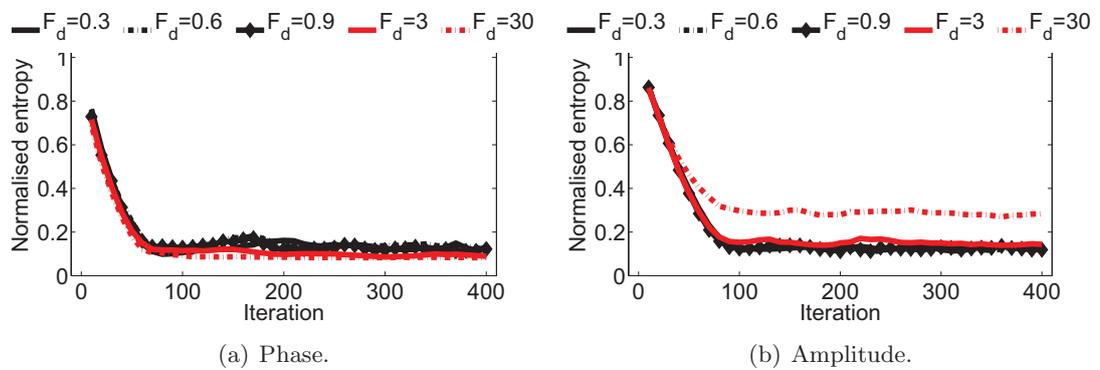


Figure 4.10: Effect of the amplitude of the driving force on the global coherence of the square lattice. When the amplitude of the driving force is increased the phase coherence of the elements of the lattice deteriorates slightly whereas an opposite behaviour is observed in the case of the amplitude coherence.

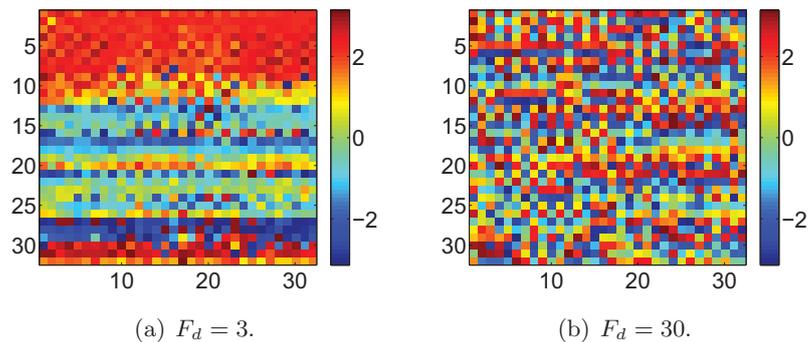


Figure 4.11: Plots of the instantaneous phase of components in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators at a coupling strength of $\kappa = 10$. The array is driven from the top edge of the lattice and the figures are plotted at $t = 3600$ with different amplitudes of the driving force F_d . It can be noticed here that when the lattice is driven ten times higher, then the phases of the elements tend to desynchronise.

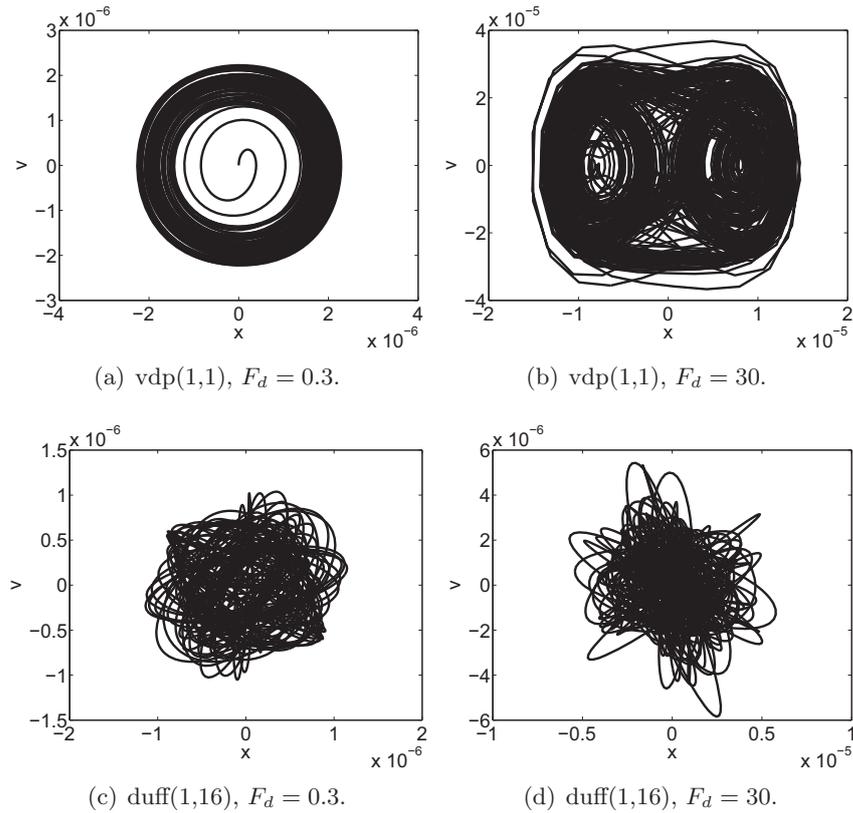


Figure 4.12: Selected phase portrait plots in an array of 32×32 x -coupled alternating Duffing resonators and van der Pol oscillators. An energy, whose amplitude is F_d , is injected from the top edge of the lattice. The coupling strength between the oscillators is $\kappa = 10$. The figure shows that increasing the amplitude of the driving force favours complicated behaviours. The remaining elements in the lattice also have a similar behaviour although the phase portraits are not shown.

The amplitude of the driving force has an opposite effect on the global coherence of the phase and amplitude of the system. More precisely, the higher the amplitude of the driving force the less coherent is the phase of the system whereas the opposite effect is observed for the amplitude of the elements of the lattice as shown by Figure 4.10. However it should be noted that the difference in global phase coherence is not very significant as the amplitude of the driving force is multiplied by hundred. By looking at the cluster synchronisation aspect, it can be observed that increasing the amplitude of the driving force will result in a randomisation of the phases of the resonators and oscillators as depicted by Figure 4.11. The latter result is also valid for the amplitudes of the elements although the plots are not shown. It can also be observed from the phase portrait plots in Figure 4.12 that increasing the amplitude of the driving force favours complicated behaviours.

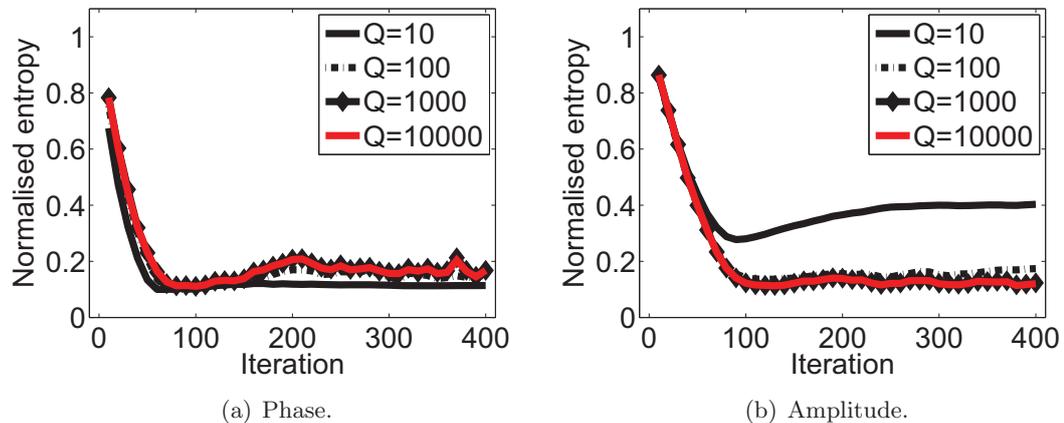


Figure 4.13: Relation between the quality factor Q of the elements of the square lattice and its global coherence. It can be noticed that the lower the value of Q , the worse the phase coherence. In contrast, the amplitude coherence of the lattice is improved as the value of Q is decreased.

The global phase coherence of the lattice is not sensitive to the value of the quality factor. However, for low quality factor value ($Q = 10$) the global amplitude coherence of the system displays better synchronisation behaviour than higher values of Q as shown by Figure 4.13. In contrast to the previous full synchronisation results, more distinguishable amplitude and phase clusters appear in the lattice when the quality factor value is high ($Q = 1000$) as can be seen in Figure 4.14. It should be noted however that for small Q values, the Duffing resonators and van der Pol oscillators have two distinct behaviours. In terms of the effect on the dynamics of the each element, increasing the quality factor Q tends to inhibit their complicated behaviour as shown by the phase portrait plots in Figure 4.15.

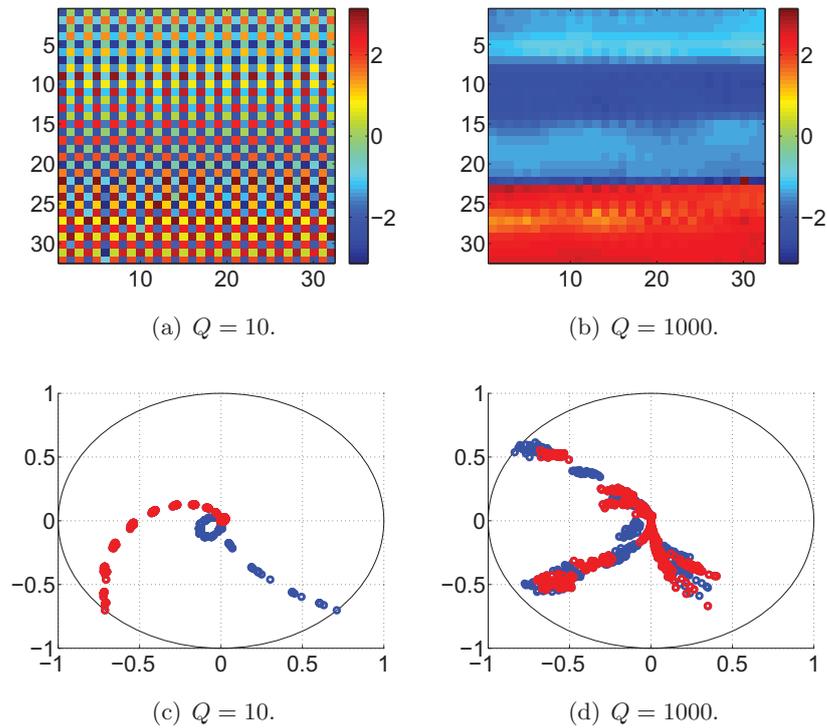


Figure 4.14: Phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The coupling between the elements is $\kappa = 10$ and the figures are plotted at $t = 3600$ with different values of the quality factor Q . The figures show that when Q is small ($Q = 10$) the Duffing resonators (red circles) and the van der Pol oscillators (blue circles) have two distinct behaviours. However at a high quality factor ($Q = 1000$) they tend to synchronise between them.

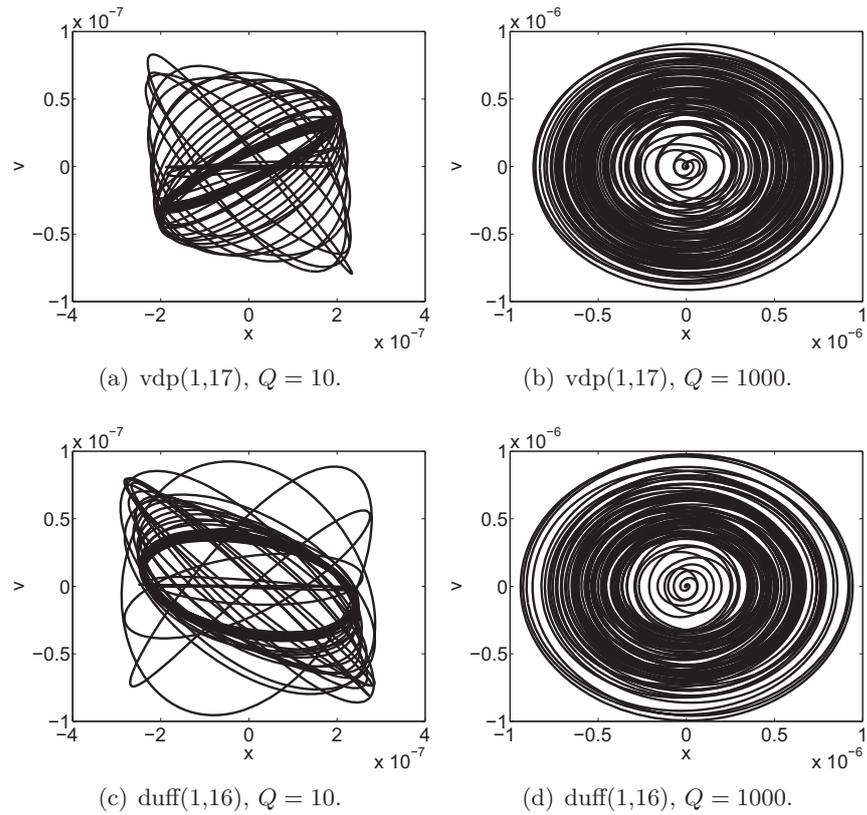


Figure 4.15: Phase portrait of selected elements in an array of 32×32 x -coupled alternating Duffing resonators and van der Pol oscillators. An energy, whose amplitude is $F_d = 0.3$, is injected from the top edge of the lattice. The coupling strength between the oscillators is $\kappa = 10$. Figures 4.15(b) and 4.15(d) show that increasing the quality factor Q ($Q = 1000$) tends to inhibit the complicated behaviour of the elements of the lattice that is illustrated by Figures 4.15(a) and 4.15(c) for example when Q is low ($Q = 10$).

4.2.3 Architecture and size effects

In the previous section, a square lattice of alternating Duffing resonators and van der Pol oscillators has been studied. In this section, we investigate the behaviour of the coupled systems when we permute the Duffing resonators and the van der Pol oscillators in the lattice and modify the size of the lattice.

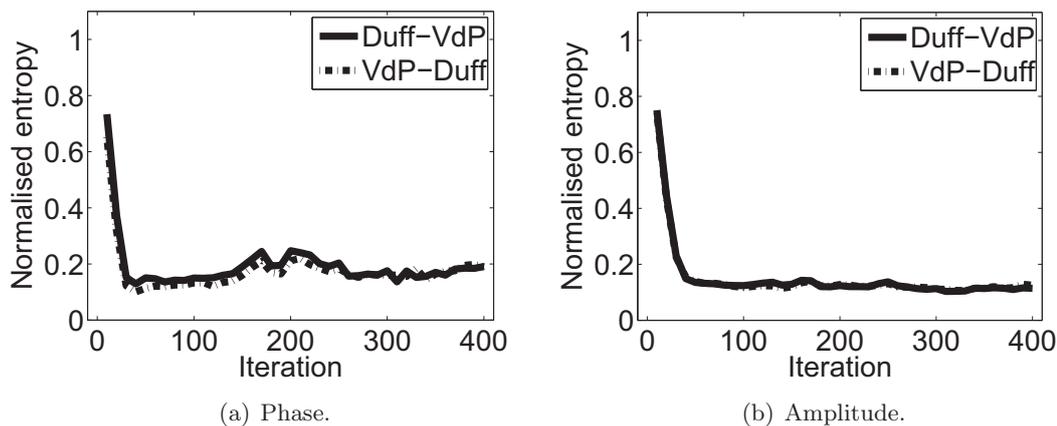


Figure 4.16: Impact of the network architecture on the square lattice global coherence. It is interesting to see that the elements in the two different architectures (Duff-VdP and VdP-Duff) have almost the same behaviour in terms of both the global phase and amplitude coherence although the elements have different behaviours from the results in Figure 4.17.

The two different architectures, characterised by a lattice of alternating Duffing resonators and van der Pol oscillators on the one hand and alternating van der Pol oscillators and Duffing resonators on the other hand, have almost the same behaviour in terms of the global coherence of their phases and amplitudes as shown by Figure 4.16. A completely different situation can be seen when looking at the cluster synchronisation where two different types of behaviour emerge for the two architectures as illustrated by Figure 4.17. Namely, in the case of a lattice composed of alternating Duffing resonators and van der Pol oscillators, different subgroups can be observed in which the phases of the Duffing resonators and van der Pol oscillators are synchronised. In the other case, the Duffing resonators and van der Pol oscillators have two distinct behaviours.

It is also interesting to be able to understand the change of behaviour as the size of the lattice varies. From the plots in Figure 4.18, it can be concluded that the global phase and amplitude coherence of the lattice decrease as the size of the lattice is increased. In the cluster synchronisation case, it can be noticed that the bigger the lattice the more clusters are present (Figure 4.19).

So far we have driven the lattice from one edge and have observed that when the array

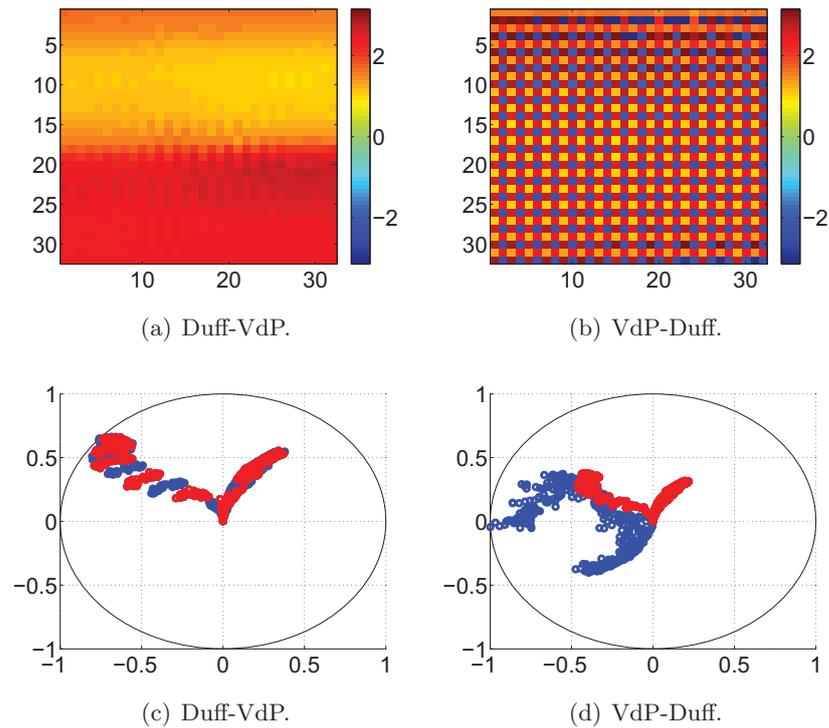


Figure 4.17: Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators, and alternating van der Pol oscillators and Duffing resonators at a coupling strength of $\kappa = 40$. The lattice is driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The figures are plotted at $t = 3600$. The phases of the elements are more synchronised in the first case (Duff-VdP) and it can be observed from the corresponding normalised polar plot that some of the Duffing resonators and van der Pol oscillators evolve in distinct subgroups. In the second case (VdP-Duff), the Duffing resonators tend to group between themselves and so do the van der Pol oscillators.

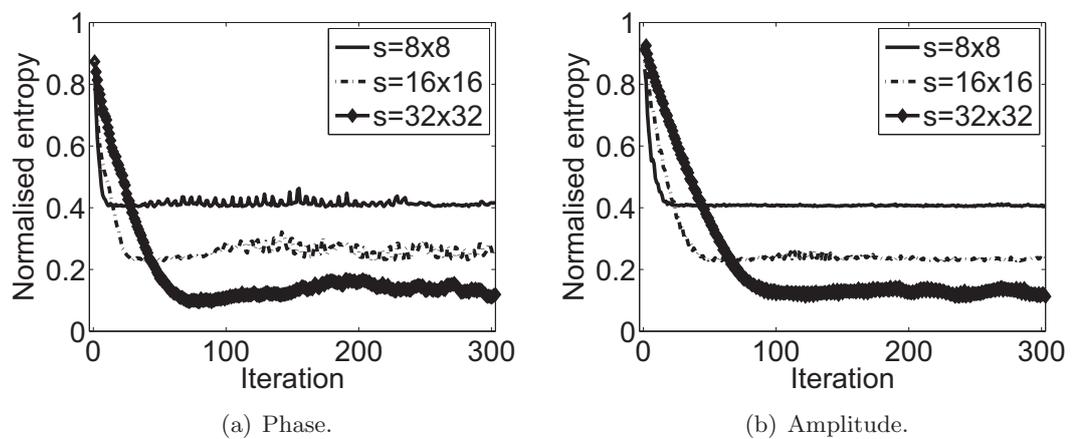


Figure 4.18: Lattice-size effect on the global coherence of the elements. The figures show that the smaller the lattice size, the quicker the convergence and the better the higher the normalised entropy of the phase and amplitude of the elements in the lattice.

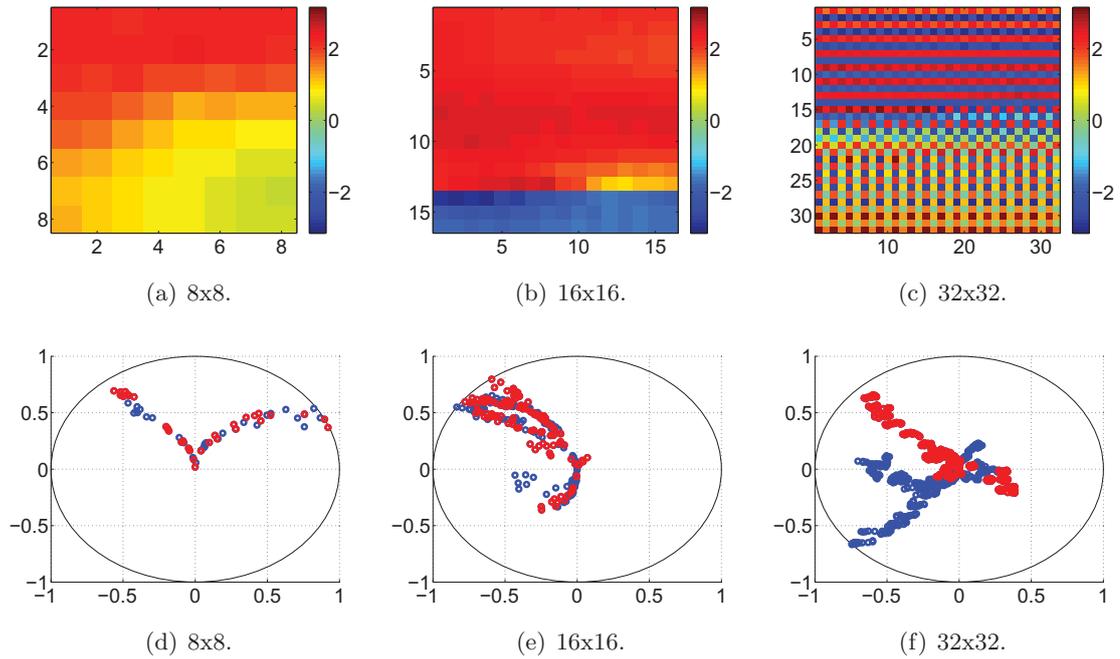


Figure 4.19: Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators at a coupling strength of $\kappa = 10$. The array is driven from the top edge of the lattice and the figures are plotted at $t = 3600$ for different sizes of the lattice. It can be observed that for small (8x8) and medium (16x16) array sizes, the phases and amplitudes of the Duffing resonators and the van der Pol oscillators are in synchrony.

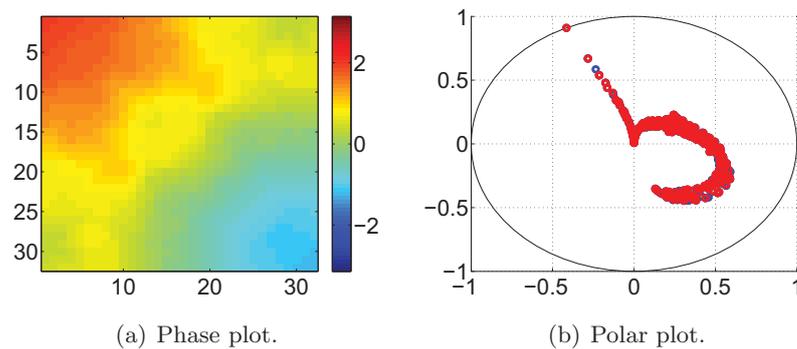


Figure 4.20: Phase and normalised polar plots of a square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators at a coupling strength of $\kappa = 40$. The array is driven from the top left corner of the lattice and the figures are plotted at $t = 3600$. The system displays an almost perfect symmetrical behaviour in terms of the phases of the elements. The symmetry axis, in this case, is given by the diagonal joining the driven Duffing resonator and the last Duffing resonator located on the opposite corner.

is driven from the top edge, the Duffing resonators and van der Pol oscillators tend to synchronise by rows. We have also looked at a different scenario where we drive the lattice only from one of its corners. Namely, we have driven the array from the Duffing resonator located on the top-left corner of the array. It can be observed from Figure 4.20 that the system displays an almost perfect symmetrical behaviour in terms of the phases of the elements. The symmetry axis, in this case, is given by the diagonal joining the driven Duffing resonator and the last Duffing resonator located on the opposite corner.

4.3 Summary

The behaviour of a Duffing resonator coupled to a van der Pol oscillator with MEMS parameters as two coupled elements and in a square lattice has been discussed in the current chapter. The two coupled elements display rich behaviours such as hysteresis, anti-resonance, stagnant response and multiple branches in their frequency-responses which have been obtained using the method of multiple time scales analysis. The stability of the system has also been investigated where two unstable branches of the Duffing resonator and van der Pol oscillator respectively have been found. Moreover the system also displays interesting transient and asymptotic behaviour. The bifurcation analysis also revealed an abrupt transition to chaos of the system. This analysis of the fundamental unit that was then reproduced into a lattice was done in order to show some of the complexity that could occur in a coupled lattice of units.

The global and local synchronisations of a square lattice of alternating Duffing resonators and van der Pol oscillators were studied numerically. Such a lattice could serve as a sensing device where the input signal is simulated with an external driving force that is injected to the Duffing resonators. The parameters of the resonators and oscillators that have been used were taken from real micro-electro-mechanical system (MEMS) devices. The effects of different system parameters was investigated. It was found that the higher the coupling strength value between the elements of the lattice the more synchronised they will become, as expected. In contrast, when the amplitude of the driving force is increased the elements of the lattice tend to randomise. It has been observed that with high quality factor values the synchronisation between the elements is favoured whereas a low quality factor value resulted in two distinct behaviours of the resonators and oscillators.

Besides the effects of the system parameters on the behaviour of the coupled systems, a different architecture has also been investigated. More precisely, a permutation of the location of the resonators and oscillators has been carried out, resulting in a square lattice

of alternating van der Pol oscillators and Duffing resonators. The behaviours of the elements in the two types of lattice are different in that the resonators and oscillators can be synchronised in the former case while they evolve quasi-independently in the latter case. The behaviour of the elements in the lattice is also dependent on its size. In particular, different clusters of synchronisation can be observed with different array sizes composed of even number of elements. However, no distinguishable clusters of synchronisation appear when the dimension of the lattice is odd. This suggests that the unit composed of a Duffing resonator coupled with a van der Pol oscillator represents an important entity in order to have rich synchronisation behaviours.

The square lattice has been driven in two different ways: from the top edge and the top left corner reflecting likely interaction in a real physical array. It has been observed that in the first case, the resonators and oscillators tend to synchronise by rows provided that the coupling strength is high enough. In the second case, a symmetric behaviour of the elements of the lattice can be seen where the axis of symmetry is given by the diagonal joining the driven Duffing resonator on the top left corner and the last Duffing resonator on the opposite corner. Different initial conditions have also been used but the system is not sensitive to them at all.

The square lattice that has been studied can be seen as an ideal 2D array of resonators and oscillators. However, it does not reflect a real physical device where random parameter mismatch and connections can occur as a result of the imperfections of the fabrication process. The behaviour of the system when taking into account those random noise should also be studied and will be the focus of the next chapter.

5 Probabilistic square lattice of Duffing resonators and van der Pol oscillators

We have seen in the previous chapter how the coupled van der Pol oscillators and Duffing resonators behave in a square lattice that is subject to a driving force from one of its edges and corners. The oscillators and resonators were simulated with MEMS parameters and each of them were connected to its four nearest neighbours except the elements located on the corner and edges of the lattice which had two and three neighbours respectively. The randomness was limited to the random initial configurations of the components in the lattice. A real MEMS device is very likely to have physical imperfections which can be regarded as noise and the substrate on which the devices are constructed can introduce parasitic capacitance tracks. These defaults can be translated into random parameter mismatch or random electrical connections between the different components of the square lattice which can then be regarded as a *probabilistic square lattice*. Therefore, the square lattice architecture that was studied in the previous chapter is an ideal one and does not reflect a real physical system. Thus, a question that arises is: how these random connections would affect the collective behaviour of the oscillators and resonators in the square lattice? The addition of random mismatch terms and connections in the square

lattice naturally leads us to consider the theory of synchronisation in complex networks in which different distributions of connections can be present and the effects of those topologies on the synchronisation of oscillatory systems are investigated. However, it should be noted that apart from the case of a square lattice topology with added random connections, the different complex networks architectures are not physically conceivable from a MEMS device perspective but constitute an interesting conceptual problem that can be present in other fields such as biology (Elowitz and Leibler, 2000; García-Ojalvo et al., 2004; Wagemakers et al., 2006; Koseska et al., 2007; Strogatz, 2003; Chialvo and Jalife, 1987) and social science (Buchanan, 2007; Chialvo and Jalife, 2005) to mention a few. This chapter is organised as follows: an overview of the study of synchronisation in random networks of oscillators will first be given in section 5.1. Next, some synchronisation indices that will be used in the study of the current probabilistic network of resonators and oscillators will be described in section 5.1.1. In section 5.2, numerical simulation results of the square lattice with random noise will be presented.

5.1 Synchronisation in random networks of oscillators

The analysis of synchronisation in the complex networks literature is mainly based on the coupling matrix, which will be defined later, and its extreme eigenvalues (Arenas et al., 2008; Dorogovtsev et al., 2008; Marro and Dickman, 1999; Gómez-Gardenes et al., 2007a,b; Gómez-Gardenes and Moreno, 2007; McGraw and Menzinger, 2005, 2007; Restrepo et al., 2005, 2006). A complex network can be defined as a graph having a set of nodes which are connected by a set of random links. The links are usually undirected and can be weighted where the weight represents the coupling strength between two nodes. On the other hand, a clustered complex network (Huang et al., 2007) consists of a number of groups, in which nodes within each group are densely connected, but the linkage among the groups is sparse.

The tools to characterise synchronisation in complex networks will therefore be applied and discussed in the context of the MEMS square lattice where random connections are introduced and their effects on synchronisation will be studied. In complex networks, random connections are usually added to a regular array, that is, an array whose elements have the same number of neighbours. The case of the square lattice is peculiar in this thesis in that it models a real MEMS physical system. In particular, the corners have two neighbours, the edges have three and all other elements have four neighbours. The corresponding MEMS system where random connections are added will therefore be referred

to as a *probabilistic square lattice*.

The analysis of synchronisation in complex networks has been focused mainly on complete synchronisation of identical oscillators in the field of nonlinear dynamics. The stability of the completely synchronised state, also called *synchronisability* (Arenas et al., 2008), has also been studied analytically (Zhou et al., 2006). The importance of the analysis of stability comes from the fact that it is an indication of the emergence of patterns when the stability is lost or it leads to a coherent state when the stability can remain (Pecora, 2008). The stability has been studied under the Master Stability Function (MSF) approach (Pecora and Carroll, 1998; Huang et al., 2009) which is a tool to analyse various synchronisation problems and will be briefly described in the next section.

In the following, a description of the tools to characterise synchronisation in the probabilistic square lattice will be given.

5.1.1 Characterisation of synchronisation in the probabilistic square lattice

The main statistics and synchronisation indices that will be borrowed from the complex networks literature are the average shortest path length, the average clustering coefficient and the extreme eigenvalues of the coupling matrix. The degree distribution will be ignored since, on the one hand, the practical effect of the degree distribution on the MEMS square lattice is not very important since it is more relevant to the differentiation of different types of complex networks. On the other hand, the degree of each node is implicitly embedded in the clustering coefficient as will be seen later. A description of each of these statistics and synchronisation indices will be given next.

The average shortest path length is denoted by $l = \langle d_{ij} \rangle$, where d_{ij} is the length of the shortest path between node i and j and $\langle . \rangle$ represents the average operator over the distances d_{ij} . The clustering coefficient C is defined as:

$$C = 1/N \sum_{i=1}^N C_i = 1/N \sum_{i=1}^N \frac{n_i}{k_i(k_i - 1)/2}, \quad (5.1)$$

where n_i is the number of connections between nearest neighbours of node i . The degree k_i of a node i represents its number of links. A large clustering coefficient implies many transitive connections and consequently redundant paths in the network, while a low value of C implies the opposite.

Before discussing the spectral properties of the probabilistic square lattice, it is necessary to define a few graphical concepts. In our case, the network is characterised by an undirected graph. The graph can be described by its *adjacency matrix* A whose entry

$a_{ij} = 1$ if the i^{th} node is connected to the j^{th} node and $a_{ij} = 0$ otherwise. The previous definition applies to unweighted undirected graphs, that is the connections between nodes have the same strength. In the case where the coupling strength varies between connections then $a_{ij} = w$, where w is the strength or weight of the connection between the nodes i and j . Besides the adjacency matrix, the graph of the network is also characterised by its diagonal matrix D of node degree. For the general case of weighted networks, the entry d_{ii} of the matrix D is the sum of the weights of edges connected to the node at index i . And finally, the coupling matrix also called the *Laplacian matrix* L of the undirected probabilistic square lattice is defined as $L = D - A$, where A is the adjacency matrix and D is the diagonal matrix of node degrees. The measures of synchronisation are based on the extreme eigenvalues of the Laplacian matrix. The first measure is given by the eigenratio $R = \lambda_N/\lambda_2$ and the second measure is just λ_2 , where λ_2 is the minimal nonzero eigenvalue and λ_N is the maximal eigenvalue of the Laplacian matrix. These two measures are called type I and type II synchronisation respectively.

The assessment of synchronisability is done under the Master Stability Function (MSF) approach which provides an objective criterion, based on λ_2 and λ_N . However, this approach only informs about the dynamics toward synchronisation from small perturbations of the synchronisation manifold which is the subspace in the state space of the dynamical equation of motion where all the oscillators evolve synchronously on the same solution of the isolated oscillator. Moreover the MSF approach assesses the linear stability of the completely synchronised state, which is a necessary but not a sufficient condition for synchronisation, and it requires that the oscillators be identical.

5.2 Numerical results

In this section, the effect of noise in the square lattice is studied numerically. In our case, the source of the noise is mainly due to the imperfections occurring during the fabrication process and can take the form of different parameter values given to the elements, such as the intrinsic frequencies of the oscillators, the quality factor or the coupling strengths between the components for example. Besides the elements of the square lattice having different parameter values, the noise in the system can also be characterised by random connections which can also be seen as possible defects of the fabrication process. It is therefore necessary to study the synchronous behaviour of the elements of the square lattice when a small random noise is added to the system.

5.2.1 Small noise added to the natural frequencies of the oscillators

The frequency mismatch between the oscillators is characterised by adding a small uniform random noise in the interval $[-5.4619 \times 10^3, 5.4619 \times 10^3]$ to their intrinsic frequencies ω_0 . The length of the previous interval corresponds to one percent of $\omega_0 = 109237$ rad/s. The amount of noise has also been increased up to ten percent of ω_0 .

The global synchronisation of the square lattice is not affected by the noise added to the frequency of the oscillators as shown by the plots of the global synchronisation index of the phase and amplitude of the elements in Figure 5.1. However, by looking at the phase and normalised polar plots of Figure 5.2, it can be observed that the introduction of a frequency mismatch improves the synchronisation between the oscillators and resonators as expected. As a comparison, in Osipov and Sushchik (1998), the collective behaviour of a one dimensional array of diffusively coupled van der Pol oscillators at weak and relatively strong nonlinearity has been investigated when a frequency noise that is uniformly distributed in the interval $[-0.5; +0.5]$ is introduced in the array. It has also been observed that the characteristics of the synchronisation are improved when an irregular distribution of natural frequencies is used.

Figures 5.2(d) and 5.2(f) show that the introduction of a frequency mismatch cancels the two independent motions of the Duffing resonators and the van der Pol oscillators shown in Figure 5.2(b). This is translated into a continuous band (Figures 5.2(c) and 5.2(e)) of the instantaneous phases instead of a mixture of stripes and discrete patterns in the noise-free case (Figure 5.2(a)). In addition, the behaviour of the system is exactly the same regardless of the amount of noise added to the natural frequency of the oscillators.

5.2.2 Small noise added to the damping

Since the quality factor Q is related to the damping γ by $\gamma = 1/Q$, in this section, a Gaussian random noise with mean and standard deviation equal to half of one and ten percent of $\gamma = 2 \times 10^{-4}$ is added to the damping term of each oscillator and resonator. The mean has been set to 1×10^{-6} and 1×10^{-5} instead of the standard zero value to avoid having negative damping coefficients as this would result in a different type of oscillator or resonator. Based on the results from the global synchronisation index in Figure 5.3, the small noise on the damping does not seem to have an effect on the global synchronisation of the elements since the time series of the global synchronisation index in both the phase and amplitude spaces are almost similar to the noise-free case. However, by looking directly at the phases and amplitudes of the elements in the square lattice, it can be noticed from the

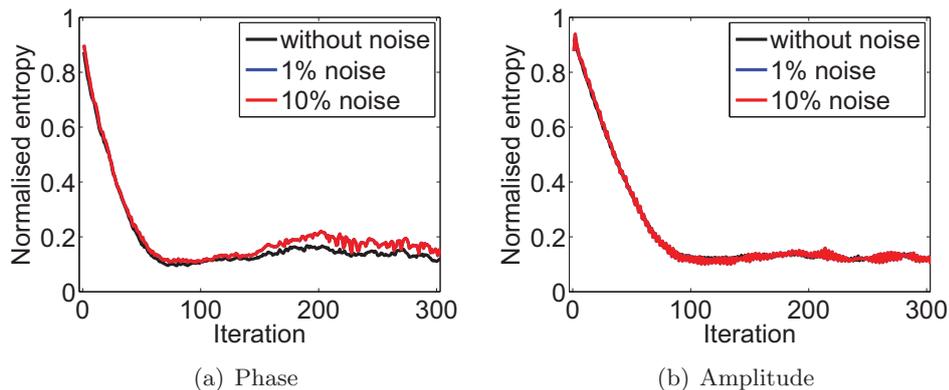


Figure 5.1: Natural frequency mismatch effect on the global coherence of the square lattice. A small uniform random noise is added to the intrinsic frequencies ω_0 of the oscillators, where the lengths of the intervals from which the uniform random noise is taken correspond respectively to one and ten percent of $\omega_0 = 109237$ rad/s. Both the phase and amplitude global coherence indices are similar to the noise-free case which indicates that the global behaviour of the system is not sensitive to the small noise added to the natural frequencies of the oscillators.

plots in Figure 5.4 that the noise added to the damping, provided that it is small enough, improves the synchronisation in the square lattice and especially in terms of the phases of the elements. The effect of adding a small noise to the damping on synchronisation was not expected: surprisingly, when one percent noise is added to the damping of the elements, the system behaves exactly as in the case where one percent noise is added to the intrinsic frequencies of the oscillators. As a matter of fact, Figures 5.2(c) and 5.2(d) are similar to Figures 5.4(c) and 5.4(d) respectively. However, when the noise is slightly bigger (10%), the system behaves as if there was no noise at all as shown by Figures 5.4(a) and 5.4(e) in one hand and Figures 5.4(b) and 5.4(f) on the other hand. In the noise-free case, a phase synchronisation by rows can only be observed on the top half of the lattice (Figure 5.4(a)), whereas it can be seen on the whole lattice when the small noise (1%) in the damping is introduced (Figure 5.4(c)). Moreover, the number of rows that are in synchrony is higher in the noisy case than in the noise-free one.

5.2.3 Random links

Besides the difference in the quality factor value as seen previously, the imperfections in the fabrication of the system can also be translated into the presence of random connections between the oscillators and resonators. The number of extra random links is here chosen in terms of the percentage of the total number of elements in the square lattice. Similarly to the previous case of noise introduced on the damping, the global coherence index does not

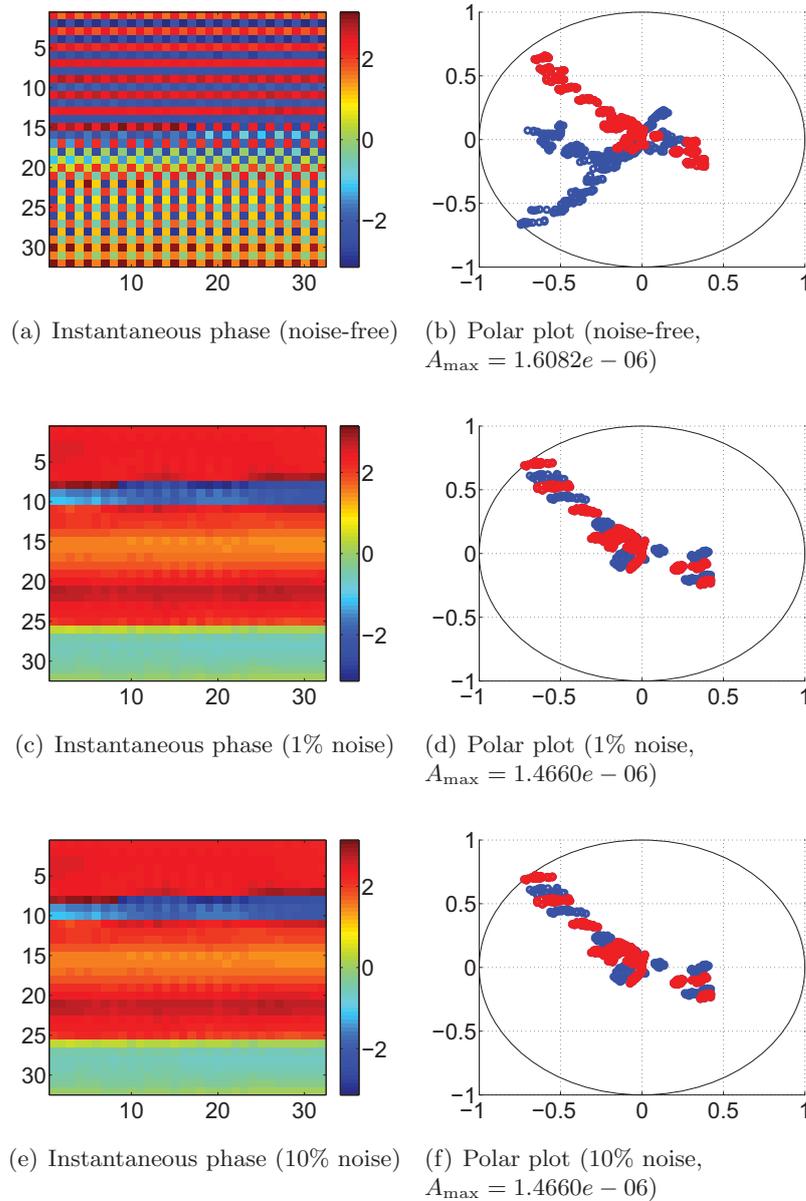


Figure 5.2: Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The figures are plotted at $t = 3600$ with and without a small uniform random noise added to the natural frequencies ω_0 of the oscillators. The lengths of the intervals from which the uniform random noise is taken correspond respectively to one and ten percent of $\omega_0 = 109237$ rad/s. The coupling strength between the elements is $\kappa = 10$. Figures 5.2(d) and 5.2(d) show that the introduction of a frequency mismatch cancels the two independent motions of the Duffing resonators and the van der Pol oscillators shown in Figure 5.2(b). This is translated into a continuous band (Figures 5.2(c) and 5.2(e)) of the instantaneous phases instead of a mixture of stripes and discrete patterns in the noise-free case (Figure 5.2(a)). In addition, the behaviour of the system is exactly the same regardless of the amount of noise added to the natural frequency of the oscillators.

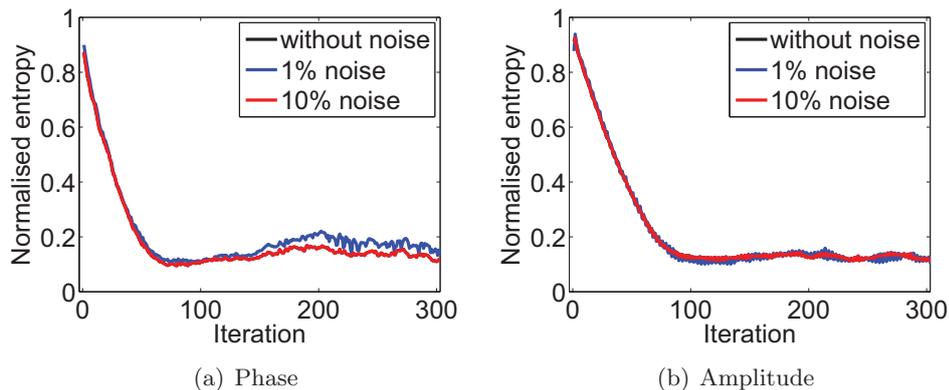


Figure 5.3: Damping noise effect on the global coherence of the square lattice. When a small random Gaussian noise is added to the dampings of the oscillators and resonators, both the phase and amplitude global coherence indices are almost similar to the noise-free case. This implies that the global behaviour of the system is not sensitive to the small noise added to the damping term.

reflect the change of behaviour of the system in the phase space with the addition of random connections. As a matter of fact, the global coherence index of the system with extra random connections in the phase space is almost the same as the lattice without random connections (Figure 5.5(a)). However, the global coherence is improved asymptotically in the amplitude space (Figure 5.5(b)). On the other hand, Figures 5.6(c) and 5.6(e) show that the addition of random connections in the square lattice completely destroys the phase synchronisation in Figure 5.6(a) where the ordered phase patterns have been transformed into completely random ones as expected. However, the lower the amount of random connections added, the more the amplitudes of the elements tend to be gathered in one group as shown by the normalised polar plots in Figures 5.6(d) and 5.6(f). By looking at the MSF synchronisability indices and the network statistics in Figure 5.7, it can be observed that neither the type I (the ratio of extreme nonzero eigenvalues λ_N/λ_2) nor the type II (the smallest positive eigenvalue λ_2) synchronisation indices best characterise the synchronisability in the square lattice as the number of random links is increased – but rather the largest eigenvalue λ_N of the Laplacian matrix. Regarding the distance and clustering statistics of the square lattice, the average shortest path length decreases with increasing number of random links as expected (Figure 5.7(d)). However the average clustering coefficient does not follow the same trend (Figure 5.7(e)).

5.2.4 Random links with various coupling strengths

In the previous section, random links are added into the square lattice to account for physical imperfections. However, the coupling strengths of the random links are equal to

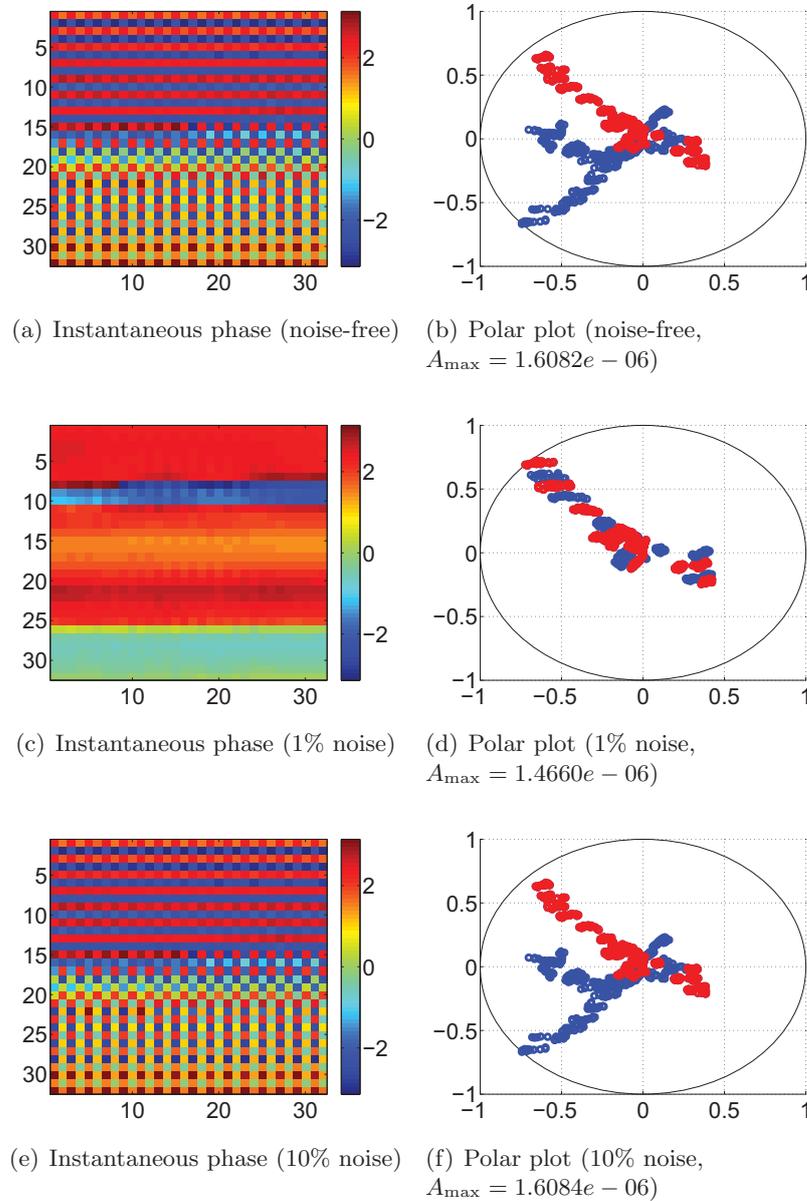
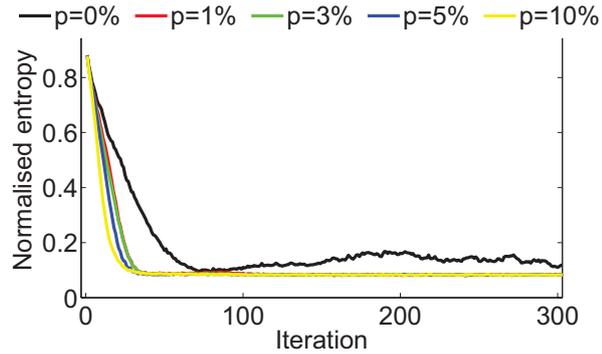
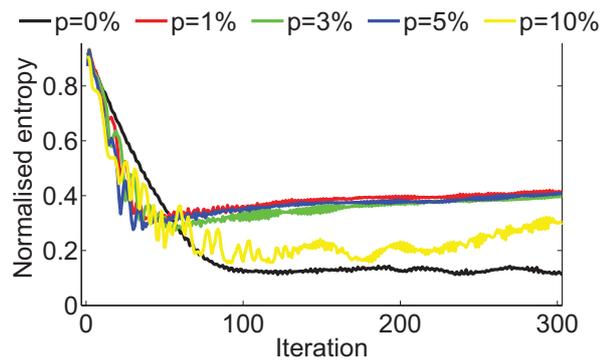


Figure 5.4: Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The figures are plotted at $t = 3600$ with and without a small Gaussian noise added to the damping term of the oscillators and resonators. The coupling strength between the elements is $\kappa = 10$. Figures 5.4(c) and 5.4(d) show that the small random noise (1%) added to the damping term improves the synchronisation between the elements similarly to the observations in Figure 5.2 where a frequency mismatch of the elements was introduced. However, when the noise is slightly bigger (10%), the system behaves as if there was no noise at all.



(a) Phase



(b) Amplitude

Figure 5.5: Effect of extra uniform random connections on the global coherence of the square lattice. The variable p corresponds to the number of random links which is associated to the percentage of the total number of oscillators. For example, with $p = 1\%$, 11 random links were added in the square lattice of 32×32 oscillators. The global phase coherence of the elements in the lattice is not sensitive to additional random connections (5.5(a)). However the global amplitude coherence index improves asymptotically with the addition of random connections (5.5(b)).

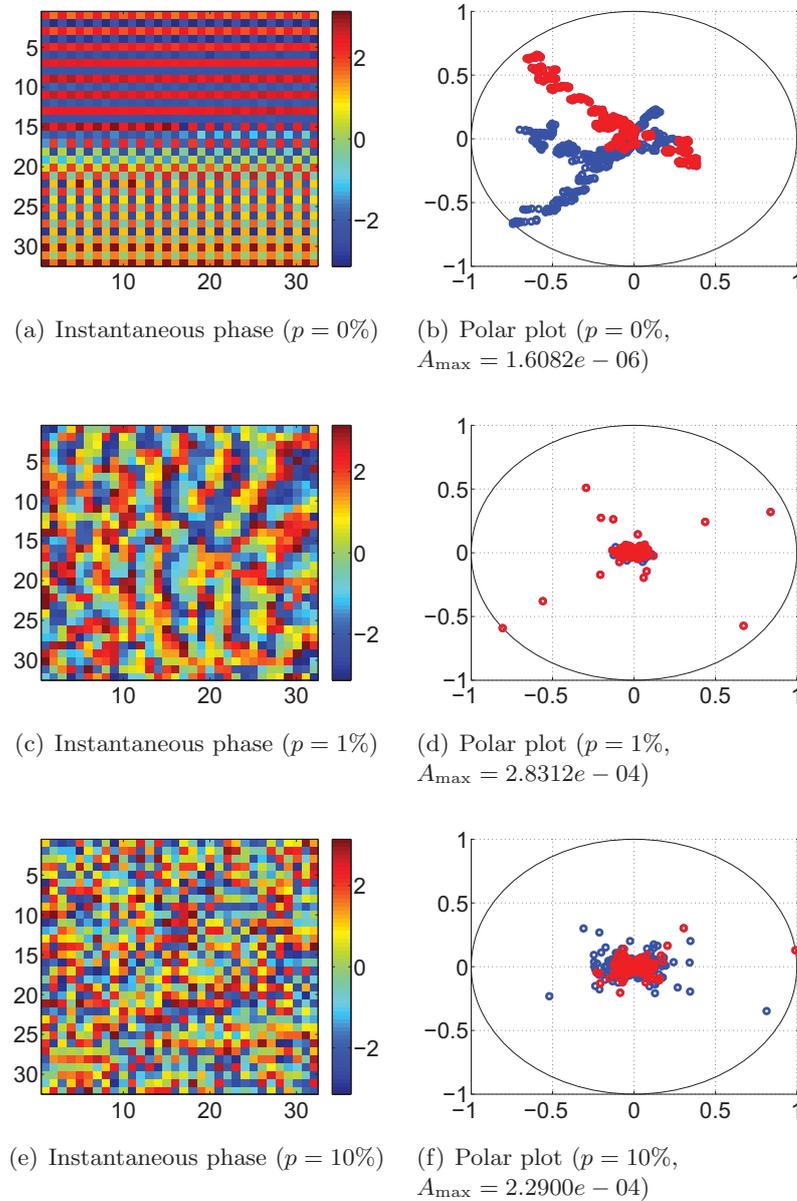


Figure 5.6: Phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The figures are plotted at $t = 3600$ with different number of uniform random links which is associated to the percentage p of the total number of oscillators. Figures 5.6(c) and 5.6(e) show that the addition of random connections in the square lattice completely destroys the phase synchronisation in Figure 5.6(a) where the ordered phase patterns have been transformed into completely random ones. However, the lower the amount of random connections added, the more the amplitudes of the elements tend to be gathered in one group as shown by the normalised polar plots in Figures 5.6(d) and 5.6(f).

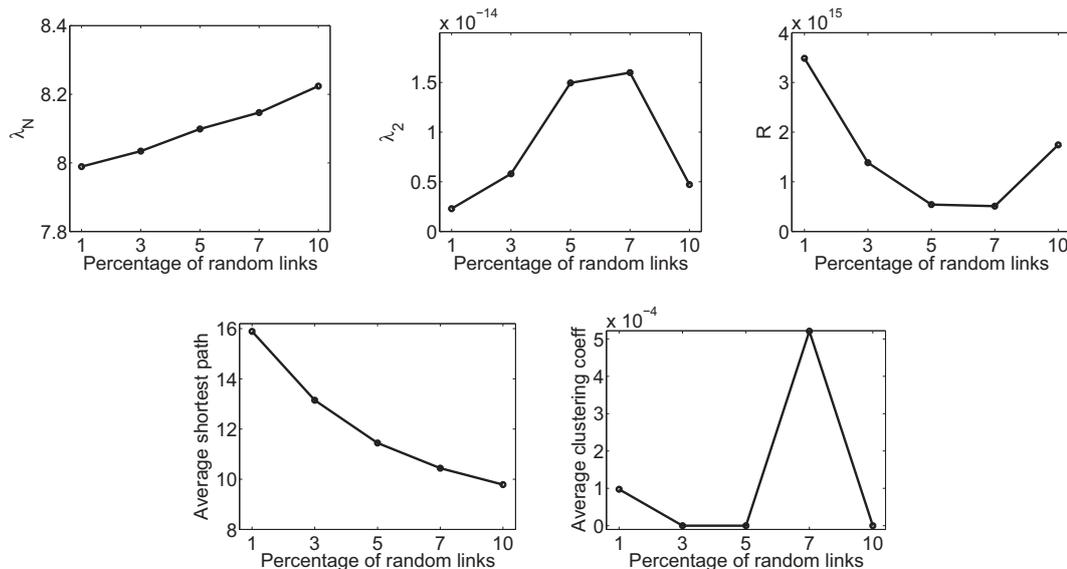


Figure 5.7: Synchronisability and statistics in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The number of uniform random links is associated to the percentage p of the total number of oscillators. The largest eigenvalue λ_N of the Laplacian matrix increases with the number of random links. Nevertheless, the average shortest path length appears to be the best measure of synchronisability: it decreases with increasing number of random links as expected. However the average clustering coefficient does not follow the same trend.

the non-random connections which may not be realistic since the strengths of the random connections should be less or greater than the actual nearest neighbour couplings. Therefore, a coefficient k is introduced in order to get a fraction or a multiple of the actual coupling strengths for the random links. Figure 5.9 shows that the global phase coherence is not affected by the strength of the random connections, whereas the global amplitude coherence is almost the same for extreme values of k (0.1 and 3) while the coherence index is improved when the strengths of the random connections are similar to those of the four nearest neighbour ones ($k = 1$). On the other hand, Figures 5.10(c)-5.10(f) show that increasing the coupling strength of the random connections does not help in improving the synchronisation when random links are introduced in the square lattice. It would be expected that, in the case of a strong coupling of the random connections, hubs of elements around those random links would appear and resulting in few distinct clusters. The reason why it does not happen here may be due to the fact that the presence of random connections destroys the unit composed of a Duffing resonator coupled to a van der Pol oscillator and hence the loss of synchronisation in the square lattice.

In the previous section, the synchronisability of the square lattice has been mainly characterised by the eigenvalues of the Laplacian matrix. However, when the coupling strengths

between the elements of the lattice are heterogeneous, another synchronisability index called *intensity* is more appropriate in this case of a weighted network S (Arenas et al., 2008). The distribution of intensity has been used to characterise the synchronisability of weighted complex networks (Barrat et al., 2004; Yook et al., 2001; Newman, 2001; Braunstein et al., 2001). The intensity S_i of a node i is defined in our case by

$$S_i = \sum_{j=1}^N a_{ij} \kappa_{ij}, \quad (5.2)$$

where N is the number of elements in the square lattice, a_{ij} is an entry of the adjacency matrix A and κ_{ij} represents the coupling strength between the elements i and j . Zhou et al. (2006) have proved analytically and verified numerically that the distribution of the intensities S_i controls the synchronisability of sufficiently random complex networks with large enough minimal degree ($k_{\min} \gg 1$).

Another synchronisation index which is related to the intensity is the *cost* C_0 of a network (Motter et al., 2005a,b) and is defined by

$$C_0 = \frac{\langle S \rangle}{\lambda_2}, \quad (5.3)$$

where $\langle S \rangle$ is the average intensity of nodes in the square lattice.

The intensity distributions are shown in Figure 5.8 where 11 extra uniform random links, which corresponds to 1% of the total number of elements, are added to the square lattice. The data points represent averages over 50 realisations of the square lattice configurations. In Figure 5.8(a), κ represents the coupling strength of the random connections and the figure shows that the synchronisability of the lattice depends on its average intensity.

5.2.5 Random driving sites

In the previous section the effect of varying the strength of additional random connections on synchronisation has been looked at. We now turn to the aspect of controlling the array by injecting random control signals. We are interested in finding whether it is necessary to drive the array through a small or large number of random driving sites. To this end, the number of random driving sites is increased from 10 to 1000. It should be noted that only sparse random control is physically feasible.

Figure 5.11(a) shows that the global behaviour of the array in phase space is not sensitive to the number of random driving sites. Therefore, random sparse control of the phases of the elements in the lattice is applicable. However, in Figure 5.11(b), the amplitudes of the elements become more synchronised globally when most of the elements are driven. This observation is expected. Moreover, the difference in behaviour in terms of the global amplitude coherence is negligible between 10 and 100 random driving sites.

In terms of local synchronisation, Figures 5.12(a) to 5.12(f) show that, regardless of the number of driving sites, the procedure of driving the square lattice from randomly selected sites has the effect of locking the phases of the elements in two different states: $\frac{3\pi}{4}$ and $-\frac{\pi}{4}$. However the amplitudes of the elements vary with different number of driving sites (Figures 5.12(b), 5.12(d) and 5.12(f)).

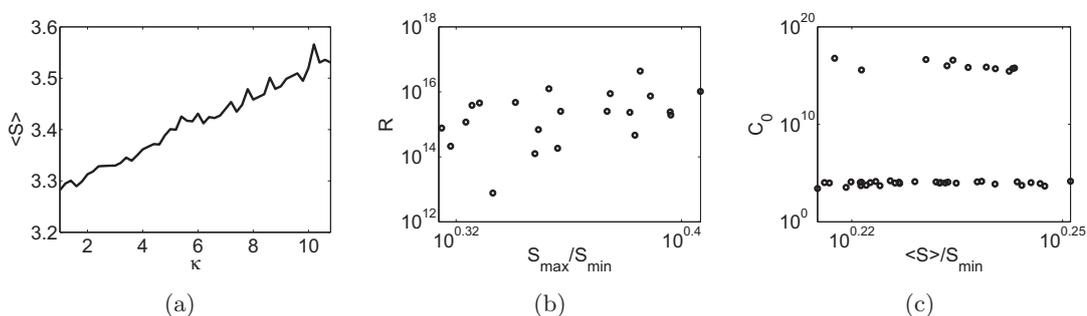


Figure 5.8: Intensity distributions in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. 11 extra uniform random links which corresponds to 1% of the total number of elements are added to the square lattice. κ is the coupling strength of the random connections. The data points represent averages over 50 realisations of the square lattice configurations. 5.8(a) shows that the synchronisability of the lattice depends on its average intensity.

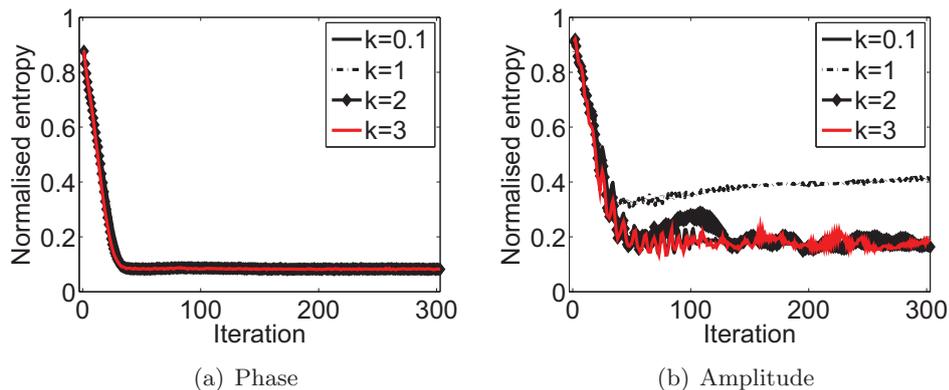


Figure 5.9: Effect of extra uniform random connections with various coupling strength on the global coherence of the square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators. 11 random links are added uniformly which corresponds to 1% of the total number of elements in the lattice. The coefficient k represents the fraction or multiple of the coupling strength κ of the non-random four nearest neighbour connections which is $\kappa = 10$ here. The global phase coherence is not affected by the strength of the random connections (Figure 5.9(a)). On the other hand, the global amplitude coherence is almost the same for extreme values of k (0.1 and 3) while the coherence index is improved when the strengths of the random connection are similar to those of the four nearest neighbour ones ($k = 1$) (Figure 5.9(b)).

5.3 Summary

In this chapter, the effect of noise in the square lattice has been studied. The noise was represented by different parameter values of the elements, namely the natural frequency, the damping coefficient and the coupling strengths, instead of having identical parameters. Random links were also added to the four nearest-neighbour connections of the square lattice and random sites were selected as the source of the driving force.

The introduction of a frequency mismatch improves the synchronisation between the oscillators and resonators especially in the phase space. In addition, the behaviour of the system is exactly the same regardless of the amount of noise added to the natural frequency of the oscillators. It was also observed that the noise added to the damping, provided that it is small enough (1%), improves the synchronisation in the square lattice and especially in terms of the phases of the elements. Interestingly, when one percent noise is added to the damping of the elements, the system behaves exactly as in the case where one percent noise is added to the intrinsic frequencies of the oscillators. However, when the noise is slightly bigger (10%), the system behaves as if there was no noise at all. Moreover, the amplitude and phase synchronisations are destroyed regardless of the value of the coupling strength of the random connections that were introduced. Finally, the procedure of driving the square lattice from randomly selected sites has the effect

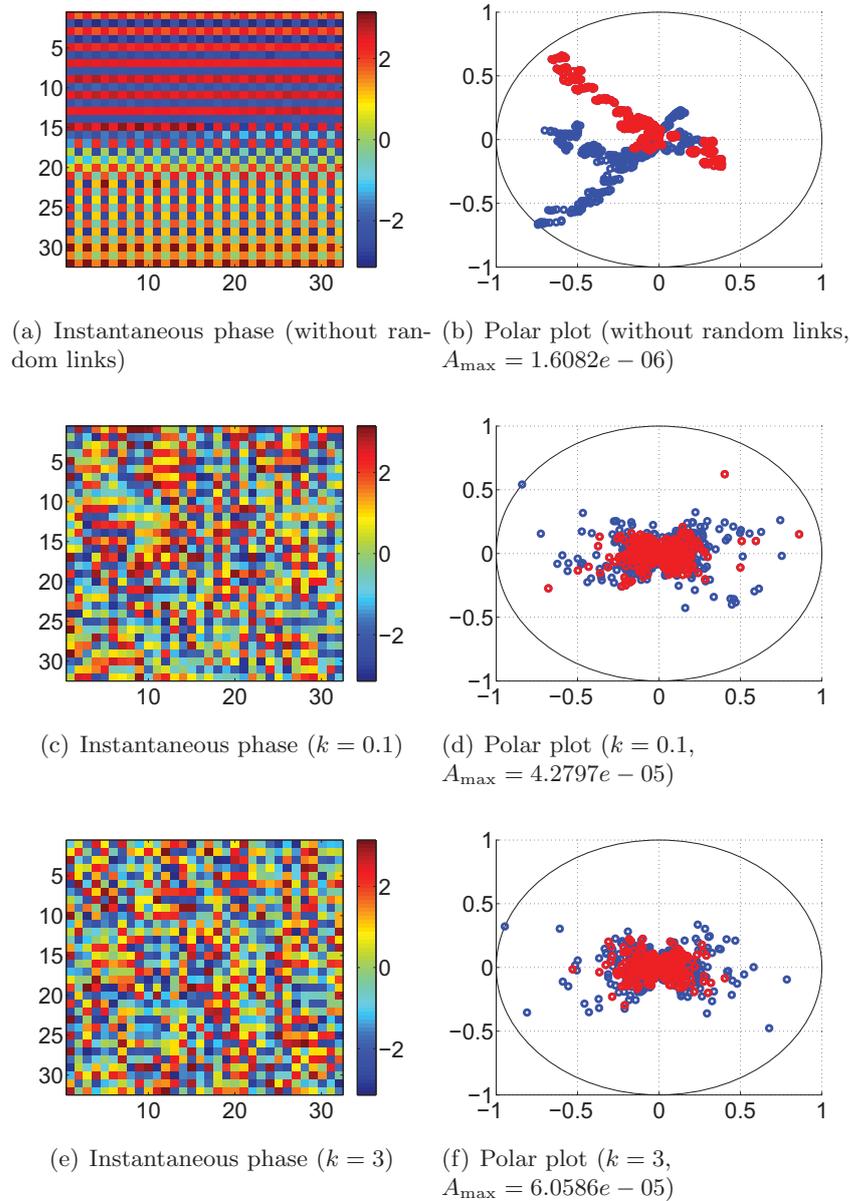


Figure 5.10: Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. The figures are plotted at $t = 3600$ with 11 extra uniform random links which corresponds to 1% of the total number of elements in the lattice. The coefficient k represents the fraction of the coupling strength κ of the non-random four nearest neighbour connections which is $\kappa = 10$ here. Figures 5.10(c)-5.10(f) show that increasing the coupling strength of the random connections does not help in improving the synchronisation when random links are introduced in the square lattice. It would be expected that, in the case of a strong coupling of the random connections, hubs of elements around those random links would appear and resulting in few distinct clusters. The reason why it does not happen here may be due to the fact that the presence of random connections destroys the unit composed of a Duffing resonator coupled to a van der Pol oscillator and hence the loss of synchronisation in the square lattice.

of locking the phases of the elements in two different states: $\frac{3\pi}{4}$ and $-\frac{\pi}{4}$. However the amplitudes of the elements are not affected.

Different synchronisation indices have been used to characterise the synchronisation of the elements of the square lattice. Among those characteristics are the global synchronisation index and the phase image and normalised polar plots that have been used in the previous chapter. The stability of the synchronisation has also been considered using indices obtained from the MSF approach. The global synchronisation index does not always reflect the real synchronous behaviour of the elements as shown by the phase images and normalised polar plots.

An important drawback of the synchronisability indices obtained from the MSF approach is that they do not take into account the dynamics of the oscillators and resonators since their computation is based on the eigenvalues of the Laplacian matrix. This observation has also been stated in Arenas et al. (2008); Huang et al. (2009). Neither the type I (the ratio of extreme nonzero eigenvalues λ_N/λ_2) nor the type II (the smallest positive eigenvalue λ_2) synchronisation indices, which are the main measures of synchronisability for complex networks, best characterise the synchronisability in the square lattice of Duffing resonators and van der Pol oscillators as the number of random links is increased but rather the largest eigenvalue λ_N of the Laplacian matrix. For heterogeneous coupling strengths in the square lattice, the synchronisability of the system depends on its average intensity.

The global and local synchronisation studies, in this chapter and the previous one, can give totally different results which implies that they can provide complementary in-

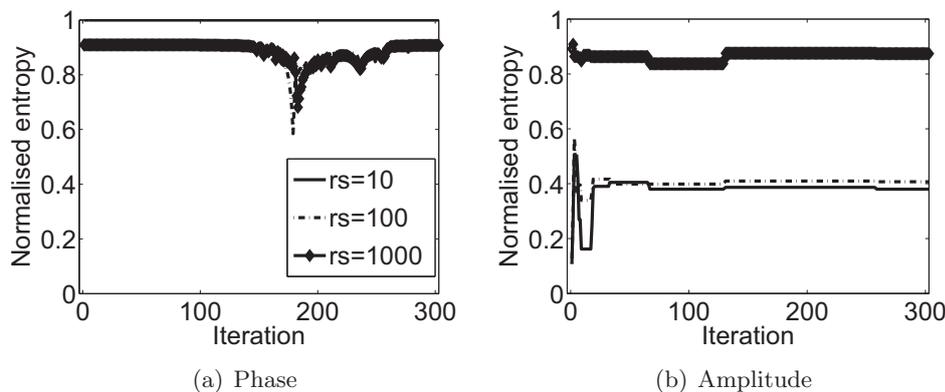


Figure 5.11: Effect of the number of random driving sites rs on the global coherence of the square lattice of 32 by 32 x -coupled alternating Duffing resonators and van der Pol oscillators. Figure 5.11(a) shows that driving the array from randomly selected sites has no effect on the global phase coherence while the global amplitude coherence is improved with more driving sites as shown in Figure 5.11(b).

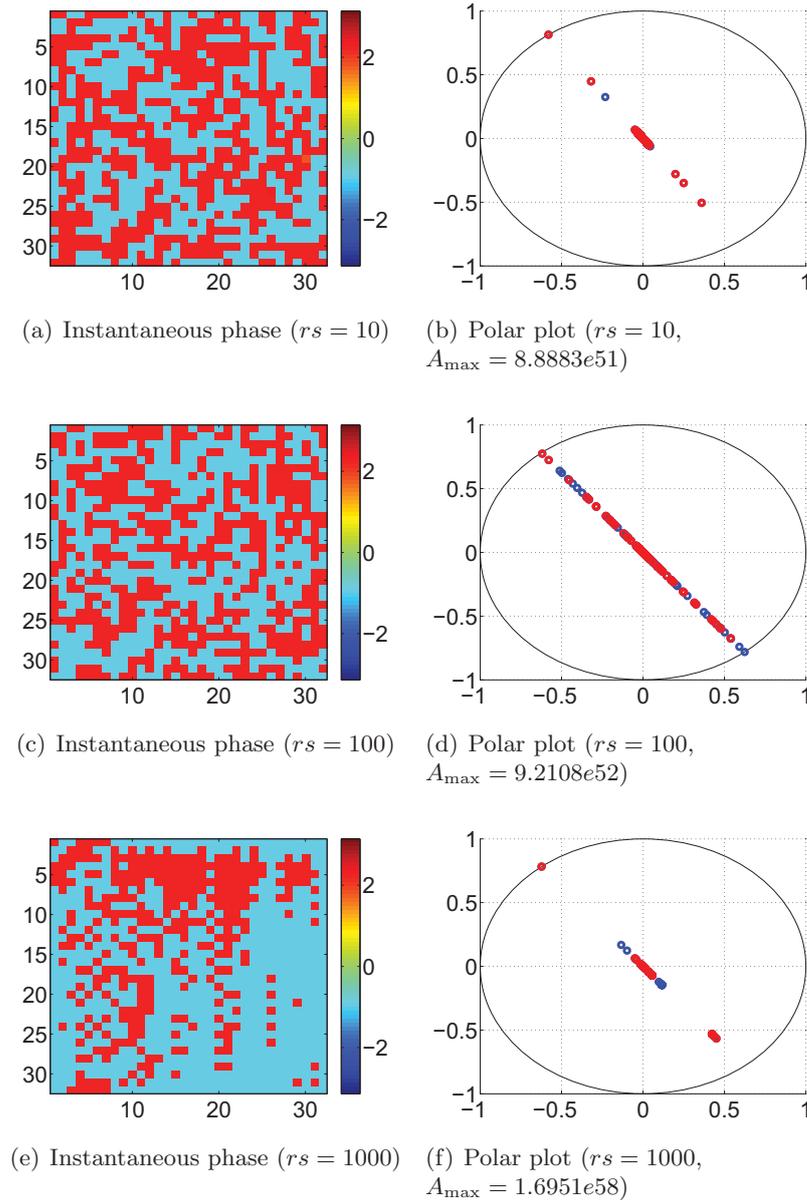


Figure 5.12: Instantaneous phase and normalised polar plots of a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 3$. The figures are plotted at $t = 3600$ with different number rs of randomly selected sites as the source of the driving. Figures 5.12(a)-5.12(f) show that the procedure of driving the square lattice from randomly selected sites has the effect of locking the phases of the elements in two different states: $\frac{3\pi}{4}$ and $-\frac{\pi}{4}$. However the amplitudes of the elements are not affected.

interpretations of the collective behaviour of the elements in the square lattice. Another aspect of synchronisation which can give more useful information than simply looking at the full synchronisation, especially for a lattice of heterogeneous systems, is called *cluster synchronisation* and will be discussed in the next chapter.

This chapter has focused on some of the emergent behaviour characteristics to be

expected from our coupled dynamical system in the presence of fixed randomness. In Chapter 6 we will return to an analysis of this behaviour using the AN model to detect synchronisation through emergent clusters.

Cluster synchronisation in a regular network of microelectromechanical system elements

6

In this chapter we provide a characterisation of cluster synchronisation in the square lattice of coupled Duffing resonators and van der Pol oscillators that was studied in the two previous chapters. The emergent collective behaviour of coupled oscillators is characterised by the phenomenon of full synchronisation as the simplest mode of spatiotemporal behaviour in a lattice of coupled oscillatory systems. In that case, the synchronisation of the elements in the lattice is studied globally and the regime of global synchronisation is associated to a stable equilibrium state of the motions of the oscillators (Osipov and Sushchik, 1997, 1998). However, more complicated synchronous behaviour than only full synchronisation has been observed based on cluster synchronisation manifolds (Belykh et al., 2003b,a). In this chapter we are interested in the characterisation of novel self-organising response patterns. The interaction of nonlinear oscillators in the square lattice allows the emergence of a much richer range of behaviours of the array beyond that expected in a traditional design strategy, where an array of isolated and linear-response sensors are present. These

behaviours include macroscopic and cluster synchronisation. These different modes of operation could in the future be used for adaptive integrated sensing and computing.

6.1 Overview

Cluster synchronisation has been defined in different ways and in terms of different system parameters such as the frequency, phase and amplitude. A cluster is defined as a coupled set of oscillators having the same average quantity (period, mean frequency) in Osipov and Sushchik (1998); Lemaître and Chaté (1999); Rabinovich et al. (1999). In globally coupled identical systems, a cluster is characterised by the vanishing difference among the elements, i.e., $|x_i - x_j| = 0$, where x_i and x_j represent the phases of the oscillators at position i and j respectively (Zhou et al., 2002). In functional magnetic resonance imaging, cluster synchronisation consists of identifying voxels in the brain which were synchronised with the voxels in the cluster that were active during a certain task. A phase synchronisation map is obtained for each voxel in the original cluster that was analysed after the recording of a synchronisation index for each voxel in the brain (Laird et al., 2002). Clusters of oscillators with infinitesimal amplitude can also be formed and this effect is called “oscillator death” or “amplitude death” (Yamaguchi and Shimizu, 1984; Bar-Eli, 1985; Ermentrout, 1986; Ermentrout and Troy, 1986; Ermentrout, 1990; Aronson et al., 1990). This type of cluster synchronisation approach has also been applied to two-coupled chaotic systems as in Rosenblum et al. (1996); Osipov and Kurths (2001) for example. And finally, the cluster synchronisation phenomenon is present in a lattice of coupled oscillators when the oscillators synchronise in groups but there is no synchronisation among the groups (Belykh et al., 2003b; Kaneko, 1990; Belykh and Mosekilde, 1996; Xie and Hu, 1997; Hasler et al., 1990; Belykh et al., 2000; Belykh and Mosekilde, 2001).

Cluster synchronisation has been studied in lattices of coupled identical periodic (Belykh et al., 2000; Belykh and Mosekilde, 2001; Belykh et al., 2003b) and chaotic oscillators (Afraimovich et al., 1997; Osipov and Sushchik, 1997; Chiu et al., 2001; Zhou et al., 2002; Belykh et al., 2003a). The analytical study is based upon bounded dissipativeness (Afraimovich et al., 1997) and linear invariant manifold theory (Belykh et al., 2003a) where the oscillators are synchronised in clusters defined by hyperplanes. In that case, the clusters synchronise in pairs symmetrically to the middle of the row and/or column of the lattice or with respect to the principal and/or secondary diagonals. The few other theoretical works on full synchronisation that exist only consider homogeneous systems such as an array of van der Pol oscillators. Among those theoretical studies are the con-

traction (Lohmiller and Slotine, 1998; Slotine and Lohmiller, 2001), partial contraction analysis (Wang and Slotine, 2004) and a few other mathematical results in the literature (Fujisaka and Yamada, 1983; Afraimovich et al., 1986; Rodrigues, 1994; Afraimovich and Rodrigues, 1994; Wu and Chua, 1994). In the above previous works, the oscillators in the lattice are of the same type, that is, they have the same dynamics. The nonidentical case is characterised by the introduction of a parameter mismatch but the oscillators still have the same equations of motion. The analytical study in the strictly identical case was conducted for mathematical convenience and hence does not reflect the physical reality. Furthermore, a restriction on the type of invariant manifold needs to be applied in the case where the oscillators are slightly nonidentical (Belykh et al., 2003a). As a matter of fact, the analytical treatment of synchronisation in lattices of coupled oscillators is rather complex and hence the lack of rigorous mathematical results in the literature as pointed in Afraimovich et al. (1997). The study of synchronisation in those lattices has therefore been confined to numerical analysis in which the focus was put on spatial features presented by stable equilibria of the oscillators in the lattice (Belykh et al., 2003b; Chow and Mallet-Paret, 1995; Nekorkin et al., 1997; Thiran, 1997).

In our case, where two different types of dynamical systems coexist in the lattice, interesting cluster synchronisation patterns could not therefore be described by these previous analysis. In order to characterise those unusual patterns, the Affinity Network is applied to learn those emergent patterns from the dynamical properties of the elements of the lattice.

6.2 Cluster synchronisation and Affinity Network

In contrast to conventional methods which study the collective behaviour of coupled oscillators at the *physical level*, using the phases of the oscillators such as in the previous section, for example, the Affinity Network (AN) that is learned from the physical characteristics of the oscillators would give an idea of their collective behaviour at a higher level: a *logical* one.

In this framework, each oscillator is regarded as one random continuous variable and the time series of its displacement and instantaneous phase can be viewed as the samples from those random variables. The displacements are taken directly from the numerical integration of the differential equations and the instantaneous phases are computed from the analytic signal obtained using the Hilbert transform (S. L Marple, 1999; Boashash, 1992a,b; Goswami and Hoefel, 2004). The AN structure that is learned from the time

series of the displacements and/or instantaneous phases of the oscillators will be such that each parent node and its child(ren) in the network can be considered as one cluster. The oscillators in each cluster can then be interpreted as forming one unit where these oscillators would have a similar behaviour. For concision, the AN structures will only be learned from the phase data although the displacement data can also be used.

6.2.1 Experimental results

The time series produced from the simulations in the two previous chapters are used in this section to compute the input similarity matrix in the AN construction algorithm. The purpose of the following experiments is to characterise the collective behaviour of the elements in the square lattice by locating clusters of synchronised elements. Moreover, we would like to evaluate the clustering capability of the AN against the DN and a standard clustering technique. In order to appreciate the clustering results, they are plotted together with the asymptotic instantaneous phase data obtained in the two previous chapters. In addition, to quantify the cluster synchronisation results, the correlation coefficient between the AN or DN cluster image and the corresponding instantaneous phase image is computed. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient.

The parameter \tilde{p} in the SCAP algorithm has been set to the largest negative number in Matlab. Each resonator and oscillator are regarded as random variables. Their instantaneous phase values, from an initial time $t_0 = 0$ to a time $t = 3600$, are considered as samples. Since the time for the series to converge varies between experiments, as shown by the normalised entropy plots in the two previous chapters, the first 900 samples have been discarded. These instantaneous phase data are then used as inputs to the AN algorithm to generate the AN cluster images where the resonators and oscillators which belong to the same cluster have the same colour.

Figure 6.1 shows that both the AN and the DN are not able to capture the phase patterns that are displayed in the lattice with different coupling strengths. In particular, the AN tends to group the elements in the lattice in one cluster as shown by the AN clusters and histograms in Figures 6.1(a) to 6.1(d). Likewise, the DN clusters in Figure 6.1(e) and 6.1(g) clearly do not reflect the row synchronisation as it appears that the elements are synchronised across columns. Moreover, the number of DN clusters, as shown in Figures 6.1(f) and 6.1(h) are far more than the number of actual phase values which can be observed from the phase images in Figures 6.1(i) and 6.1(j).

The checkerboard-like and ordered behaviours that are observed for low ($Q = 10$) and

high (1000) Q respectively are made clear from the AN clustering results. In particular, the AN clusters and histogram in Figures 6.2(c) and 6.2(d), respectively, show that most of the elements are grouped in one cluster. The corresponding phase image in Figure 6.2(j) agrees with this observation since the phase is circular and hence the elements coloured in blue and red are close in phase space. On the other hand, the DN clusters in Figure 6.2(e) and 6.2(g) tend to show synchronisation across columns which is in contradiction with the phase images in Figure 6.2(i) and 6.2(j) respectively. The DN histograms in Figure 6.2(f) and 6.2(h) show that in the case of different Q values, the DN clusters still exceeds the actual number of different phase values that are observed from the phase images in Figures 6.2(i) and 6.2(j).

A rather atypical behaviour is observed in the system when the external driving force is injected from one corner of the lattice since a symmetrical behaviour along the diagonal joining the corner from which the energy is injected and its opposite corner is present. In spite of this different ordered behaviour, the AN clustering can still reflect this symmetry as shown in Figure 6.3: when the square lattice is driven from its top left corner, the phases of the elements are symmetric along the main diagonal as shown in Figure 6.3(e). The same diagonal represents the axis of symmetry of the corresponding AN clusters in Figure 6.3(a). Similarly, it can be said that the DN reflects the fact that the array is driven from the top left corner in the sense that a line representing the main diagonal can be observed in Figure 6.3(c). However, a symmetric behaviour along the main diagonal cannot be observed as the DN clusters display synchronisation across rows. The AN and DN histograms in Figure 6.3(b) and 6.3(d), respectively, show that both the AN and the DN do not give sensible number of clusters compared to the phase image in Figure 6.3(e) where around five clusters can be observed.

The AN has also been tested on the random networks of resonators and oscillators and Figure 6.4 shows that the clusters of synchronisation captured by the algorithm are sensible when completely random patterns occur. Unlike the AN clusters, the random phase patterns shown in Figure 6.4(e) are not captured by the DN clusters in Figure 6.4(c) where a noisy synchronisation across columns can be observed. This is also confirmed by the AN clusters in Figure 6.5 where randomly selected sites are chosen as the source of the driving signal in the square lattice. The AN clusters in Figures 6.5(e) and 6.5(g) are very close to the phase images in Figures 6.5(i) and 6.5(j) respectively whereas the DN clusters are less accurate. This is confirmed by the AN histograms in Figures 6.5(f) and 6.5(h) where the two AN clusters exactly match with the two different phase values in Figures 6.5(i) and 6.5(j) whereas the DN clusters are slightly more.

The standard k-means clustering algorithm with the same number of centres as are found in the images has also been applied in the experiments in order to independently validate the AN clustering results. The correlation measure ρ has been computed to evaluate the results. It has been found that sometimes k-means can outperform the AN. This is illustrated by the results in Figure 6.6 where the k-means with seven centres gives more meaningful clusters than the AN when the array of oscillators and resonators is driven from the top-left corner. However, the k-means with two centres performs poorly on the array that is randomly driven at 1000 different sources whereas the AN produces sensible clusters. The average correlation $\bar{\rho}$ over the different graphs in the experiments of this chapter has been computed in order to compare the overall performance of the different methods: $\bar{\rho}_{\text{AN}} = 0.39$, $\bar{\rho}_{\text{DN}} = 0.18$ and $\bar{\rho}_{\text{k-means}} = -0.17$ for the AN, DN and k-means, respectively.

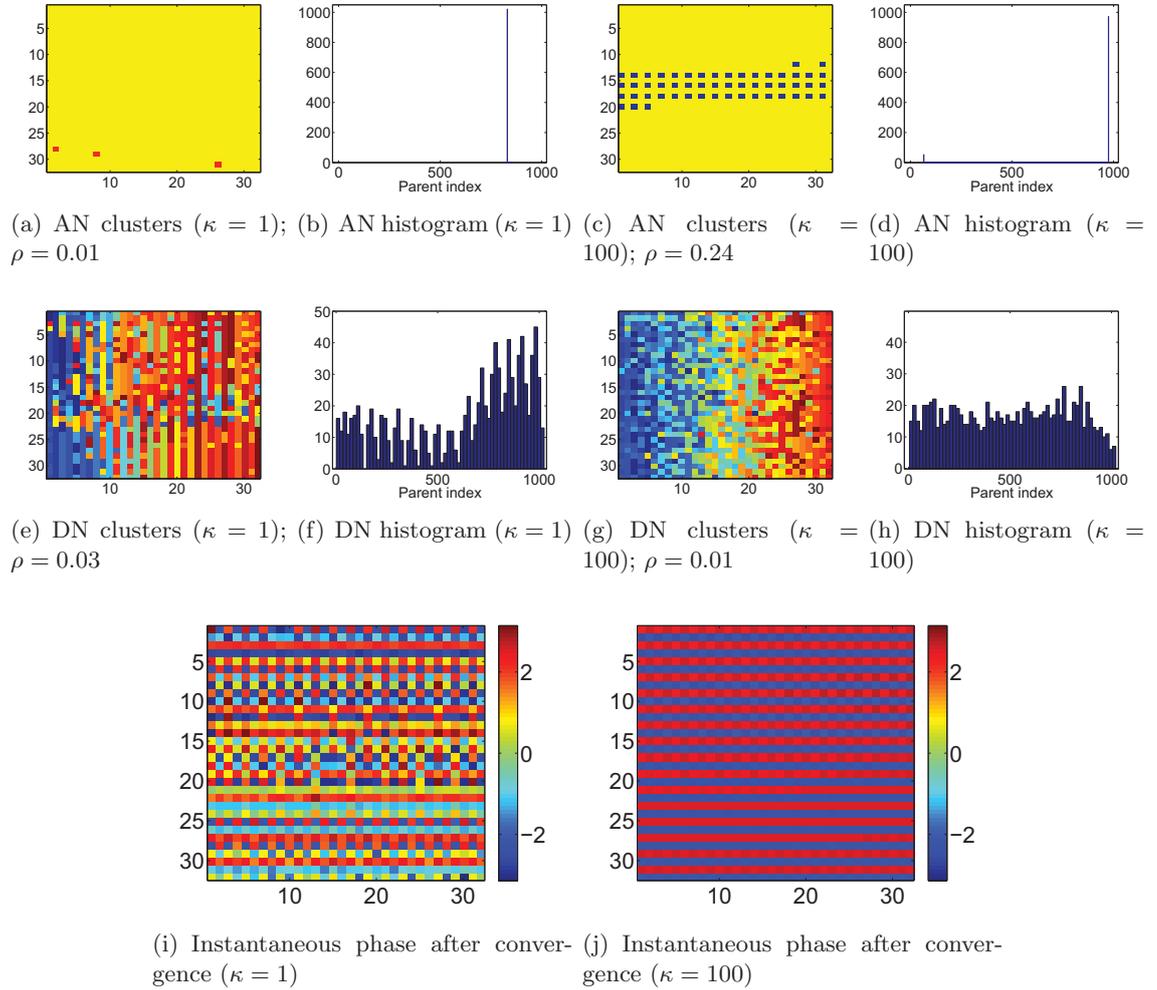


Figure 6.1: Characterisation of cluster synchronisation using AN and DN with varying coupling strength κ in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. ρ represents the correlation coefficient between the AN or DN cluster image and the corresponding instantaneous phase image. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient. Both the AN and the DN are not able to capture the phase patterns that are displayed in the lattice with different coupling strengths. In particular, the AN tends to group the elements in the lattice in one cluster as shown by the AN clusters and histograms in Figures 6.1(a) to 6.1(d). Likewise, the DN clusters in Figure 6.1(e) and 6.1(g) clearly do not reflect the row synchronisation as it appears that the elements are synchronised across columns. Moreover, the number of DN clusters, as shown in Figures 6.1(f) and 6.1(h) are far more than the number of actual phase values which can be observed from the phase images in Figures 6.1(i) and 6.1(j).

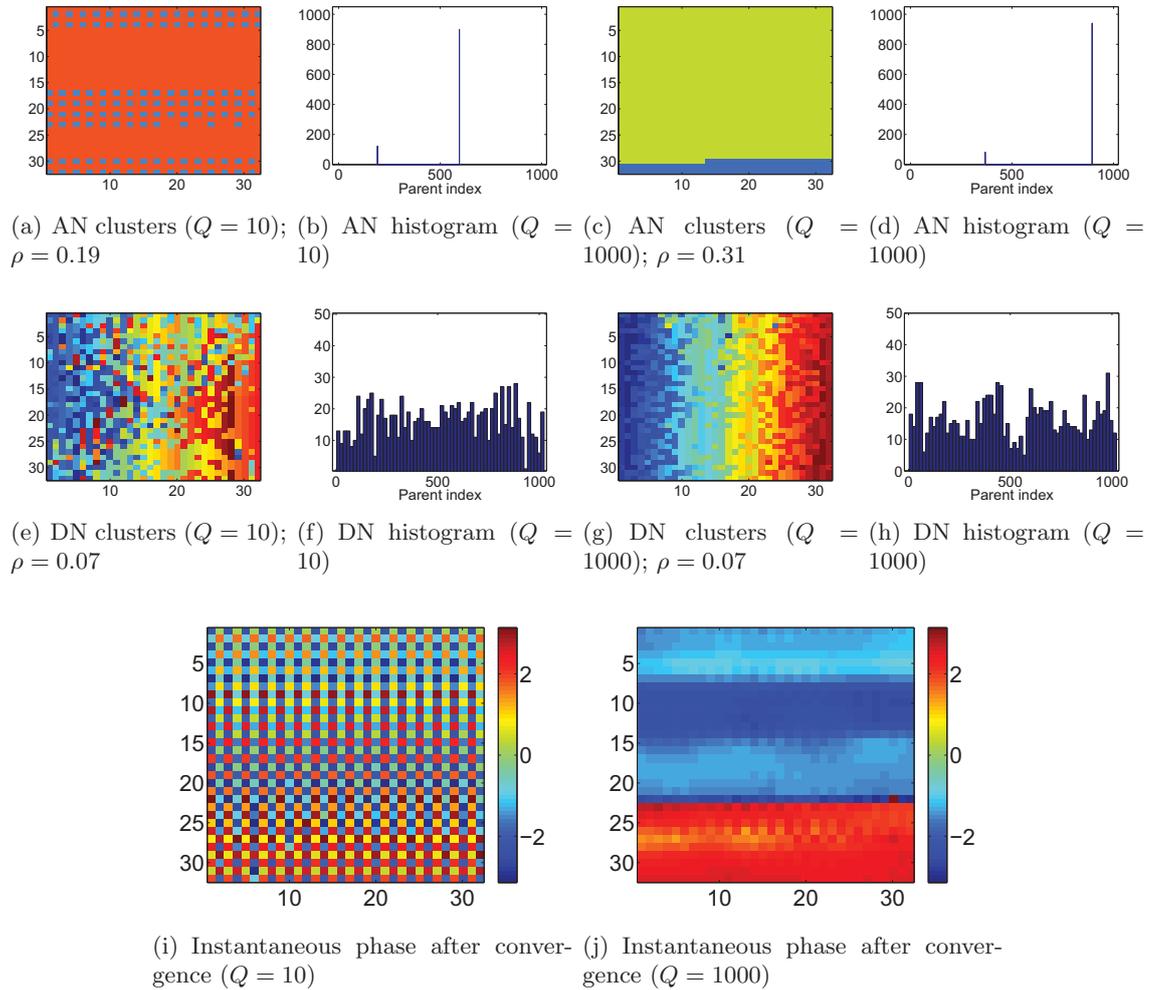


Figure 6.2: Characterisation of cluster synchronisation using AN and DN with varying damping through the quality factor Q in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. ρ represents the correlation coefficient between the AN or DN cluster image and the corresponding instantaneous phase image. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient. Similarly to the previous figure, the checkerboard-like and ordered behaviours that are observed for low ($Q = 10$) and high (1000) Q respectively are made clear from the AN clustering results. In particular, the AN clusters and histogram in Figures 6.2(c) and 6.2(d), respectively, show that most of the elements are grouped in one cluster. The corresponding phase image in Figure 6.2(j) agrees with this observation since the phase is circular and hence the elements coloured in blue and red are close in phase space. On the other hand, the DN clusters in Figure 6.2(e) and 6.2(g) tend to show synchronisation across columns which is contradiction with the phase images in Figure 6.2(i) and 6.2(j) respectively. The DN histograms in Figure 6.2(f) and 6.2(h) show that in the case of different Q values, the DN clusters still exceeds the actual number of different phase values that are observed from the phase images in Figures 6.2(i) and 6.2(j).

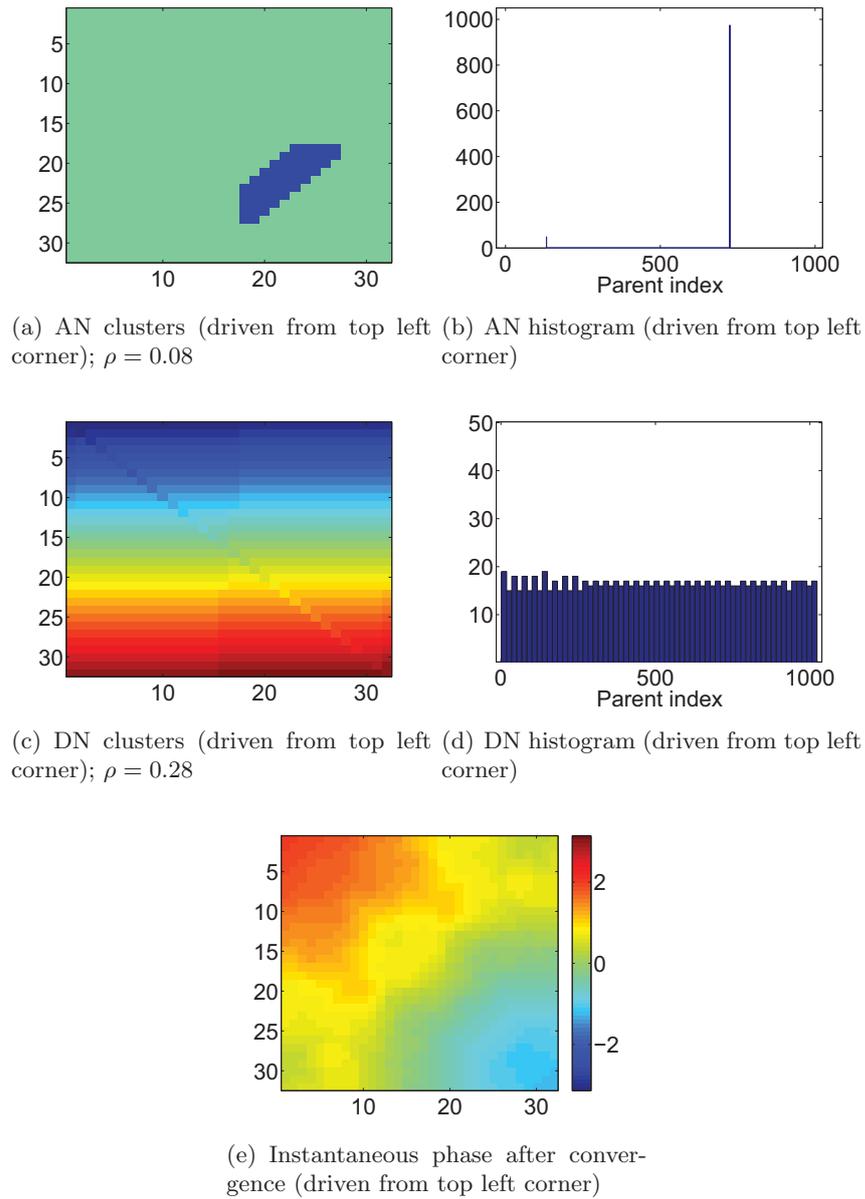


Figure 6.3: Characterisation of cluster synchronisation using AN and DN with a different location of the external driving force in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators. ρ represents the correlation coefficient between the AN or DN cluster image and the corresponding instantaneous phase image. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient. When the square lattice is driven from its top left corner, the phases of the elements are symmetric along the main diagonal as shown in Figure 6.3(e). The same diagonal represents the axis of symmetry of the corresponding AN clusters in Figure 6.3(a). Therefore, the symmetrical behaviour of the system is also captured by the AN algorithm. Similarly, it can be said that the DN reflects the fact that the array is driven from the top left corner in the sense that a line representing the main diagonal can be observed in Figure 6.3(c). However, a symmetric behaviour along the main diagonal cannot be observed as the DN clusters display synchronisation across rows. The AN and DN histograms in Figure 6.3(b) and 6.3(d), respectively, show that both the AN and the DN do not give sensible number of clusters compared to the phase image in Figure 6.3(e) where around five clusters can be observed.

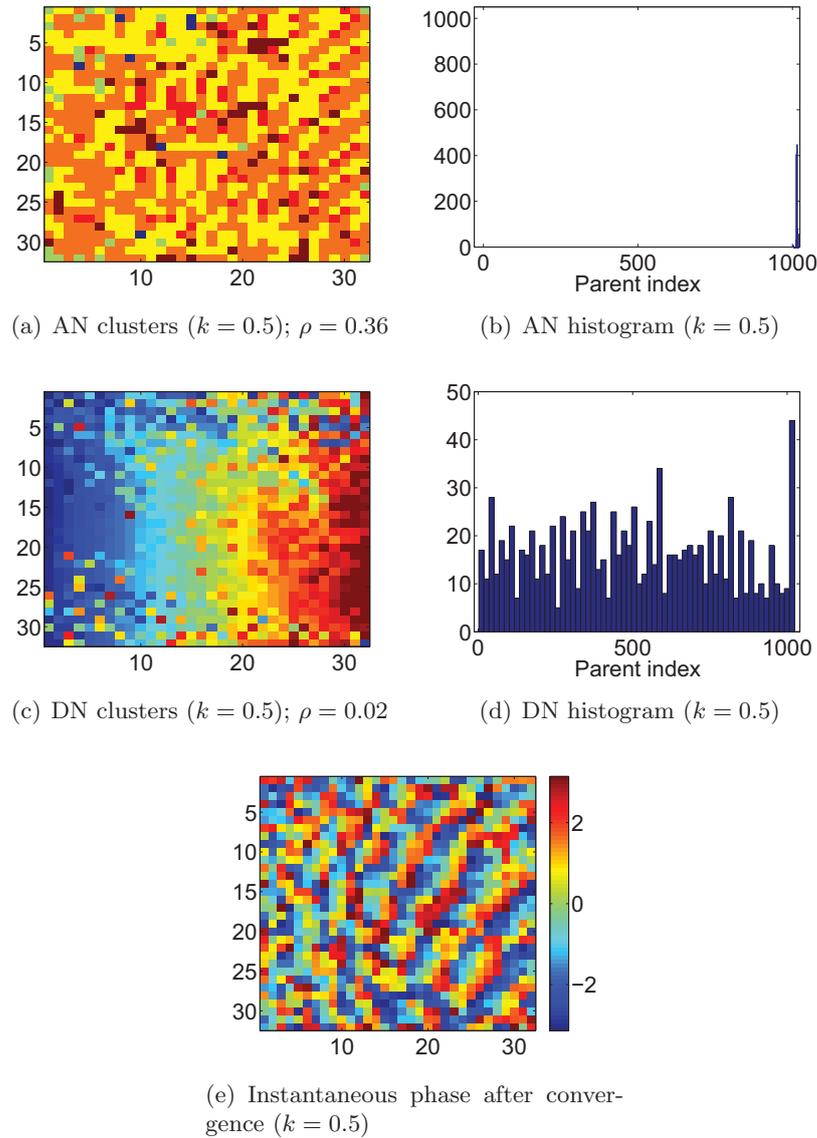


Figure 6.4: Characterisation of cluster synchronisation using AN and DN in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 0.3$. ρ represents the correlation coefficient between the AN or DN cluster image and the corresponding instantaneous phase image. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient. Figure 6.4(e) is plotted at $t = 3600$ with 11 extra uniform random links which corresponds to 1% of the total number of elements in the lattice. The coefficient k represents the fraction of the coupling strength κ of the non-random four nearest neighbour connections which is $\kappa = 10$ here. In this case, the AN algorithm produces six sensible clusters. Unlike the AN clusters, the random phase patterns shown in Figure 6.4(e) are not captured by the DN clusters in Figure 6.4(c) where a noisy synchronisation across columns can be observed. In addition, the number of DN clusters as shown by the histogram in Figure 6.4(d) is far beyond the number of different phase values in the phase image of Figure 6.4(e).

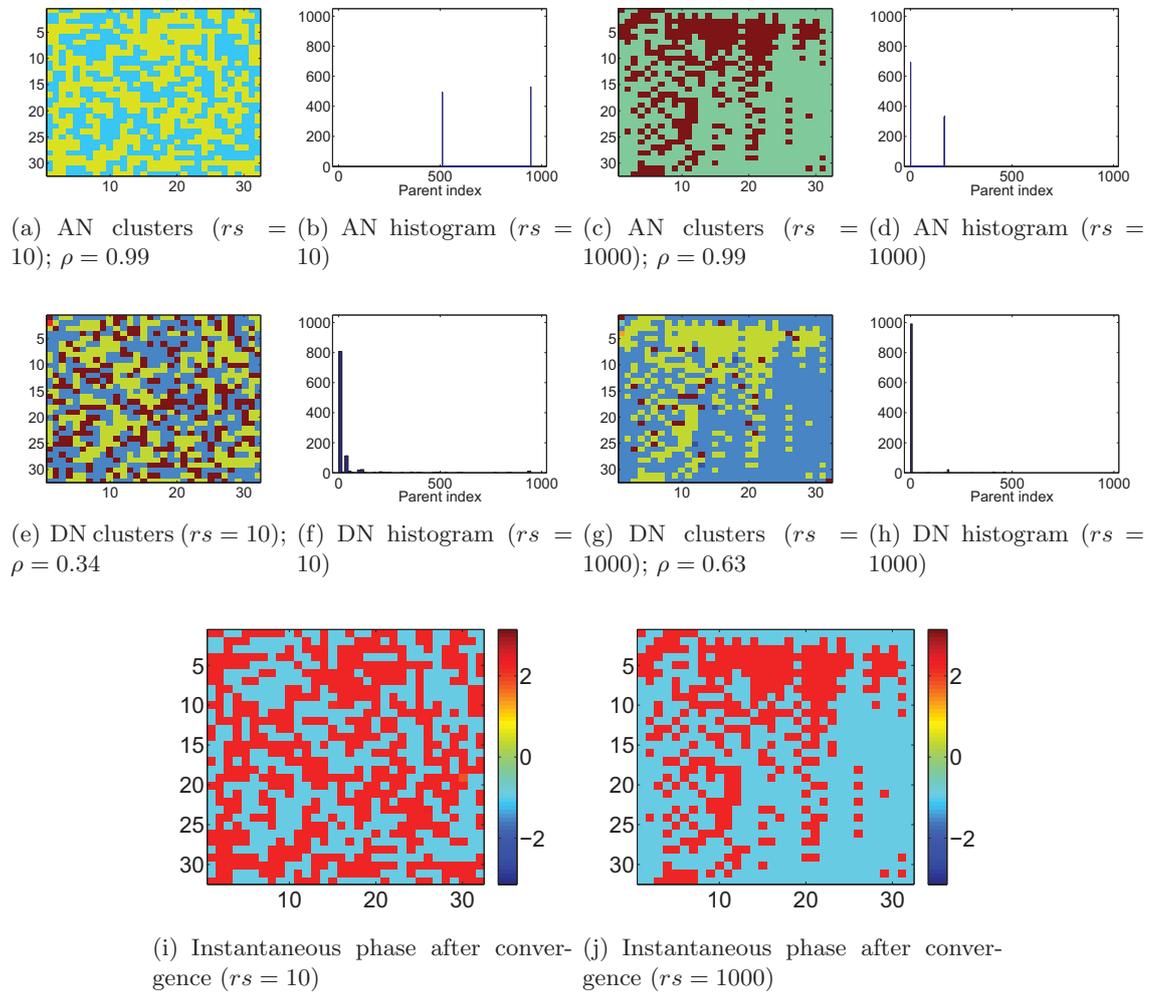


Figure 6.5: Characterisation of cluster synchronisation using AN and DN with different number rs of randomly selected sites as the source of the driving in a square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators driven from the top edge with an energy whose amplitude is $F_d = 3$. ρ represents the correlation coefficient between the AN or DN cluster image and the corresponding instantaneous phase image. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient. The AN clusters in Figures 6.5(e) and 6.5(g) are very close to the phase images in Figures 6.5(i) and 6.5(j) respectively whereas the DN clusters are less accurate. The AN histograms in Figures 6.5(f) and 6.5(h) show that the two AN clusters exactly match with the two different phase values in Figures 6.5(i) and 6.5(j) whereas the DN clusters are slightly more.

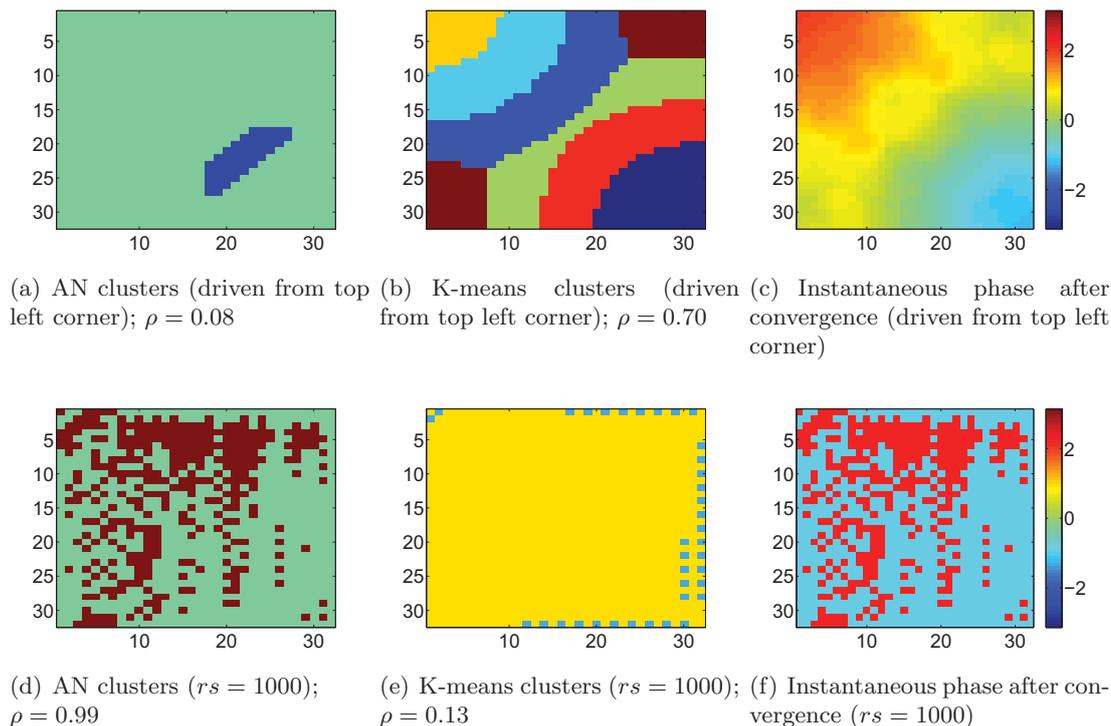


Figure 6.6: Comparison between AN and k-means. In Figures 6.6(b)-6.6(c), the square lattice of 32 by 32 x-coupled alternating Duffing resonators and van der Pol oscillators is driven from its top left corner with an energy whose amplitude is $F_d = 3$. In Figures 6.6(e)-6.6(f), different number rs of randomly selected sites as the source of the driving are applied on the square lattice. ρ represents the correlation coefficient between the AN or k-means cluster image and the corresponding instantaneous phase image. Each cluster and instantaneous phase images have been rescaled before computing the correlation coefficient. The k-means with seven centres (Figure 6.6(b)) gives more meaningful clusters than the AN (Figure 6.6(a)) when the array of oscillators and resonators is driven from the top-left corner. However, the k-means with two centres (Figure 6.6(e)) performs poorly on the array that is randomly driven at 1000 different sources whereas the AN produces sensible clusters (Figure 6.6(d)).

6.3 Summary

The application of the Affinity Network (AN) as an inference, or interpreting, logic layer to analyse the physical output of the dynamical system lattice was examined in this chapter since the analytic investigation in such a large system is complex and most theoretical studies deal with homogeneous systems. It was found that the clustering ability of the AN can produce sensible results on the structures of the emergent patterns across the lattice, even when the clusters were fragmented and distributed. This is evidenced by the correlation coefficient as a quantitative measure of the clustering results. In contrast, the DN clusters tend to show phase synchronisations across columns when synchronisation across rows was displayed by the actual phase images. Moreover, less accurate DN clusters

are observed when random links and driving sites are introduced in the lattice. However, the AN can also fail to characterise well the cluster synchronisation in the lattice. For instance, the AN was not able to capture the symmetric phase pattern which occurs when the array is driven from the top left corner. In that case, a standard clustering method, such as k-means, can be applied.

In sensing applications it would be an enormous advantage to be able to predict accurately the behaviour of the system depending on its parameter values and the characteristics of the driving signal that is applied to it. Unfortunately, no analytical tool is currently available to achieve this purpose for lattices of coupled oscillators to the best of our knowledge and further studies are therefore needed.

7 Conclusion

This thesis has considered problems of inference and macroscopic clustering derived from lower physical or data layers with inherent similarity relationships. We considered both non-model based problems such as discovering relationships in medical data, and a strongly physics-oriented problems of dynamical systems synchronisation. To provide a statistical analysis level of these problems, a new class of graphical model called an Affinity Network (AN) was introduced and compared with other state-of-the-art models, and in particular the Dependency Network and Bayesian graphical model.

7.1 Graphical model and existing instances

A graphical model is characterised by a *network structure*, which is a graph where each node corresponds to a random variable from a dataset, and some *parameters* represented by the conditional probabilities of the variables. Bayesian Networks (BNs) and Dependency Networks (DNs) are the main instances of a graphical model. BNs usually display causal relationships between random variables whereas DNs were devised to relax this hard constraint associated with the causal relationships in BNs. Since BNs do not have cycles, they cannot discriminate spurious causal relations between confounding variables.

In contrast, AN and DN graphs can be cyclic and can therefore capture more general dependency relationships.

The structure of a BN is usually learned by combining the BDe scoring function with a hill-climbing heuristic search although a conditional independence test can also be used to construct a BN structure. On the other hand, a classification/regression algorithm governs the structure learning of a DN. The conjugacy property of the Dirichlet distribution for discrete variables, and the Normal distribution properties for continuous variables enable the computation of the conditional probabilities in a BN. In contrast, these conditional probabilities are directly estimated from the data using a classification or regression techniques in a DN. The structure and parameters of a BN and a DN are both learned from data although expert knowledge can also be used to specify the conditional probabilities in a BN.

7.2 The AN graphical model

An AN is a directed, possibly cyclic, graphical model in which the nodes represent random variables and the links between those random variables denote probabilistic relationships based on similarities between two connected nodes. The AN shares some properties of a BN and a DN in the sense that the conditional probabilities of an AN are learned using a similar approach to learn the conditional probabilities of a BN. In addition, the joint probability of an AN can be computed the same way as in a DN. The AN differs from the BN and the DN in its graph learning, also called network structure learning.

The AN graphical model structure learning algorithm uses the soft-constraint affinity propagation (SCAP) algorithm combined with the Bayesian Dirichlet (BD) scoring function. The SCAP algorithm is derived from affinity propagation (AP) which is an optimisation algorithm that is used to drive the AN structure search. SCAP is able to reconstruct the possible loops in the network structure contrary to AP because of its soft constraints. The parameter learning algorithm for an AN is similar to a BN parameter learning.

Contrary to a BN, an AN and a DN do not necessarily represent a factorisation of the joint distribution of the data. In a DN, inconsistencies are likely to occur when there is no joint distribution from which the conditional probabilities may be obtained via the rules of probability. However, when the sample size is large, inconsistencies will be rare given that the conditional distribution is learned from the same data set, which is assumed to be generated from a single underlying distribution. Hence, when the data has many

samples, an AN structure, similarly to a DN, can also be seen as a representation of the joint distribution of the data since the conditionals are directly learned from the data. A DN displays the dependencies between variables whereas an AN shows the closeness between pairs of variables with respect to the chosen scoring function.

7.3 AN for medical data

The proposed AN approach has also been compared with the DN where a classification tree is used to construct the DN structure. The combination of SCAP and the BD score to construct an AN always recovers the essential graph, that is, the undirected graph of the benchmark networks obtained from an exhaustive search whereas the classification tree adds more parents than required. In particular, based on a comparison with an exhaustive structure search, the AN structure, which mainly had reversed arcs, was found better than the DN structure, which had more added and deleted arcs on the Chest, Ferrara and Wisconsin breast cancer data. However, when only the root of the tree is taken as the parent of the corresponding child variable, the classification tree performs better than the normal DN. The AN and DN networks of the Ferrara breast cancer data were particularly interesting because they respectively showed that the progesterone (PR) and oestrogen (ER) receptors are the main variables in this dataset. This fact is sensible since breast cancer is considered to be a hormone-dependent cancer and that the balance between these two hormones is vital to the health of every woman.

SCAP is a good alternative to heuristic or stochastic search algorithms that are used in BNs and DNs since the SCAP algorithm is based on a message-passing scheme. Furthermore, the algorithm has been proved to be efficient, fast and especially suited for huge data such as the medical alarm network where the proposed method is faster than a classification tree. It was observed that the AN structure of the alarm network was constructed in 20 seconds with SCAP and the BDe score, and in 135 seconds with a DN and a classification tree on a machine with two 2.40 GHz Intel Xeon processors with 512 KB memory each. The corresponding BN structure was constructed in 998 seconds on a Macintosh II running LightSpeed Pascal version 2.0. Moreover, this alternative method to learn an AN structure provides extra information by quantifying the strength of each arc in the network and also by providing a simple way to analyse the sensitivity of the scoring function used in conjunction with the SCAP algorithm.

7.4 AN of MEMS dynamical data

Besides static medical data, the AN graphical model was applied to dynamic time series data from Duffing resonators and van der Pol oscillators coupled alternatively in a square lattice. Such a lattice can serve as a sensing device where the input signal is simulated with an external driving force that is injected to the Duffing resonators. The physical behaviour of the system has been studied and it has been found that it displays rich behaviours such as hysteresis, anti-resonance, stagnant response and multiple branches in their frequency responses which have been obtained using the method of multiple time scales analysis. The stability of the system has also been investigated where two unstable branches of the Duffing resonator and van der Pol oscillator respectively have been found. Moreover the system also displays interesting transient and asymptotic behaviour. The bifurcation analysis also revealed an abrupt transition to chaos of the system. The global and local synchronisations of the square lattice of up to 1024 devices was also studied numerically. The parameters of the resonators and oscillators that have been used come from real micro-electro-mechanical system (MEMS) devices. The effects of different system parameters have been investigated. The square lattice that has been studied can be seen as an ideal 2D array of resonators and oscillators. However, it does not reflect a real physical device where random parameter mismatch and connections can occur as a result of the imperfections of the fabrication process. For this reason, the lattice has been called a probabilistic square lattice when random noise is present. The behaviour of the system when taking into account random noise was also briefly considered. In particular, the noise was represented by different parameter values of the elements, namely the natural frequency, the damping coefficient and the coupling strengths, instead of having identical parameters, on one hand. On the other hand, random links were added to the four nearest-neighbour connections of the square lattice and random sites were selected as the source of the driving force.

The global and local synchronisation studies can give totally different results whereas the cluster synchronisation aspect provides more useful information than simply looking at the full synchronisation especially for a lattice of heterogeneous systems. Nevertheless, no cluster synchronisation characterisation is available to the best of our knowledge as opposed to the case of global synchronisation in which various synchronisation indices can be used. Moreover, the analytic investigation in such a large system is complex and most theoretical studies deal with homogeneous systems. The application of the Affinity Network as an inference, or interpreting, logic layer to analyse the physical output of the

dynamical system lattice was examined since the analytic investigation in such a large system is complex and most theoretical studies deal with homogeneous systems. It was found that the clustering ability of the AN can produce sensible results on the structures of the emergent patterns across the lattice, even when the clusters were fragmented and distributed. This is evidenced by the correlation coefficient as a quantitative measure of the clustering results. In contrast, the DN clusters tend to show phase synchronisations across columns when synchronisation across rows was displayed by the actual phase images. Moreover, less accurate DN clusters are observed when random links and driving sites are introduced in the lattice.

7.5 Limitations of the AN and future directions

The shortcomings of the AN stem from the setting of some parameter values in the structure learning algorithm on one hand. On the other hand, the AN is mainly used to find meaningful relationships between variables. The aspect of probabilistic inference through the computation of the joint probability of the variables has received a lesser attention. This also leads to additional limitations regarding inference and control tasks in the MEMS domain.

7.5.1 AN parameters

The value of the penalty term \tilde{p} in the SCAP equations controls the AN structure. However, an automatic procedure which sets the value of \tilde{p} is not currently available. In our experiments we have set \tilde{p} to a large negative number in order to relax the AP hard constraint and to remove the possibility of having self-loops. Moreover, a node can only have one parent in the AN graph. Multiple parents can be obtained by introducing a threshold parameter on the maximum value of the affinity function. The choice of this parameter also remains an open question.

7.5.2 Inference and belief propagation

It has been discussed in this thesis that the joint distribution is not computed in an AN since the network structure may not represent a factorisation of the joint probability density function (pdf) as in a BN. Therefore, the belief propagation algorithm for BNs may not be suitable to use with ANs. Hence the inference part which is an important aspect of a graphical model is missing in the AN approach. However, the pseudo-Gibbs sampling for DNs can be used in order to answer probabilistic queries since conditional

probability tables are available in ANs. Nevertheless, it has been shown empirically in Heckerman et al. (2000) that computing the joint probability in this way is not efficient compared to belief propagation in BNs.

7.5.3 Inference in MEMS

In the context of cluster synchronisation, the AN can also fail to produce good clusters in the lattice of oscillators and resonators. In that case, a standard clustering algorithm, such as k-means, can sometimes complement the AN. In addition, it would also be interesting to be able to predict the collective behaviour of the resonators and oscillators for a given set of initial conditions and system parameters. The AN is currently unable to perform this task due to the absence of a temporal factor. The analytical study of such inference problem remains a huge challenge because of its complexity. It has been mentioned earlier that most theoretical works consider identical elements although some noise has been introduced in the system. Moreover some works such as Wang and Slotine (2004) only treat the problem from a purely theoretical point of view without any numerical validation.

7.5.4 Pinning control

Another interesting problem which is related to the square lattice of resonators and oscillators is the ability to control the cluster synchronisation patterns. This is referred to as *pinning control* where for an arbitrarily selected cluster synchronisation pattern, a controller is introduced for each cluster and sufficient conditions are derived such that the given cluster synchronisation pattern is achieved for any initial conditions (Wu et al., 2009; Li et al., 2004; Grigoriev et al., 1997; Greilich et al., 2005). The different applications of control in chaotic systems include for instance turbulence (Katz et al., 1994), instabilities in plasma (Pentek et al., 1996), multimode lasers (Colet et al., 1994) and reaction-diffusion systems (Petrov et al., 1994). Linear control theory (Romeiras et al., 1992; Petrov et al., 1995) has been used so far to address the problem of pinning control. This is therefore a topic for future research.

A AN message-update equations

In this appendix we derive the AN update equations for the availability and responsibility messages. We follow the steps in Frey and Dueck (2007) but in the context of learning an AN structure. In other words, a data point and a cluster centre in Frey and Dueck (2007) are replaced by a child node and a parent node respectively.

We recall from Section 3.1 that the problem of learning a graphical network structure can then be summarised as finding the set of parents \mathbf{PA}_i for each variable X_i , $i = 1, \dots, n$ in a dataset. To facilitate the discussion, we suppose that a variable has only one parent and the general case will be described later. The identification of the parents can also be viewed as searching over the space of valid configurations of $\mathbf{PA} = \{pa_1, \dots, pa_n\}$, where the scalar pa_i , $i = 1, \dots, n$ is the parent of X_i , so as to minimise the energy

$$E(\mathbf{PA}) = - \sum_{i=1}^n s(X_i, pa_i), \quad (\text{A.1})$$

where $s(X_i, pa_i)$ indicates how well the variable pa_i is suited to be the parent of the variable X_i . Usually the maximisation of the network similarity, S_{net} , which is the negative energy

plus a constraint function that enforces valid parent configurations, is performed:

$$S_{net}(\mathbf{PA}) = -E(\mathbf{PA}) + \sum_{k=1}^n \delta_k(\mathbf{PA}), \quad (\text{A.2})$$

$$= \sum_{i=1}^n s(X_i, pa_i) + \sum_{k=1}^n \delta_k(\mathbf{PA}), \quad (\text{A.3})$$

where $\delta_k(\mathbf{PA})$ is a penalty term that equals $-\infty$ if some variable X_k , whose parent is different from itself, has been chosen to be the parent of variable X_i and 0 otherwise. This means that whenever a variable is selected as a parent by another variable, it is not allowed to have a parent other than itself. Equation (A.3) can be represented using a factor graph \mathcal{F} in which each term in (\mathbf{PA}) is represented by a function node and each parent pa_i is represented by a variable node. Function nodes and variable nodes are connected by edges, and a variable node is connected to a function node if and only if its corresponding term depends on the variable. For example, the term $s(X_i, pa_i)$ in the above expression has a corresponding function node that is connected to the single variable pa_i and the term $\delta_k(\mathbf{PA})$ has a corresponding function node that is connected to all variables pa_1, \dots, pa_n as shown in Figure 3.1. The global function $S(\mathbf{PA})$ in the factor graph \mathcal{F} is given by the sum of all the functions represented by function nodes. The max-sum algorithm, which is the log-domain version of the max-product algorithm, can be used to search over configurations of the parents \mathbf{PA} in the factor graph that maximise $S(\mathbf{PA})$. The max-sum algorithm for the factor graph \mathcal{F} consists of recursively sending messages from variables to functions and from functions to variables as follows:

- the message sent from pa_i to $\delta_k(\mathbf{PA})$ consists of n real numbers $\begin{pmatrix} \rho_{i \rightarrow k}(X_1) \\ \vdots \\ \rho_{i \rightarrow k}(X_n) \end{pmatrix}$ (Figure 3.1(b)).

- The message sent from $\delta_k(\mathbf{PA})$ to pa_i also consists of n numbers $\begin{pmatrix} \alpha_{i \leftarrow k}(X_1) \\ \vdots \\ \alpha_{i \leftarrow k}(X_n) \end{pmatrix}$ (Figure 3.1(c)).

The value of pa_i can be estimated at any time by summing all incoming availability α and similarity S messages (Figure 3.1(d)). The responsibility ρ messages are computed as the element-wise sum of all incoming messages because they are outgoing from variables:

$$\rho_{i \rightarrow k}(pa_i) = s(X_i, pa_i) + \sum_{k': k' \neq k} \alpha_{i \leftarrow k'}(pa_i). \quad (\text{A.4})$$

The sum of the incoming messages and the maximum over all variables except the variable the message is being sent to, give the messages sent from functions to variables. For instance, the message sent from the penalty function δ_k to the parent variable pa_i is:

$$\alpha_{i \leftarrow k}(pa_i) = \overbrace{\max_{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n} \left[\delta_k(X_1, \dots, X_{i-1}, pa_i, X_{i+1}, \dots, X_n) + \sum_{i': i' \neq i} \rho_{i' \rightarrow k}(X_{i'}) \right]}^{\text{best possible configuration satisfying } \delta_k \text{ given } pa_i}.$$

$$= \left\{ \begin{array}{l} \text{best configuration with or without parent } X_k \\ \sum_{i': i' \neq k} \max_{j'} \rho_{i' \rightarrow k}(X_{j'}) \quad , \text{ for } pa_i = k = i \\ \text{best configuration with no parent } X_k \\ \sum_{i': i' \neq k} \max_{j': j' \neq k} \rho_{i' \rightarrow k}(X_{j'}) \quad , \text{ for } pa_i \neq k = i \\ X_k \text{ is a parent} \quad \text{best configuration of others} \\ \rho_{k \rightarrow k}(X_k) + \sum_{i': i' \notin \{i, k\}} \max_{j'} \rho_{i' \rightarrow k}(X_{j'}) \quad , \text{ for } pa_i = k \neq i \\ \text{best configuration with no parent } X_k \\ \max \left[\max_{j': j' \neq k} \rho_{k \rightarrow k}(X_{j'}) + \sum_{i': i' \notin \{i, k\}} \max_{j'} \rho_{i' \rightarrow k}(X_{j'}) \right] \\ \text{best configuration with a parent } X_k \\ \rho_{k \rightarrow k}(X_k) + \sum_{i': i' \notin \{i, k\}} \max_{j'} \rho_{i' \rightarrow k}(X_{j'}) \quad , \text{ for } pa_i \neq k \neq i \end{array} \right. \quad (\text{A.5})$$

The vector messages $\begin{pmatrix} \rho_{i \rightarrow k}(X_1) \\ \vdots \\ \rho_{i \rightarrow k}(X_n) \end{pmatrix}$ and $\begin{pmatrix} \alpha_{i \leftarrow k}(X_1) \\ \vdots \\ \alpha_{i \leftarrow k}(X_n) \end{pmatrix}$ are easier to interpret if they

are viewed as the sum of constant and variable components (with respect to pa_i), that is,

$$\rho_{i \rightarrow k}(pa_i) = \tilde{\rho}_{i \rightarrow k}(pa_i) + \bar{\rho}_{i \rightarrow k}, \quad (\text{A.6})$$

and

$$\alpha_{i \leftarrow k}(pa_i) = \tilde{\alpha}_{i \leftarrow k}(pa_i) + \bar{\alpha}_{i \leftarrow k}. \quad (\text{A.7})$$

For convenience, if we let

$$\bar{\rho}_{i \rightarrow k} = \max_{j: j \neq k} \rho_{i \rightarrow k}(X_j), \quad (\text{A.8})$$

then

$$\max_{j': j' \neq k} \tilde{\rho}_{i \rightarrow k}(X_{j'}) = 0, \quad (\text{A.9})$$

and

$$\max_{j'} \tilde{\rho}_{i \rightarrow k}(X_{j'}) = \max(0, \tilde{\rho}_{i \rightarrow k}(X_k)). \quad (\text{A.10})$$

By setting

$$\bar{\alpha}_{i \leftarrow k} = \alpha_{i \leftarrow k}(pa_i : pa_i \neq X_k), \quad (\text{A.11})$$

we have

$$\tilde{\alpha}_{i \leftarrow k}(pa_i) = 0 \text{ for all } pa_i \neq X_k. \quad (\text{A.12})$$

This also means that

$$\sum_{k':k' \neq k} \tilde{\alpha}_{i \leftarrow k'}(pa_i) = \tilde{\alpha}_{i \leftarrow pa_i}(pa_i) \text{ for } pa_i \neq X_k. \quad (\text{A.13})$$

and

$$\sum_{k':k' \neq k} \tilde{\alpha}_{i \leftarrow k'}(pa_i) = 0 \text{ for } pa_i = X_k. \quad (\text{A.14})$$

Using (A.6) and (A.7), we solve for

$$\tilde{\rho}_{i \rightarrow k}(pa_i = X_k) = \rho_{i \rightarrow k}(pa_i = X_k) - \bar{\rho}_{i \rightarrow k} \quad (\text{A.15})$$

and

$$\tilde{\alpha}_{i \leftarrow k}(pa_i = X_k) = \alpha_{i \leftarrow k}(pa_i = X_k) - \bar{\alpha}_{i \leftarrow k} \quad (\text{A.16})$$

to obtain simple update equations where $\bar{\rho}$ and $\bar{\alpha}$ terms cancel as follows:

$$\tilde{\rho}_{i \rightarrow k}(pa_i = X_k) = \rho_{i \rightarrow k}(pa_i = X_k) - \bar{\rho}_{i \rightarrow k}, \quad (\text{A.17})$$

$$= \rho_{i \rightarrow k}(X_k) - \max_{j:j \neq k} \rho_{i \rightarrow k}(X_j), \quad (\text{A.18})$$

$$= s(X_i, X_k) + \sum_{k':k' \neq k} \bar{\alpha}_{i \leftarrow k'} - \max_{j:j \neq k} [s(X_i, X_j) + \tilde{\alpha}_{i \leftarrow j}(X_j) + \sum_{k':k' \neq k} \bar{\alpha}_{i \leftarrow k'}], \quad (\text{A.19})$$

where (A.18) is obtained using (A.8). (A.19) results from (A.4), (A.13) and (A.14).

$$\tilde{\alpha}_{i \leftarrow k}(pa_i = X_k) = \alpha_{i \leftarrow k}(pa_i = X_k) - \bar{\alpha}_{i \leftarrow k} \quad (\text{A.20})$$

$$= \alpha_{i \leftarrow k}(X_k) - \alpha_{i \leftarrow k}(X_j : X_j \neq X_k) \quad (\text{A.21})$$

$$= \begin{cases} \sum_{i':i' \neq k} \max(0, \tilde{\rho}_{i' \rightarrow k}(X_k)) \\ + \sum_{i':i' \neq k} \bar{\rho}_{i' \rightarrow k} - \sum_{i':i' \neq k} \bar{\rho}_{i' \rightarrow k}, & \text{for } k = i \\ \tilde{\rho}_{k \rightarrow k}(X_k) + \sum_{i':i' \notin \{i,k\}} \max(0, \tilde{\rho}_{X_{i'} \rightarrow X_k}(X_{j'})) + \sum_{i':i' \neq i} \bar{\rho}_{X_{i'} \rightarrow X_k} \\ - \max[0, \tilde{\rho}_{k \rightarrow k}(X_k) \\ + \sum_{i':i' \notin \{i,k\}} \max(0, \tilde{\rho}_{X_{i'} \rightarrow X_k}(X_{j'})] - \sum_{i':i' \neq i} \bar{\rho}_{i' \rightarrow k}, & \text{for } k \neq i \end{cases} \quad (\text{A.22})$$

where (A.21) is obtained using (A.11). (A.22) results from (A.5), (A.6), (A.9) and (A.10).

The vector messages are of the form $\begin{pmatrix} 0 \\ \vdots \\ \rho_{i \rightarrow k}(X_k) \\ \vdots \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ \vdots \\ \alpha_{i \leftarrow k}(X_k) \\ \vdots \\ 0 \end{pmatrix}$ given that $\tilde{\rho}_{i \rightarrow k}(pa_i)$ and $\tilde{\alpha}_{i \leftarrow k}(pa_i)$ for $pa_i \neq X_k$ are not used in the updates. Hence messages can be considered to be scalar. This allows to define the responsibilities $r(X_i, X_k)$ and availabilities $a(X_i, X_k)$ as:

$$r(X_i, X_k) = \tilde{\rho}_{i \rightarrow k}(pa_i = X_k) = s(X_i, X_k) - \max_{j: j \neq k} [s(X_i, X_j) + a(X_i, X_j)], \quad (\text{A.23})$$

and

$$a(X_i, X_k) = \begin{cases} \tilde{\alpha}_{i \leftarrow k}(pa_i = X_k) = \sum_{i': i' \neq k} \max(0, r(X_{i'}, X_k)), & \text{for } k = i \\ \min[0, r(X_k, X_k) + \sum_{i': i' \notin \{i, k\}} \max(0, r(X_{i'}, X_k))], & \text{for } k \neq i \end{cases} \quad (\text{A.24})$$

where the $\min[0, \cdot]$ in the availability update comes from the fact that $x - \max(0, x) = \min(0, x)$.

B Dimensionless equations

B.1 Duffing resonator

The dimensional form of the driven Duffing equation is given by

$$m\ddot{x} + k_1\dot{x} + k_2x + \delta x^3 = F \cos(\omega t), \quad (\text{B.1})$$

where m is the mass of the resonator, k_1 is the damping term (ω_0 is the natural frequency of the resonator and Q is its quality factor), k_2 is the spring constant, δ is the cubic nonlinear spring constant. F and ω represent the amplitude and the frequency of the driving force, respectively.

By setting $x = zd$, where d is the maximum allowable displacement of the oscillators or characteristic dimension, and $t = \frac{\tau}{\omega_0}$ and using the dimensionless variables z and τ , Equation (B.1) then becomes

$$m\omega_0^2 d \frac{d^2 z}{d\tau^2} + m\omega_0^2 \frac{d}{Q} \frac{dz}{d\tau} + m\omega_0^2 dz + \delta d^3 z^3 = F \cos\left(\frac{\omega}{\omega_0} \tau\right). \quad (\text{B.2})$$

Using the relations $m\omega_0^2 = k_2$ and $k_2d = \mathcal{F}$, and setting $\Omega = \frac{\omega}{\omega_0}$, we have,

$$\mathcal{F} \frac{d^2z}{d\tau^2} + \frac{\mathcal{F}}{Q} \frac{dz}{d\tau} + \mathcal{F}z + \delta d^3 z^3 = F \cos(\Omega\tau). \quad (\text{B.3})$$

Dividing the above equation by $\mathcal{F} = k_2d$, we obtain

$$\frac{d^2z}{d\tau^2} + \frac{1}{Q} \frac{dz}{d\tau} + z + \frac{\delta}{k_2} d^2 z^3 = F_d \cos(\Omega\tau), \quad (\text{B.4})$$

where $F_d = \frac{F}{\mathcal{F}}$ is the normalised driving force. Hence the normalised form of the Duffing equation can be written as

$$\boxed{\ddot{x} + \gamma\dot{x} + x + \delta x^3 = F_d \cos(\Omega\tau)}, \quad (\text{B.5})$$

where $\gamma = \frac{1}{Q}$ and $\delta = \frac{\delta}{k_2} d^2$.

B.2 Van der Pol oscillator

The van der Pol equation is given by:

$$m\ddot{x} + \mu(x^2 - d^2)\dot{x} + kx = 0, \quad (\text{B.6})$$

where μ is the damping term and k is the spring constant. To match the fundamental frequency of Duffing resonator and Van der Pol oscillator, we assume that $\mu = k_1$ and $k = k_2$.

By using the scaled variables for the displacement and time as previously, we have

$$m\omega_0^2 d \frac{d^2z}{d\tau^2} + \mu d^3 \omega_0 (z^2 - 1) \frac{dz}{d\tau} + kdz = 0, \quad (\text{B.7})$$

$$\frac{d^2z}{d\tau^2} + \frac{\mu d^2}{\sqrt{mk}} (z^2 - 1) \frac{dz}{d\tau} + kdz = 0. \quad (\text{B.8})$$

The term $\frac{\mu}{\sqrt{mk}}$ is equivalent to $\frac{1}{Q}$ which implies

$$\frac{d^2z}{d\tau^2} + \frac{d^2}{Q} (z^2 - 1) \frac{dz}{d\tau} + z = 0. \quad (\text{B.9})$$

Therefore the normalised form of the Van der Pol equation is given by

$$\boxed{\ddot{x} + \gamma^*(x^2 - 1)\dot{x} + x = 0}, \quad (\text{B.10})$$

where $\gamma^* = \frac{\mu d^2}{\sqrt{mk}}$.

B.3 Coupled driven Duffing resonator and van der Pol oscillator

When the driven Duffing resonator and van der Pol oscillators are coupled together through displacement, the following two equations are obtained using the above re-scaled time and displacement variables.

$$\ddot{x}_{i,j} + \gamma \dot{x}_{i,j} + x_{i,j} + \delta x_{i,j}^3 = \kappa \left(\sum_{n=1}^N x_n - N x_{i,j} \right) + F_d \cos(\Omega\tau), \quad (\text{B.11})$$

$$\ddot{x}_{i,j} + \gamma^* (x_{i,j}^2 - 1) \dot{x}_{i,j} + x_{i,j} = \kappa \left(\sum_{n=1}^N x_n - N x_{i,j} \right), \quad (\text{B.12})$$

where κ is the coupling strength between the resonators and oscillators, $x_{i,j}$ represents the displacement of the resonator or oscillator located at the i -th row and j -th column in the square lattice, N is the number of nearest neighbours of the element at indices (i, j) , and x_n , $n = 1, \dots, N$, are the displacements of its nearest neighbours. A resonator coupled to an oscillator that is located in the middle of the array, for example, have four neighbours each and the equation of motion of the coupled system is given by:

$$\begin{aligned} \ddot{x}_{i,j} + \gamma \dot{x}_{i,j} + x_{i,j} + \delta x_{i,j}^3 &= \kappa (x_{i,j+1} + x_{i,j-1} + x_{i-1,j} + x_{i+1,j} - 4x_{i,j}) \\ &+ F_d \cos(\Omega\tau), \end{aligned} \quad (\text{B.13})$$

$$\ddot{x}_{i,j} + \gamma^* (x_{i,j}^2 - 1) \dot{x}_{i,j} + x_{i,j} = \kappa (x_{i,j+1} + x_{i,j-1} + x_{i-1,j} + x_{i+1,j} - 4x_{i,j}). \quad (\text{B.14})$$

Multiple time scales analysis of the driven Duffing resonator weakly coupled to a van der Pol oscillator

C

The method of multiple time scales (Nayfeh and Mook, 1995) is a global perturbation scheme, that is, a method that is used to find an approximate solution to a problem which cannot be solved exactly. It is useful in systems characterised by disparate time scales, such as weak dissipation in an oscillator. In Nayfeh and Mook (1995), the basic method has been described to characterise the behaviour of a single driven Duffing resonator and a single van der Pol oscillator. In this work, the derivations are related to a driven Duffing resonator coupled to a van der Pol oscillator, which is novel and has appeared in the following two journal papers: Wei et al. (2010); Randrianandrasana et al. (2010).

In order to introduce the method of multiple time scales, let us recall the dimensionless equation of motion of the coupled system when the van der Pol oscillator is coupled to

the driven Duffing resonator, as seen in Chapter 4:

$$\ddot{x} + \gamma_x \dot{x} + x + \delta x^3 = \kappa(y - x) + f_d \cos\left(\frac{\omega_d}{\omega_x} \tau\right), \quad (\text{C.1a})$$

$$\ddot{y} + \gamma_y(y^2 - 1)\dot{y} + \left(\frac{\omega_y}{\omega_x}\right)^2 y = \kappa(x - y), \quad (\text{C.1b})$$

where, x, γ_x, ω_x and x_2, γ_y, ω_y are the dimensionless displacement, damping coefficient and fundamental frequency of the Duffing resonator and the van der Pol oscillator respectively. f_d and ω_d are the amplitude and frequency of the external driving force. δ is the non-linearity of the Duffing resonator and κ is the coupling stiffness between the two coupled elements. We consider the case where both the internal and external resonances coincide ($\omega_x = \omega_y$). The method of multiple time scales is valid when the system parameters are small and are of the same order so that a solvable linear problem around a small perturbation is obtained. Equation (C.1) is therefore rewritten as:

$$\ddot{x} + \epsilon\mu_x \dot{x} + x + \epsilon\alpha x^3 = \epsilon\beta(y - x) + \epsilon F \cos(\Omega\tau), \quad (\text{C.2a})$$

$$\ddot{y} + \epsilon\mu_y(y^2 - 1)\dot{y} + y = \epsilon\beta(x - y), \quad (\text{C.2b})$$

where the coefficients $\mu_x, \mu_y, \alpha, \beta$ and F are such that $\gamma_x = \epsilon\mu_x, \gamma_y = \epsilon\mu_y, \delta = \epsilon\alpha, \kappa = \epsilon\beta$ and $f_d = \epsilon F$. $\epsilon \ll 1$ is a positive parameter and $\Omega = \frac{\omega_d}{\omega_x}$ is the ratio of the frequency of the excitation to the linear natural frequency of the system. Equation (C.2) could not be solved using a standard perturbation method in which a solution expressed as an expansion in a power series is assumed to exist. This is due to the presence of *small divisors* and *secular terms* in the expansion. On the one hand, the small divisor terms occur for some values of the frequency Ω which make some of the denominators in the expansion to be very small. Normally the expansion is composed of two terms and when $\Omega \approx 1$, small divisors first appear in the first term hence the name *primary* or *main resonance* when $\Omega \approx 1$. For other values of Ω which cause the small divisors to first appear in the second term, these cases are referred to as *secondary resonances*. On the other hand, the name “secular terms” comes from the astronomy literature. Etymologically, the word secular is derived from the French word siècle, which means century. For the expansion to be uniform, the corrections must be free of secular terms as will be explained later.

As the name of the method suggests, in the multiple scales analysis, the scales $\tau_0 = \tau$ and $\tau_1 = \epsilon\tau$ are introduced to determine an approximate solution to Equation (C.2). τ_0 is called the fast time and τ_1 the slow time since it only becomes significant when $t \sim 1/\epsilon$.

Hence the derivatives are transformed according to

$$\begin{aligned}\frac{d}{d\tau} &= D_0 + \epsilon D_1 + \dots, \\ \frac{d^2}{d\tau^2} &= D_0^2 + 2\epsilon D_0 D_1 + \dots,\end{aligned}$$

where $D_n = \partial/\partial\tau_n$.

Since we consider the coupled system as one unit, we set the natural frequencies (ω_0) of the Duffing and the van der Pol to be the same.

Since τ appears explicitly in the governing equation, we check whether the dependence on τ is fast or slow. In this case, we check $\cos(\Omega\tau)$. At the primary resonance, $\Omega \approx 1$. Since Ω is away from zero, $\cos(\Omega\tau)$ is fast varying, and we write

$$\cos(\Omega\tau) = \cos(\Omega\tau_0),$$

that is, τ is represented in terms of τ_0 .

Then Equation (C.2) becomes

$$D_0^2 x + x = \epsilon [\beta(y - x) + F \cos(\Omega\tau_0) - 2D_0 D_1 x - \mu_x D_0 x - \alpha x^3], \quad (\text{C.3a})$$

$$D_0^2 y + y = \epsilon [\beta(x - y) - 2D_0 D_1 y - \mu_y (y^2 - 1) D_0 y]. \quad (\text{C.3b})$$

Clearly, the left-hand side of Equation (C.3) is a simple harmonic oscillator whose solutions can be approximated in the form

$$x = x_0(\tau_0, \tau_1) + \epsilon x_1(\tau_0, \tau_1) + \dots, \quad (\text{C.4a})$$

$$y = y_0(\tau_0, \tau_1) + \epsilon y_1(\tau_0, \tau_1) + \dots \quad (\text{C.4b})$$

Substituting (C.4a) and (C.4b) into (C.3a) and (C.3b) respectively and equating coefficients of like powers of ϵ , we obtain

$$D_0^2 x_0 + x_0 = 0, \quad (\text{C.5a})$$

$$D_0^2 y_0 + y_0 = 0, \quad (\text{C.5b})$$

$$D_0^2 x_1 + x_1 = -2D_0 D_1 x_0 - \mu_x D_0 x_0 - \alpha x_0^3 + \beta(y_0 - x_0) + F \cos(\Omega\tau_0), \quad (\text{C.5c})$$

$$D_0^2 y_1 + y_1 = -2D_0 D_1 y_0 - \mu_y (y_0^2 - 1) D_0 y_0 + \beta(x_0 - y_0). \quad (\text{C.5d})$$

The general solution of (C.5a) and (C.5b) can be expressed as

$$x_0(\tau_0, \epsilon) = A_x(\tau_1)e^{i\tau_0} + cc, \quad (\text{C.6a})$$

$$y_0(\tau_0, \epsilon) = A_y(\tau_1)e^{i\tau_0} + cc, \quad (\text{C.6b})$$

where cc denotes the complex conjugate. The quantities $A_x(\tau_1)$ and $A_y(\tau_1)$ are arbitrary, complex functions which are determined from Equations (C.5c) and (C.5d) by imposing solvability or secular conditions.

Then (C.5c) and (C.5d) become

$$\begin{aligned} D_0^2 x_1 + x_1 &= -2\left(ie^{i\tau_0} \frac{dA_x}{d\tau_1} + cc\right) - \mu_x i A_x e^{i\tau_0} - \alpha(A_x^3 e^{i3\tau_0} \\ &\quad + 3|A_x|^2 A_x e^{i\tau_0} + cc) + \beta((A_y - A_x)e^{i\tau_0} + cc) \\ &\quad + \frac{F}{2}(e^{i(1+\epsilon\sigma)\tau_0} + cc), \quad (\text{C.7a}) \\ D_0^2 y_1 + y_1 &= -2\left(ie^{i\tau_0} \frac{dA_y}{d\tau_1} + cc\right) + \beta(A_x - A_y)e^{i\tau_0} + cc \\ &\quad - \mu_y (A_y^2 e^{2i\tau_0} + 2|A_y|^2 + \overline{A_y}^2 e^{-2i\tau_0} - 1)(A_y i e^{i\tau_0} + cc), \end{aligned}$$

where a detuning parameter σ has been introduced since the case $\Omega \approx 1$ is considered. σ is defined by

$$\Omega = 1 + \epsilon\sigma. \quad (\text{C.8})$$

Then

$$\Omega\tau_0 = (1 + \epsilon\sigma)\tau_0 = \tau_0 + \epsilon\sigma\tau_0 = \tau_0 + \sigma\tau_1. \quad (\text{C.9})$$

In Equation (C.7a), for example, there are terms whose common factor is $e^{i\tau_0}$ and other terms whose common factor is $e^{i3\tau_0}$, where τ_0 is a fast time scale. In that case, the secular terms correspond to the factors of $e^{i\tau_0}$ since $e^{i\tau_0}$ becomes appreciable compared to $e^{i3\tau_0}$ after very long times, given that τ_0 is fast varying. Thus, for the expansion to be uniform, the corrections must be free of secular terms.

Eliminating the secular terms from (C.7), yields

$$2\frac{dA_x}{d\tau_1} + \mu_x A_x - 3i\alpha|A_x|^2 A_x + \beta i(A_y - A_x) + \frac{i}{2}F e^{i\sigma\tau_1} = 0, \quad (\text{C.10a})$$

$$\frac{dA_y}{d\tau_1} + i\frac{\beta}{2}(A_x - A_y) + \frac{\mu_y}{2}(|A_y|^2 - 1)A_y = 0. \quad (\text{C.10b})$$

Expressing $A_x(\tau_1, \epsilon)$ and $A_y(\tau_1, \epsilon)$ in the polar forms $A_x(\tau_1, \epsilon) = a_x(\tau_1, \epsilon)e^{i\theta_x(\tau_1, \epsilon)}$ and $A_y(\tau_1, \epsilon) = a_y(\tau_1, \epsilon)e^{i\theta_y(\tau_1, \epsilon)}$, where $a_x(\tau_1, \epsilon)$ and $\theta_x(\tau_1, \epsilon)$, respectively, $a_y(\tau_1, \epsilon)$ and $\theta_y(\tau_1, \epsilon)$ are the amplitudes and the phases of the fundamental solutions. Separating real and imag-

inary parts in (C.10) yields

$$2\frac{da_x}{d\tau_1} = -\mu_x a_x + \beta a_y \sin \Gamma_y - \frac{F}{2} \sin \Gamma_x, \quad (\text{C.11a})$$

$$2\frac{da_y}{d\tau_1} = -\mu_y (a_y^2 - 1)a_y - \beta a_1 \sin \Gamma_y, \quad (\text{C.11b})$$

$$2a_x \frac{d\Gamma_x}{d\tau_1} = -(\beta - 2\sigma)a_x + 3\alpha a_x^3 - \beta a_y \cos \Gamma_y - \frac{F}{2} \cos \Gamma_x, \quad (\text{C.11c})$$

$$2a_y \frac{d\Gamma_y}{d\tau_1} = (\beta - 2\sigma)a_y - \beta a_x \cos \Gamma_y - 2a_y \frac{d\Gamma_x}{d\tau_1}, \quad (\text{C.11d})$$

where the following transformations have been introduced:

$$\Gamma_x = \theta_x - \sigma\tau_1, \quad (\text{C.12})$$

and

$$\Gamma_y = \theta_y - \theta_x. \quad (\text{C.13})$$

Hence

$$\frac{d\Gamma_x}{d\tau_1} = \frac{d\theta_x}{d\tau_1} - \sigma, \quad (\text{C.14})$$

and

$$\frac{d\Gamma_y}{d\tau_1} = \frac{d\theta_y}{d\tau_1} - \frac{d\theta_x}{d\tau_1}. \quad (\text{C.15})$$

The previous derivatives in Equations (C.7a) to (C.11d) are computed with respect to τ_1 . To determine the steady-state motions, we use the fact that a_x , a_y and Γ_x , Γ_y are constants, set $\frac{da_x}{d\tau_1} = \frac{da_y}{d\tau_1} = 0$ and $\frac{d\Gamma_x}{d\tau_1} = \frac{d\Gamma_y}{d\tau_1} = 0$. From (C.11), we find, for the Duffing resonator,

$$\left(a_x(\beta - 2\sigma) + 3\alpha a_x^3 - \frac{a_y^2(\beta - 2\sigma)}{a_x} \right)^2 + \left(a_x \mu_x - \frac{a_y^2 \mu_y}{a_x} (1 - a_y^2) \right)^2 = \frac{1}{4} F^2. \quad (\text{C.16})$$

Similarly, for the Van der Pol oscillator, we have

$$(\beta - 2\sigma)^2 a_y^2 + \mu_y^2 a_y^2 (1 - a_y^2)^2 = a_x^2 \beta^2. \quad (\text{C.17})$$

Equations (C.16) and (C.17) correspond to the steady-state response of the amplitudes of the Duffing resonator and van der Pol oscillator respectively, from which the frequency-response curves in Figure 4.1 are constructed.

Bibliography

- Juan A. Acebron, L. L. Bonilla, Conrad J. Pérez Vicente, Félix Ritort, and Renato Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Rev. Mod. Phys.*, 77(1):137–185, Apr 2005. doi: 10.1103/RevModPhys.77.137.
- S. Acid and L. M. de Campos. Learning right sized belief networks by means of a hybrid methodology. *Lecture Notes in Artificial Intelligence*, 1910:309–315, 2000.
- S. Acid and L. M. de Campos. A hybrid methodology for learning belief networks: Benedict. *International Journal of Approximate Reasoning*, 27:235–262, 2001.
- V. S. Afraimovich and W. W. Lin. Synchronization in lattices of coupled oscillators with Neumann/periodic boundary conditions. *Dynamics and Stability Systems*, 13:237–264, 1998.
- V. S. Afraimovich and H. M. Rodrigues. Uniform ultimate boundedness and synchronization for nonautonomous equations. Technical report, Georgia Institute of Technology, 1994. Preprint CDSNS 94-202.
- V. S. Afraimovich, N. N. Verichev, and M. I. Rabinovich. Stochastic synchronization of oscillations in dissipative systems. *Radiophysics and Quantum Electronics*, 29(9): 795–803, 1986.
- V. S. Afraimovich, S. N. Chow, and J. K. Hale. Synchronisation in lattices of coupled oscillators. *Physica D*, 103:445–1490, 1997.
- H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723, 1974.
- F. Ambrogi, E. Biganzoli, P. Querzoli, S. Ferretti, P. Boracchi, S. Alberti, E. Marubini, and I. Nenci. Molecular subtyping of breast cancer from traditional tumor marker profiles using parallel clustering methods. *Clinical Cancer Research*, 12:781–790, February 2006.

- S. A. Anderson, D. Madigan, and M. D. Perlman. A characterisation of Markov equivalence classes for acyclic digraphs. Technical report, Dept of Statistics, University of Washington, 1995. #287.
- A. Arenas, A. Daz-Guilera, J. Kurths, Y. Moreno, and C. Zhou. Synchronization in complex networks. *Physics Reports*, 469:93–153, 2008.
- D. G. Aronson, G. B. Ermentrout, and N. Koppel. Amplitude response of coupled oscillators. *Physica D*, 41(403):403–449, 1990.
- K. Aubin, M. K. Zalalutdinov, T. Alan, R. B. Reichenbach, R. Rand, A. Zehnder J. M. Parpia, and H. Craighead. Limit cycle oscillations in CW laser-driven NEMS. *Journal of Microelectromechanical Systems*, 13:1018–1026, 2004.
- K. Bar-Eli. On the stability of coupled chemical oscillators. *Physica D*, 14(242):242–252, 1985.
- A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. In *Proceedings of the National Academy of Science USA*, volume 101, pages 3747–3752, 2004.
- M. A. Barron and M. Sen. Synchronisation of four coupled van der Pol oscillators. *Non-linear Dynamics*, 56:357–367, 2009.
- I. A. Beinlich, H. J. Suermondt, R. M. Chavez, and G. F. Cooper. The alarm monitoring system: a case study with two probabilistic inference techniques for belief networks. In *In Proceedings of the European Conference on Artificial Intelligence in Medicine*, pages 247–256, 1989.
- I. Belykh, v. Belykh, K. Nevidin, and M. Hasler. Persistent clusters in lattices of coupled nonidentical chaotic systems. *Chaos*, 13(1):165–178, 2003a.
- V. N. Belykh and E. Mosekilde. Cluster synchronization modes in an ensemble of coupled chaotic oscillators. *Physical Review E*, 63:036216, 2001.
- V. N. Belykh and E. Mosekilde. One-dimensional map lattices: synchronization, bifurcations, and chaotic structures. *Physical Review E*, 54:3196–3203, 1996.
- V. N. Belykh, I. V. Belykh, and M. Hasler. Hierarchy and stability of partially synchronous oscillations of diffusively coupled dynamical systems. *Physical Review E*, 62(5):6332–6345, 2000.

- V. N. Belykh, I. V. Belykh, M. Hasler, and K. V. Nevidin. Cluster synchronization modes in three-dimensional lattices of diffusively coupled oscillators. *International Journal of Bifurcation and Chaos*, 13(4):755–779, 2003b.
- J. H. Bommel and M. A. Musen. *Handbook of Medical Informatics*. Springer Verlag, Heidelberg, Germany, 1997.
- D. Di Bernardo, M. G. Signorini, and S. Cerutti. A model of two nonlinear coupled oscillators for the study of heartbeat dynamics. *International Journal of Bifurcation and Chaos*, 8(10):1975–1985, 1998.
- C. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.
- B. Boashash. Estimating and interpreting the instantaneous frequency of a signal – Part 1: Fundamentals. *IEEE Transactions on Signal Processing*, 80(4):520–538, 1992a.
- B. Boashash. Estimating and interpreting the instantaneous frequency of a signal – Part 2: Algorithms and Applications. *IEEE Transactions on Signal Processing*, 80(4):540–568, 1992b.
- R. R. Bouckaert. Belief networks construction using the minimum description length principle. *Lecture Notes in Computer Science*, 747:41–48, 1993.
- R. R. Bouckaert. *Bayesian Belief Networks: from Construction to Inference*. PhD thesis, University of Utrecht, 1995.
- L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley. Optimal paths in disordered complex networks. *Physical Review Letters*, 91:168701, 2001.
- Jorge N. Breaa, Leslie M. Kayb, and Nancy J. Kopella. Biophysical model for gamma rhythms in the olfactory bulb via subthreshold oscillations. *PNAS*, 106(51):21954–21959, December 2009.
- M. Buchanan. *The social atom*. Bloomsbury, 2007.
- Eyal Buks and Michael L. Roukes. Electrically tunable collective response in a coupled micromechanical array. *Journal of MicroElectroMechanical Systems*, 11(6):802–807, 2002.
- W. Buntine. Theory refinement of Bayesian networks. In *Proceedings of the Seventh Conference on Uncertainty in Artificial Intelligence*, pages 52–60, 1991.

- Thomas Burwick. Temporal coding: competition for coherence and new perspectives on assembly formation. In *IJCNN'09: Proceedings of the 2009 international joint conference on Neural Networks*, pages 3467–3476, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-3549-4.
- J. Cheng, R. Greiner, J. Kelly, D. A. Bell, and W. Liu. Learning Bayesian networks from data: an information-theory based approach. *Artificial Intelligence*, 137:43–90, 2002.
- J. Chevrolat, J. Golmard, S. Ammar, R. Jouvent, and J. Boisvieux. Modeling behaviour syndromes using Bayesian networks. *Artificial Intelligence in Medicine*, 14(3):259–277, 1998.
- D. R. Chialvo and J. Jalife. Non-linear dynamics of cardiac excitation and impulse propagation. *Nature*, 330:749–752, 1987.
- D. R. Chialvo and J. Jalife. Changing opinions in a changing world: a new perspective in sociophysics. *International Journal of Modern Physics C*, 16:515, 2005.
- D. M. Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, 2002a.
- D. M. Chickering. Learning equivalence classes of Bayesian network structures. *J. Mach. Learn. Res.*, 2:445–498, 2002b.
- C-H. Chiu, W-W. Lin, and C-S. Wang. Synchronization in a lattice of coupled van der Pol systems. *International Journal of Bifurcation and Chaos*, 8(12):2353–2373, 1998.
- C-H. Chiu, W-W. Lin, and C-S. Wang. Synchronization in a lattice of coupled oscillators with various boundary conditions. *Nonlinear Analysis*, 46:213–229, 2001.
- C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14:462–467, 1968.
- S. Chow and J. Mallet-Paret. Pattern formation and spatial chaos in lattice dynamical systems – Part I. *IEEE Transactions on Circuits. 1, Fundamental theory and applications*, 42:746–751, 1995.
- P. Colet, R. Roy, and K. Weisenfeld. Controlling hyperchaos in a multimode laser model. *Physical Review E*, 50:3453–3457, 1994.
- G. F. Cooper and E. Herskovits. Kutató: An entropy-driven system for construction of probabilistic expert systems from databases. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pages 54–62, 1990.

- G. F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–348, 1992.
- G.F. Cooper and E.H. Herskovits. A Bayesian method for the induction of probabilistic networks from data. Technical report, Section of Medical Informatics, Univ. of Pittsburgh, 1991.
- D. Dash and M. Druzdzal. A hybrid anytime algorithm for the construction of causal models from sparse data. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 142–149. Morgan Kaufmann, 1999.
- L. M. de Campos. Independency relationships and learning algorithms for singly connected networks. *Journal of Experimental and Theoretical Artificial Intelligence*, 10:511–549, 1998.
- L. M. de Campos and J. F. Huete. A new approach for learning belief networks using independence criteria. *International Journal of Approximate Reasoning*, 24:11–37, 2000.
- L. M. de Campos, J. M. Fernández-Luna, and J. M. Puerta. An iterated local search algorithm for learning Bayesian networks with restarts based on conditional independence tests. *International Journal of Intelligent Systems*, 18:221–235, 2003.
- S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Review of Modern Physics*, 80:1275–1335, 2008.
- G. Duffing. *Erzwungene Schwingungen bei Vernderlicher Eigenfrequenz*. F. Vieweg u. Sohn, 1918.
- DXPLAIN. Massachusetts General Hospital, Boston, 1992.
- M. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.
- G. B. Ermentrout. Loosing amplitude and saving phase. In H. G. Othmer, editor, *Non-linear Oscillations in Biology and Chemistry*. Springer-Verlag, Berlin, 1986.
- G. B. Ermentrout. Oscillator death in populations of all to all coupled nonlinear oscillators. *Physica D*, 41(219):219–231, 1990.
- G. B. Ermentrout and W. C. Troy. Phase locking in a reaction-diffusion system with a linear frequency gradient. *SIAM Journal on Applied Mathematics*, 39(623):359–367, 1986.

- R. Feynman. There's plenty of room at the bottom, 1959. <http://www.zyvex.com/nanotech/feynman.html>.
- B. J. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 215:972–976, 2007.
- B.J. Frey, F.R. Kschischang, H.A Loeliger, and N. Wiberg. Factor graphs and algorithms. In *Proceedings of the 35th Allerton Conference on Communication, Control and Computing*, 1997.
- N. Friedman and M. Goldszmidt. Learning Bayesian networks with local structure. In Michael I. Jordan, editor, *Learning in Graphical Models*, pages 421–459. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
- N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyse expression data. In *Proceedings of the fourth annual international conference on Computational molecular biology*, pages 127–135. ACM, 2000.
- H. Fujisaka and T. Yamada. Stability theory of synchronized motion in coupled oscillator systems. *Progress of Theoretical Physics*, 69:32–47, 1983.
- S. F. Galán, F. Aguado, F. J. Díez, and Mira J. NasoNet. Modelling the spread of nasopharyngeal cancer with temporal Bayesian networks. *Artificial Intelligence in Medicine*, 25: 247–254, 2002.
- J. García-Ojalvo, M. Elowitz, and S. Strogatz. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. In *Proceedings of the National Academy of Science USA*, volume 101, pages 10955–10960, 2004.
- A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, second edition, 2004.
- S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–742, 1983.
- R. Gentleman, B. Ding, S. Dudoit, and J. Ibrahim. Distance measures in DNA microarray data analysis. In R. C. Gentleman, V. J. Carey, W. Huber, R. Irizarry, and S. Dudoit, editors, *Bioinformatics and computational biology solutions using R and Bioconductor*, chapter 12, pages 189–208. Springer-Verlag, 2005.

- J. Gómez-Gardenes and Y. Moreno. Synchronization in networks with variable local properties. *International Journal of Bifurcation and Chaos*, 17:2501–2507, 2007.
- J. Gómez-Gardenes, Y. Moreno, and A. Arenas. Paths to synchronization on complex networks. *Physical Review Letters*, 98:034101, 2007a.
- J. Gómez-Gardenes, Y. Moreno, and A. Arenas. Synchronisability determined by coupling strengths and topology on complex networks. *Physical Review E*, 75:066106, 2007b.
- J. C. Goswami and A. E. Hoefel. Algorithms for estimating instantaneous frequency. *Signal processing*, 84(8):1423–1427, 2004.
- A. Greilich, M. Markus, and E. Goles. Control of spatiotemporal chaos: dependence of the minimum pinning distance on the spatial measure entropy. *European Physical Journal D*, 33(2):279–283, 2005.
- R. O. Grigoriev, M. C. Cross, and H. G. Schuster. Pinning control of spatiotemporal chaos. *Physical Review Letters*, 79(15):2795–2798, 1997.
- M. H. De Groot. *Optimal Statistical Decisions*. McGraw-Hill, 1970.
- J. K. Hale. Diffusive coupling, dissipative and synchronization. *Journal of Dynamics and Differential Equations*, 9(1):1–52, 1997.
- M. Hasler, Y. Maistrenko, and O. Popovich. Simple example of partial synchronization of chaotic systems. *Physical Review E*, 58(5):6843–6846, 1990.
- D. Heckerman, E. Horvitz, and B. Nathwani. Toward normative expert systems: Part I The Pathfinder Project. *Methods of Information in Medicine*, 31:90–105, 1992.
- D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. In *KDD Workshop*, pages 85–96, 1994.
- D. Heckerman, D. Geiger, and D. M. Chickering. Learning Bayesian networks: the combination of knowledge and statistical data. *Machine Learning*, 20:197–243, 1995.
- D. Heckerman, C. Meek, and G. Cooper. A Bayesian approach to causal discovery. In C. Glymour and G. F. Cooper, editors, *Computation, Causation, and Discovery*, pages 141–165. AAAI Press, Menlo Park, California, 1999.
- D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. M. Kadie. Dependency networks for inference, collaborative filtering, and data visualization. *Journal of Machine Learning Research*, 1:49–75, 2000.

- M. Henrion. Propagating uncertainty in Bayesian networks by probabilistic logic sampling. In *Uncertainty in Artificial Intelligence 2*, pages 149–163. Elsevier Science, 1988.
- E. H. Herskovits and A. P. Dagher. Applications of Bayesian networks to health care. Technical report, Noetic Systems Incorporated, Baltimore, Maryland, 1997. NSI-TR-1997-02.
- L. Huang, Y.-C. Lai, K. Park, and X.-G Wang. Synchronization in complex clustered networks. *Front. Phys. China*, 2(4):1–14, 2007.
- L. Huang, Q. Chen, Y. C. Lai, and L. M. Pecora. Generic behaviour of master-stability functions in coupled nonlinear dynamical systems. *Physical Review E*, 80:036204, 2009.
- ILIAD. Applied Informatics, Salt Lake City, 1992.
- F. Jensen, S. Lauritzen, and K. Olesen. Bayesian updating in recursive graphical models by local computations. *Computational Statistics Quarterly*, 4:269–282, 1990.
- H. Joe. Relative entropy measures of multivariate association. *Journal of the American Statistical Association*, 84:157–164, 1989.
- K. Kaneko. Clustering, coding, switching, hierarchical ordering, and control in a network of chaotic elements. *Physica D*, 41:137–172, 1990.
- T. Kapitaniak. Transition to hyperchaos in chaotically forced coupled oscillators. *Physical Review E*, 47:R2975–R2978, 1993.
- H. J. Kappen and J. P. Neijt. Promedas, a probabilistic decision support system for medical diagnosis. Technical report, Foundation for Neural Networks (SNN) - University Medical Centre Utrecht (UMCU), The Netherlands, 2002.
- R. A. Katz, T. Galib, and J. Cembrola. Mechanisms underlying transitional and turbulent boundary layer (tbl) flow-induced noise in underwater acoustics (ii). *Journal de Physique IV*, 4(5):1063–1065, 1994.
- M. Kayaalp and G. F. Cooper. A Bayesian network scoring metric that is based on globally uniform parameter priors. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, pages 251–258, 2002.
- R. Korf. Linear-space best-first search. *Artificial Intelligence*, 62(1):41–78, 1993.
- A. Koseska, E. Volkov, A. Zaikin, and J. Kurths. Inherent multistability in arrays of autoinducer coupled genetic oscillators. *Physical Review E*, 75:031916, 2007.

- F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.
- A. R. Laird, B. P. Rogert, and M. E. Meyerand. Cluster synchronisation in functional magnetic resonance imaging. In *Proc. Intl. Soc. Mag. Reson. Med*, volume 10, page 1, 2002.
- W. Lam and F. Bacchus. Learning Bayesian belief networks. An approach based on the MDL principle. *Computational Intelligence*, 10:269–293, 1994.
- S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of Royal Statistics Society B*, 50(2):157–194, 1988.
- A. Lemaître and H. Chaté. Phase ordering and onset of collective behaviour in chaotic coupled map lattices. *Physical Review Letters*, 82:1140–1143, 1999.
- M. Leone, Sumedha, and M. Weigt. Clustering by soft-constraint affinity propagation: applications to gene-expression data. *Bioinformatics*, 23(20):2708–2715, 2007.
- X. Li, X. F. Wang, and G. R. Chen. Pinning a complex dynamical network to its equilibrium. *IEEE Transactions on Circuits and Systems–I: Regular papers*, 51(10):2074–2087, 2004.
- Y. Lin and M. J. Druzdzel. Stochastic sampling and search in belief updating algorithms for very large Bayesian networks. In *Working notes of the AAAI spring symposium on search techniques for problem solving under uncertainty and incomplete information*, pages 77–82, 1989.
- W. Lohmiller and J.-J. E. Slotine. On contraction analysis for nonlinear systems. *Automatica*, 34(6):683–696, 1998.
- S. Mani, S. McDermott, and M. Vltort. MENTOR: A Bayesian model for prediction of mental retardation in newborns. *Research in Developmental Disabilities*, 18(5):303–318, 1997.
- J. Marro and R. Dickman. *Nonequilibrium Phase Transitions in Lattice Models*. Cambridge University Press, 1999.
- P. McCullagh and J. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall, New York, 1989.

- P. N. McGraw and M. Menzinger. Clustering and the synchronization of oscillator networks. *Physical Review E*, 72:015101, 2005.
- P. N. McGraw and M. Menzinger. Analysis of nonlinear synchronization dynamics of oscillator networks by Laplacian spectral methods. *Physical Review E*, 75:027104, 2007.
- Meditel. Devon, Pa. Meditel: Computer assisted diagnosis, 1991.
- C. Meek. Causal inference and causal explanation with background knowledge. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 403–410, 1995.
- C. Meek. *Graphical models: selecting causal and statistical models*. PhD thesis, Carnegie Mellon University, 1997.
- L. Mendelowitz, A. Verdugo, and R. Rand. Dynamics of three coupled limit cycle oscillators with application to artificial intelligence. *Communications in Nonlinear Science and Numerical Simulation*, 14:270–283, 2009.
- N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculation by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
- A. E. Motter, C. Zhou, and J. Kurths. Enhancing complex-network synchronization. *Europhysical Letters*, 69:334–340, 2005a.
- A. E. Motter, C. S. Zhou, and J. Kurths. Network synchronization, diffusion, and the paradox of heterogeneity. *Physical Review E*, 71:016116, 2005b.
- K. Murphy, Y. Weiss, and M. Jordan. Loopy belief propagation for approximate inference: an empirical study. In *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence (UAI-99)*, pages 467–47, San Francisco, CA, 1999. Morgan Kaufmann.
- A. H. Nayfeh and D. T. Mook. *Nonlinear Oscillations*. Wiley Classics Library, New York, 1995.
- R. M. Neal. Probabilistic inference using Markov Chain Monte Carlo methods. Technical report, Dept. of Computer Science, University of Toronto, 1993.
- R. E. Neapolitan. *Learning Bayesian Networks*. Pearson Prentice Hall, 2004.

- V. I. Nekorkin, V. A. Makarov, V. B. Zaantsev, and M. G. Velarde. Spatial disorder and pattern formation in lattices of coupled bistable elements. *Physica D*, 100:330–342, 1997.
- M. E. J. Newman. Scientific collaboration networks. II. shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132, 2001.
- A. E. Nicholson. Fall diagnosis using dynamic belief networks. In *Proceedings of the 4th Pacific Rim International Conference on Artificial Intelligence (PRICAI-96)*, pages 218–227, 1996.
- O. Ogunyemi, J. Clarke, N. Ash, and B. Webber. Combining geometric and probabilistic reasoning for computer-based penetrating-trauma assessment. *Journal of the American Medical Informatics Association*, 9(3):273–282, 2002.
- A. Onisko. Evaluation of the Hepar II system for diagnosis of liver disorders. In *Working Notes on the European Conference on Artificial Intelligence in Medicine (AIME-01: Workshop Bayesian Models in Medicine)*, pages 59–64, 2001.
- V.I. Oseledets. A multiplicative ergodic theorem. lyapunov characteristic numbers for dynamical systems. *Transactions of the Moscow Mathematical Society*, 19:197–231, 1968.
- G. V. Osipov and J. Kurths. Regular and chaotic phase synchronisation of coupled circle maps. *Physical Review E*, 65(016216):1–13, 2001.
- G. V. Osipov and M. M. Sushchik. The effect of natural frequency distribution on cluster synchronisation in oscillator arrays. *IEEE Transactions on Circuits and Systems – I: fundamental theory and applications*, 44(10):1006–1010, 1997.
- G. V. Osipov and M. M. Sushchik. Synchronised clusters and multistability in arrays of oscillators with different natural frequencies. *Physical Review E*, 58(6):7198–7207, 1998.
- J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, San Francisco, CA, 1988.
- J. Pearl. Why there is no statistical test for confounding, why many think there is, and why they are almost right. Technical report, UCLA Computer Science Department, 1998. R-256.

- J. Pearl and T. S. Verma. A theory of inferred causation. In James F. Allen, Richard Fikes, and Erik Sandewall, editors, *KR'91: Principles of Knowledge Representation and Reasoning*, pages 441–452, San Mateo, California, 1991a. Morgan Kaufmann.
- J. Pearl and T. S. Verma. Equivalence and synthesis of causal models. In James F. Allen, Richard Fikes, and Erik Sandewall, editors, *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence*, pages 220–227, San Mateo, California, 1991b. M. Henrion, R. Shachter, L. Kanal and J. Lemmer.
- J. Pearl, D. Geiger, and T. S. Verma. The logic of Influence Diagrams. In R. M. Oliver and J. Q. Smith, editors, *Influence Diagrams, Belief Networks and Decision Analysis*, pages 480–487. Wiley Ltd., Sussex, England, 1990.
- L. M. Pecora. Synchronisation of oscillators in complex networks. *Pramana Journal of Physics*, 70(6):1175–1198, 2008.
- L. M. Pecora and T. L. Carroll. Master stability functions for synchronized coupled systems. *Physical Review Letters*, 80:2109–2112, 1998.
- A. Pentek, J. B. Kadtko, and Z. Toroczkai. Stabilizing chaotic vortex trajectories: an example of high-dimensional control. *Physics Letters A*, 224:85–92, 1996.
- V. Petrov, M. J. Crowley, and K. Showalter. Tracking unstable periodic orbits in the Belousov-Zhabotinsky reaction. *Physical Review Letters*, 72:2955–2958, 1994.
- V. Petrov, E. Mihaliuk, S. K. Scott, and K. Showalter. Stabilizing and characterizing unstable states in high-dimensional systems from time series. *Physical Review E*, 51:3988–3996, 1995.
- J. Platt. Fast training of support vector machines using sequential minimal optimization. In *Advances in Kernel Methods – Support Vector Learning*, pages 185–208. MIT Press, 1999.
- QMR. CAMDAT, Pittsburgh. QMR (Quick Medical Reference), 1992.
- M. I. Rabinovich, J. J. Torres, P. Varona, R. Huerta, and P. Weidman. Origin of coherent structures in a discrete chaotic medium. *Physical Review E*, 60(2):R1130–R1133, 1999.
- R. Rand and P. J. Holmes. Bifurcation of periodic motions in two weakly coupled van der Pol oscillators. *International Journal of Non-Linear Mechanics*, 15:387–399, 1980.

- R. Rand and J. Wong. Dynamics of four coupled phase-only oscillators. *Communications in Nonlinear Science and Numerical Simulation*, 13(3):501–507, 2008a.
- R. Rand and J. Wong. Dynamics of four coupled phase-only oscillators. *Communications in Nonlinear Science and Numerical Simulation*, 13:501–507, 2008b.
- M. F. Randrianandrasana, X. Wei, and D. Lowe. A preliminary study into emergent behaviours in a lattice of interacting nonlinear resonators and oscillators. *Communications in Nonlinear Science and Numerical Simulations*, 2010. In press.
- J. G. Restrepo, E. Ott, and B. R. Hunt. Synchronization in large directed networks of coupled phase oscillators. *Chaos*, 16:015107, 2005.
- J. G. Restrepo, E. Ott, and B. R. Hunt. Emergence of synchronization in large complex networks of interacting dynamical systems. *Physica D*, 224:114–122, 2006.
- J. Rissanen. Stochastic complexity and modeling. *Annals of Statistics*, 14:1080–1100, 1986.
- H. M. Rodrigues. Uniform ultimate boundedness and synchronization. Technical report, Georgia Institute of Technology, 1994. Preprint CDSNS 94-160.
- F. J. Romeiras, C. Grebogi, E. Ott, and W. P. Dayawansa. Controlling chaotic dynamical systems. *Physica D*, 58:165–192, 1992.
- M. G. Rosenblum, A. S. Pikovsky, and J. Kurths. Phase synchronization of chaotic oscillators. *Physical Review Letters*, 76(11):1804–1807, 1996.
- Jr. S. L. Marple. Computing the discrete-time analytic signal via FFT. *IEEE Transactions on Signal Processing*, 47(9):2601–2603, 1999.
- G. C. Sakellropoulos and G. C. Nikiforidis. Development of a Bayesian network in the prognosis of head injuries using graphical model selection techniques. *Methods of Information in Medicine*, 38:37–42, 1999.
- M. Sato, B. E. Hubbard, A. J. Sievers, B. Ilic, D. A. Czaplewski, and H. G. Craighead. Observation of locked intrinsic localized vibrational modes in a micromechanical oscillator array. *Phys. Rev. Lett.*, 90(4):044102, Jan 2003. doi: 10.1103/PhysRevLett.90.044102.
- R. D. Shachter and Kenley. Gaussian influence diagrams. *Management Science*, 35:527–550, 1989.

- M. Singh and M. Valtorta. Construction of Bayesian network structures from data: A brief survey and an efficient algorithm. *International Journal of Approximate Reasoning*, 12: 294–299, 1995.
- J.-J. E. Slotine and W. Lohmiller. Modularity, evolution, and the binding problem: a view from stability theory. *Neural Networks*, 14(2):137–145, 2001.
- P. Spirtes and C. Meek. Learning Bayesian networks with discrete variables from data. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, pages 294–299, 1995.
- P. Spirtes, C. Glymour, and R. Scheines. Causation, prediction and search. In *Lecture Notes in Statistics 81*. Springer Verlag, New York, 1993.
- S. H. Strogatz. *Sync: Emerging Science of spontaneous order*. Hyperion, 2003.
- J. Suzuki. A construction of Bayesian networks from databases based on the MDL principle. In *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence*, pages 294–299, 1993.
- P. Tass, M. G. Rosenblum, J. Weule, J. Kurths, A. S. Pikovsky, J. Volkmann, A. Schnitzler, and H.-J. Freund. Detection of n:m phase locking from noisy data: application to magnetoencephalography. *Physical Review Letters*, 81(15):3291–3294, 1998.
- P. Thiran. *Dynamics and self-organisation of locally coupled neural networks*. Presses Polytechniques et Universitaires Romandes, Lausanne, 1997.
- Patrick A. Truitt, Jared B. Hertzberg, C. C. Huang, and Keith C. Schwab Kamil L. Ekinici and. Efficient and sensitive capacitive readout of nanomechanical resonator arrays. *Nano Letters*, 7(1):120–126, 2007.
- Mauro Ursino, Elisa Magosso, and Cristiano Cuppini. Recognition of abstract objects via neural oscillators: interaction among topological organization, associative memory and gamma band synchronization. *Trans. Neur. Netw.*, 20(2):316–335, 2009. ISSN 1045-9227. doi: <http://dx.doi.org/10.1109/TNN.2008.2006326>.
- B. van der Pol. A theory of the amplitude of free and forced triode vibrations. *Radio Review*, 1:701–710, 754–762, 1920.
- T. Verma and J. Pearl. Causal networks: semantics and expressiveness. In R. D. Shachter, T. S. Lewitt, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, 4, pages 69–76. Elsevier Science, North-Holland, Amsterdam, 1990.

- A. Wagemakers, J. M. Buldú, J. García-Ojalvo, and M. A. F. Sanjuán. Synchronisation of electronic genetic networks. *Chaos*, 16:3127, 2006.
- K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo. Adaptive affinity propagation clustering. *Acta Automatica Sinica*, 33(12):1242–1246, 2007.
- W. Wang and J.-J. E. Slotine. On partial contraction analysis for coupled nonlinear oscillators. *Biological Cybernetics*, 92(1):38–53, 2004.
- X. Wei, M. F. Randrianandrasana, and D. Lowe. Nonlinear dynamics of a periodically driven duffing resonator coupled to a van der pol oscillator. *Mathematical Problems in Engineering*, 2010. In press.
- Y. Weiss. Belief propagation and revision in networks with loops. Technical Report AIM-1616, Dept. of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 1997.
- N. Wermuth and S. Lauritzen. Graphical and recursive models for contingency tables. *Biometrika*, 72:537–552, 1983.
- P. Woafu, J. C. Chedjou, and H. B. Fotsin. Dynamics of a system consisting of a van der pol oscillator coupled to a duffing oscillator. *Physical Review E*, 54(6):5929–5934, 1996.
- W. H. Wolberg and O. L. Mangasarian. Multisurface method of pattern separation for medical diagnosis applied to breast cancer cytology. In *Proceedings of the National Academy of Sciences*, volume 87, pages 9193–9196, 1990.
- C. W. Wu and L. O. Chua. A unified framework for synchronization and control of dynamical systems. *International Journal of Bifurcation and Chaos*, 4:979–988, 1994.
- W. Wu, W. Zhou, and T. Chen. Cluster synchronisation of linearly coupled complex networks under pinning control. *IEEE Transactions on Circuits and Systems–I: Regular papers*, 56(4):829–839, 2009.
- F. Xie and G. Hu. Spatiotemporal periodic and chaotic patterns in a two-dimensional coupled map lattice system. *Physical Review E*, 55:79–86, 1997.
- Y. Yamaguchi and H. Shimizu. Theory of self-synchronization in the presence of native frequency distribution and external noises. *Physica D*, 11(212):212–226, 1984.
- S. H. Yook, H. Jeong, A.-L. Barabási, and Y. Tu. Weighted evolving networks. *Physical Review Letters*, 86:5835–5838, 2001.

-
- M. K. Zalalutdinov, J. W. Baldwin, M. H. Marcus, R. B. Reichenbach, J. M. Parpia, and B. H. Houston. Two-dimensional array of coupled nanomechanical resonators. *Applied Physics Letters*, 88:143504, 2006a.
- Maxim K. Zalalutdinov, Jeffrey W. Baldwin, Martin H. Marcus, Robert B. Reichenbach, Jeevak M. Parpia, and Brian H. Houston. Two-dimensional array of coupled nanomechanical resonators. *Appl. Phys. Lett.*, 88:1143504, 2006b.
- J. Zhang. Selecting typical instances in instance-based learning. In *Proceedings of the Ninth International Machine Learning Conference*, pages 470–479, Aberdeen, Scotland, 1992. Morgan Kaufmann.
- C. Zhou, , and J. Kurths. Spatiotemporal coherence resonance of phase synchronisation in weakly coupled chaotic oscillators. *Physical Review E*, 65:040101, 2002.
- C. Zhou, A. E. Motter, and J. Kurths. Universality in the synchronization of weighted random networks. *Physical Review Letters*, 96:034101, 2006.