Networks of Coupled VO$_2$ Oscillators for Neuromorphic Computing

Présentée le 28 avril 2021

Faculté des sciences et techniques de l'ingénieur
Laboratoire des dispositifs nanoélectroniques
Programme doctoral en génie électrique

pour l'obtention du grade de Docteur ès Sciences

par

Elisabetta CORTI

Acceptée sur proposition du jury

Prof. E. D. N. Matioli, président du jury
Prof. M. A. Ionescu, Dr S. Karg, directeurs de thèse
Prof. S. Datta, rapporteur
Prof. B. Linares-Barranco, rapporteur
Prof. C. Enz, rapporteur
Acknowledgements

This work would have never been possible without the guidance and the support of colleagues, family and friends which helped me walking this path and keeping up with the difficulties as well the success of this exciting field of research.

In particular, I would like to thank my supervisor Dr. Siegfried Karg for his support during the past three and a half years. I really appreciated how you followed me closely when I first joined the lab and dedicated so much of your time to train me in all the aspects of the work which were unfamiliar to me. At the same time, you gave me the freedom to try my own ideas and to learn from mistakes. Your guidance in this project was essential for all the progress we made.

My deepest gratitude goes to my manager, Dr. Kirsten Moselund, for establishing a collaborative and motivating environment in our group, as well as for the many technical discussions which greatly contributed to the results presented in this thesis. Thank you for your support and the sincere interest you take in creating an inspiring, friendly and also fun working atmosphere.

I would like to thank Prof. Adrian Mihai Ionescu for his academic supervision and valuable advice. The fresh perspective you brought to our exchanges has been a great contribution to the research.

Special thanks to Dr. Bernd Gotsmann, for all the time he dedicated to my project, for the sincere feedback on my work and for keeping my spirits up when I was facing this or that challenge in my research. Thank you also for all the coffee breaks you organized in our neighborhood in the home office times, they have been the highlight of my weeks while I was drafting this thesis.

Thank you to the members of the PhD jury, Prof. Suman Datta from Notre Dame University, Prof. Christian Enz and Prof. Elison Matioli from EPFL, Prof. Bernabé Linares-Barranco from the Seville Microelectronics Institute, which took the time to evaluate this work and for their helpful feedback.

I am grateful to the S&T department manager Dr. Heike Riel for closely following my research and providing valuable advice.

I would like to acknowledge the work of two formidable students, Joaquin Cornejo and Federico Balduini, which joined me in this path during the period of their Master’s theses. You started your work at the beginning of the pandemic, but despite the confusion and delays
you were both able to convey extraordinary results in this research. It was a pleasure to work with you.

A special acknowledgment to Abhishek Khanna from Notre Dame University. During your short period in the lab you accelerated my work, shared with me long evenings in the lab, powered by low quality pizzas and coffees to see a very important part of the results coming to life.

I am also grateful to our collaborators, which provided me with the material to fabricate the devices presented in this thesis. Many thanks Matteo Cavalieri from EPFL for your PLD work. A very special thank to the team in Cambridge University, Prof. John Robertson, Dr. Kham Niang, Guadong Bai and Haichang Lu. Not only you have exchanged with me valuable ideas and samples, but you also hosted me as a visitor scientist in your lab. It was a great experience, a lot of good work and also quite some fun.

Thank you to our collaborators in the NeurONN project, for scientific discussions and the ideas they brought to the project. In particular, I am grateful to Dr. Aida Todri-Sanial, for additionally providing such an exceptional organization to the project. A sincere acknowledgment also to the other partners and in particular to Corentin Delacour, Dr. Stefania Carapezzi, Prof. Maria J. Avedillo, Juan N. Martinez and Dr. Ahmed Nejim for the insightful exchange and for the essential expertise they brought to the project.

I would like to acknowledge the partners in the Phase Change Switch project, for providing important insights on the material characteristics and the device processing. Many successful ideas rise from discussions with experienced colleagues. I would like to thank Dr. Bert Offrein and Dr. Abu Sebastian, together with their teams in IBM for the multiple exchanges and the scientific input they provided me during the past three years.

This work would have not been possible without the combined effort of the engineers and scientists which provide support and take care of the numerous tools that are needed for our research. Special thanks to my colleague and friend Marilyne Sousa, for performing ellipsometry and TEM characterization on my samples. To Ronald Grundbacher, for setting up and optimizing the ALD deposition. To Steffen Reidt, for performing unconventional anneals in the PLD tool. To Ute Dreschler, Dr. Diana Davila Pineda and Daniele Caimi for all their advice and help in the cleanroom processes.

I was very lucky to be part of an amazing team of smart colleagues and friends at IBM. Thank you Heinz, for your technical input and for the discussions about technology and life. Fabian, Federico and Nele, thank you for your work on the SThM set-up and for the incredible characterization you provided on the devices. Pengyan, I am grateful for your work on the device simulations. I would like to thank Clarissa and Cezar, my two office mates, with whom I shared so many serious and fun discussions and to the other half of my lab, Svenja, Preksha and Markus S., for the reciprocal support when our experiments failed and for cheering with me when results came out for a change. Thank you to my skilled colleagues and partners in crime in parties and adventures, Philipp, Fabian, Morten, Tushar, Preksha and Andrea, for the good times inside and outside work. A special acknowledgment to Markus R. for being present when he would have rather not, to Daniel for the young fresh air he brought to our lab and to Simon for the interesting experiments and conversations. Thank you to the friends and
colleagues Noelia, Alan and Kasia, for the travels, the ski adventures and the amazing dinners we had together. Most importantly, I would like to acknowledge my family for being with me through thick and thin. In particular I thank my parents Marialuisa and Maurizio, Nonna Rosanna and my sister Francesca, for their continuous support through my long study path. Special thanks to our friends Carla and Fabio, for shipping multiple packages of essential Italian food which gave me the energy to write this thesis. Lastly, thank you to Philipp, for sharing this adventure with me and to all the friends in Italy and in other parts of the world, for being there for me.

Zurich, April 10, 2021

Elisabetta
Abstract

Neuromorphic computing is a wide research field aimed to the realization of brain-inspired hardware, apt to tackle computation of unstructured data more efficiently than currently done with standard computer architectures. Oscillatory neural networks are known for their associative memory capability, which enables to retrieve the information stored in the system from noisy or incomplete data. The development of phase-transition materials such as vanadium dioxide (VO$_2$) allows to design compact relaxation oscillator units which can be coupled in frequency and phase to realize an oscillatory neural network in hardware. In this thesis, we investigate the oscillatory neural network technology from the realization of the basic oscillator components with VO$_2$ to the exploitation of the coupled oscillators as analog filters in convolutional neural networks applications.

VO$_2$ phase-transition devices are realized in a CMOS compatible process in two geometries, a planar and a crossbar configuration. The impact of the polycrystallinity of the VO$_2$ film on the insulator-to-metal transition of the device is analyzed; through the contacting of a single grain we demonstrate the realization of a VO$_2$ device with a single, sharp phase transition. The VO$_2$ devices are connected in circuits to build networks of coupled oscillators. Through coupling with resistive and capacitive elements, experimental demonstrations of a 4-VO$_2$ coupled oscillator network is shown. The network encodes the input and output information in the relative phase of the oscillators. The associative memory capability of the system is used to extract features from hand-written digits. By expanding the network to a 3×3 coupled oscillator system, we demonstrate in simulations how an oscillatory neural network can replace up to five digital filters in a convolutional neural network, retaining the same image processing capabilities.

**Keywords:** oscillatory neural network • vanadium dioxide (VO$_2$) • relaxation oscillator • frequency locking • phase locking • associative memory • neuromorphic computing • convolutional neural networks • time encoded information
Zusammenfassung

Neuromorphic Computing ist ein weites Forschungsfeld, das auf die Realisierung von hirn-inspirierter Hardware abzielt und dazu geeignet ist, unstrukturierte Daten effizienter als mit derzeitigen Prozessoren zu berechnen. Oszillierende neuronale Netzwerke sind für ihre assoziativen Speicherfähigkeiten bekannt, die es ermöglichen, die im System gespeicherten Informationen aus verrauschten oder unvollständigen Daten zu gewinnen. Die Nutzung von Phasenübergangsmaterialien wie Vanadiumdioxid (VO₂) ermöglicht den Entwurf kompakter Relaxationsoszillatoreinheiten, die in Frequenz und Phase gekoppelt werden können, um ein oszillierendes neuronales Netzwerk in Hardware zu realisieren. In dieser Arbeit untersuchen wir die oszillierende neuronale Netzwerktechnologie von der Realisierung der grundlegenden Oszillatorkomponenten mit VO₂ bis zur Nutzung der gekoppelten Oszillatoren als analoge Filter in Anwendungen von faltenden neuronalen Netzwerken.


Stichwörter: Oszillierende neuronale Netzwerke • Vanadiumdioxid (VO₂) • Relaxationsoszillatoreinheiten • Frequenz locking • Phase locking • assoziativen Speicherfähigkeiten • neuromorphic computing • convolutional neural networks • zeitcodierte Information
Contents

Acknowledgements iii

Abstract vii

Zusammenfassung ix

List of Publications xiii

List of Figures and Tables xv

List of Acronyms xix

1 Introduction 1

1.1 The Computational Challenge of Deep Learning . . . . . . . . . . . . . . . . . . . 1
1.2 Neuromorphic Computing: an Overview of the State of the Art . . . . . . . . . . . 3
  1.2.1 Biologically-Inspired Platforms . . . . . . . . . . . . . . . . . . . . . . . . . 3
  1.2.2 Neuromorphic Accelerators for Neural Networks . . . . . . . . . . . . . . 4
  1.2.3 Beyond Neural and Synaptic Behaviors . . . . . . . . . . . . . . . . . . . . 5
  1.2.4 Motivation for Oscillatory Neural Networks . . . . . . . . . . . . . . . . . . 6
1.3 Aim of the Thesis . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

2 VO₂ Oscillators: from the Material to the Applications 11

2.1 Vanadium Dioxide . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
  2.1.1 VO₂ Fabrication . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
  2.1.2 VO₂ Electrically-Triggered Transition . . . . . . . . . . . . . . . . . . . . . . 16
2.2 VO₂ Relaxation Oscillators . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
  2.2.1 Coupled Oscillators . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 21
2.3 Oscillatory Neural Networks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22
  2.3.1 Hopfield Neural Network . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
  2.3.2 Models of Oscillatory Neural Networks . . . . . . . . . . . . . . . . . . . . . 24
2.4 Technologies that Compute with Oscillators: an Overview . . . . . . . . . . . . . 30
  2.4.1 Materials and Devices for Oscillatory Neural Networks . . . . . . . . . . . . 30
  2.4.2 Image Processing with Oscillators . . . . . . . . . . . . . . . . . . . . . . . . 32
# Contents

2.5 Convolutional Neural Networks ................................................. 35

3 Experimental Methods ......................................................... 41
  3.1 Device Fabrication .......................................................... 41
    3.1.1 Deposition Techniques ............................................... 41
    3.1.2 Raman Spectroscopy .................................................. 44
    3.1.3 Annealing Techniques ................................................ 46
    3.1.4 Device Processing ..................................................... 49
  3.2 Scanning Thermal Microscopy ............................................. 51
  3.3 Device Model for Circuit Simulations .................................... 53

4 Characterization of the Phase Transition in scaled VO$_2$ Devices .... 55
  4.1 Characterization of VO$_2$ Planar Devices ............................... 55
    4.1.1 Characterization via Scanning Thermal Microscopy .............. 59
    4.1.2 Simulation of Phase Transitions in Planar Devices ............ 61
  4.2 Crossbar Devices ........................................................... 63
  4.3 Single Grain Devices ....................................................... 65
  4.4 Main Achievements .......................................................... 69

5 Coupled Oscillator Networks based on VO$_2$ Devices ..................... 71
  5.1 VO$_2$ Oscillators: Characteristic and Performances .................. 71
  5.2 Coupled Oscillators ........................................................ 72
    5.2.1 Two Coupled Oscillators based on VO$_2$ Devices ............... 74
    5.2.2 Pattern Recognition with Three Coupled Oscillators .......... 78
    5.2.3 Feature Edge Extraction with Four Coupled Oscillators ....... 81
  5.3 Main Achievements .......................................................... 84

6 VO$_2$ Coupled Oscillators as Filters in Convolutional Neural Networks .. 85
  6.1 VO$_2$ Coupled Oscillators as Analog Filters in Convolutional Neural Networks .................................................. 85
  6.2 Backpropagation Algorithm applied to the ONN ....................... 88
  6.3 Phase-Detector Circuit for ONN Second Layer .......................... 90
  6.4 Benchmark ....................................................................... 93
  6.5 Main Achievements .......................................................... 94

7 Conclusion and Outlook .......................................................... 95
  7.1 Future Directions .............................................................. 99

A Appendix ................................................................................. 101
  A.1 SThM Characterization ..................................................... 101
  A.2 Simulation of Multi-Grain Switching in VO$_2$ Devices ............... 103

References .................................................................................. 107

Curriculum Vitae .......................................................................... 133
List of Publications

Most of the results presented in this thesis have been published in peer-reviewed scientific journals. For the papers presented here, I planned the experiments, performed the majority of the experimental work and data analysis, and wrote the largest part of the manuscripts:


In the scope of this thesis I filed 2 patent applications:


List of Publications

At the same time I contributed to several other projects, which are not included in this thesis. A continued collaboration within the NeurONN European Project resulted in the following manuscripts (in preparation):


# List of Figures and Tables

<table>
<thead>
<tr>
<th>Figure/Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Coupled Oscillators as Hardware Accelerators for Neural Networks in the Neuro-morphic Computing Landscape</td>
<td>6</td>
</tr>
<tr>
<td>2.1</td>
<td>Properties of Vanadium Oxides</td>
<td>11</td>
</tr>
<tr>
<td>2.2</td>
<td>VO$_2$ structural phase change</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>VO$_2$ band diagram</td>
<td>13</td>
</tr>
<tr>
<td>2.4</td>
<td>Representation of a Peierls instability</td>
<td>15</td>
</tr>
<tr>
<td>2.5</td>
<td>Current-voltage characteristic of a VO$_2$ device</td>
<td>17</td>
</tr>
<tr>
<td>2.6</td>
<td>Oscillator based on a VO$_2$ Device</td>
<td>19</td>
</tr>
<tr>
<td>2.7</td>
<td>Frequency-locking of two coupled oscillators</td>
<td>22</td>
</tr>
<tr>
<td>2.8</td>
<td>Periodic attractor in an oscillatory neural network</td>
<td>25</td>
</tr>
<tr>
<td>2.9</td>
<td>Image recognition performed with a system of coupled oscillators</td>
<td>27</td>
</tr>
<tr>
<td>2.10</td>
<td>Fully-Connected and Frequency-Modulated Oscillatory Neural Networks</td>
<td>28</td>
</tr>
<tr>
<td>2.11</td>
<td>Benchmark of oscillators technologies</td>
<td>31</td>
</tr>
<tr>
<td>2.12</td>
<td>Image filtering operations performed with ONNs</td>
<td>34</td>
</tr>
<tr>
<td>2.13</td>
<td>Summary of demonstration of image filtering operations performed with ONNs</td>
<td>35</td>
</tr>
<tr>
<td>3.1</td>
<td>Scheme of a PLD system</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>Characteristics of VO$_2$ PLD films</td>
<td>43</td>
</tr>
<tr>
<td>3.3</td>
<td>Scheme of an ALD process</td>
<td>44</td>
</tr>
<tr>
<td>3.4</td>
<td>Raman Spectra of Vanadium Oxides</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>Effect of the Annealing Time on VO$_2$ ALD Films</td>
<td>46</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary of flash anneal parameters</td>
<td>48</td>
</tr>
<tr>
<td>3.7</td>
<td>Comparison between flash annealed and slow annealed VO$_2$ films</td>
<td>48</td>
</tr>
<tr>
<td>3.8</td>
<td>Processing of VO$_2$ Planar and Crossbar Devices</td>
<td>50</td>
</tr>
<tr>
<td>Figure/Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.9</td>
<td>Simplified schematic of the SThM setup</td>
<td>52</td>
</tr>
<tr>
<td>3.10</td>
<td>VO₂ circuit model employed in simulations</td>
<td>54</td>
</tr>
<tr>
<td>4.1</td>
<td>Hysteresis curves of VO₂ planar devices</td>
<td>56</td>
</tr>
<tr>
<td>4.2</td>
<td>IV curve of a VO₂ planar device</td>
<td>58</td>
</tr>
<tr>
<td>4.3</td>
<td>SThM imaging of the filament formation in the IMT of VO₂ planar devices</td>
<td>59</td>
</tr>
<tr>
<td>4.4</td>
<td>Evolution of the metallic filament in a VO₂ planar device for increasing bias voltage</td>
<td>60</td>
</tr>
<tr>
<td>4.5</td>
<td>Simulation of a planar VO₂ device</td>
<td>62</td>
</tr>
<tr>
<td>4.6</td>
<td>Illustration of a planar and a crossbar device</td>
<td>63</td>
</tr>
<tr>
<td>4.7</td>
<td>Hysteresis curve of crossbar devices with different dimensions</td>
<td>64</td>
</tr>
<tr>
<td>4.8</td>
<td>I-V characteristic of a crossbar device</td>
<td>64</td>
</tr>
<tr>
<td>4.9</td>
<td>Electrical characterization of a VO₂ single-grain device</td>
<td>66</td>
</tr>
<tr>
<td>4.10</td>
<td>TEM Characterization of a VO₂ single-grain device</td>
<td>67</td>
</tr>
<tr>
<td>4.11</td>
<td>IV curve of a VO₂ single-grain device</td>
<td>68</td>
</tr>
<tr>
<td>5.1</td>
<td>Circuit realizations of a single oscillator unit</td>
<td>72</td>
</tr>
<tr>
<td>5.2</td>
<td>Characteristics of the best-performing devices</td>
<td>73</td>
</tr>
<tr>
<td>5.3</td>
<td>Schematic of the oscillatory neural network circuit</td>
<td>74</td>
</tr>
<tr>
<td>5.4</td>
<td>Resistively-Coupled Oscillators Circuit</td>
<td>75</td>
</tr>
<tr>
<td>5.5</td>
<td>Experiment of phase encoding in a two-resistively-coupled oscillators circuit</td>
<td>75</td>
</tr>
<tr>
<td>5.6</td>
<td>Effect of VO₂ multi-step switching on the oscillating waveform</td>
<td>76</td>
</tr>
<tr>
<td>5.7</td>
<td>Grey-scale pixel encoding in two-coupled oscillators</td>
<td>78</td>
</tr>
<tr>
<td>5.8</td>
<td>Time-Delay inference with two-coupled oscillators</td>
<td>79</td>
</tr>
<tr>
<td>5.9</td>
<td>Pattern recognition with 3-coupled oscillators</td>
<td>80</td>
</tr>
<tr>
<td>5.10</td>
<td>Edge extraction experiments with 4-coupled oscillators</td>
<td>81</td>
</tr>
<tr>
<td>5.11</td>
<td>Edge extraction simulations with 4-coupled oscillators</td>
<td>83</td>
</tr>
<tr>
<td>6.1</td>
<td>Structure of the CNN under consideration</td>
<td>87</td>
</tr>
<tr>
<td>6.2</td>
<td>A single ONN performs the filtering actions of 5 digital CNN filters</td>
<td>87</td>
</tr>
<tr>
<td>6.3</td>
<td>Schematic of the backpropagation algorithm for ONNs</td>
<td>89</td>
</tr>
<tr>
<td>6.4</td>
<td>Phase detector and connector circuit between two ONN layers</td>
<td>92</td>
</tr>
<tr>
<td>7.1</td>
<td>Summary of results</td>
<td>97</td>
</tr>
<tr>
<td>7.2</td>
<td>Summary of the achievements of this work compared to the state of the art</td>
<td>99</td>
</tr>
</tbody>
</table>
List of Figures and Tables

A.1  IV Characteristic of a device measured with the SThM  ..................................... 102
A.2  SThM Measurement of the Metallic Filament Expansion in a VO$_2$ Device .... 102
A.3  Example of discrimination between metallic and insulating paths in VO$_2$ planar devices with SThM characterization ................................................................. 104
A.4  Parameters for the simulation of VO$_2$ granular devices ............................. 105
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASIC</td>
<td>application-specific integrated circuit</td>
</tr>
<tr>
<td>AFM</td>
<td>atomic force microscopy</td>
</tr>
<tr>
<td>ALD</td>
<td>atomic layer deposition</td>
</tr>
<tr>
<td>CMOS</td>
<td>complementary metal-oxide-semiconductor</td>
</tr>
<tr>
<td>CNN</td>
<td>convolutional neural network</td>
</tr>
<tr>
<td>CPU</td>
<td>central processing unit</td>
</tr>
<tr>
<td>DL</td>
<td>deep learning</td>
</tr>
<tr>
<td>DNN</td>
<td>deep neural network</td>
</tr>
<tr>
<td>DOM</td>
<td>degree of match</td>
</tr>
<tr>
<td>DTC</td>
<td>digital to time converter</td>
</tr>
<tr>
<td>FA</td>
<td>flash anneal</td>
</tr>
<tr>
<td>FC</td>
<td>fully-connected</td>
</tr>
<tr>
<td>FFT</td>
<td>fast Fourier transform</td>
</tr>
<tr>
<td>FM</td>
<td>frequency modulated</td>
</tr>
<tr>
<td>FPGA</td>
<td>field-programmable gate array</td>
</tr>
<tr>
<td>GPU</td>
<td>graphics processing unit</td>
</tr>
<tr>
<td>ICP</td>
<td>inductively coupled plasma</td>
</tr>
<tr>
<td>IMT</td>
<td>insulator to metal transition</td>
</tr>
<tr>
<td>I-V</td>
<td>current voltage characteristic</td>
</tr>
<tr>
<td>MEMS</td>
<td>micro electro-mechanical system</td>
</tr>
<tr>
<td>MIT</td>
<td>metal to insulator transition</td>
</tr>
<tr>
<td>NEMS</td>
<td>nano electro-mechanical system</td>
</tr>
<tr>
<td>ONN</td>
<td>oscillatory neural network</td>
</tr>
<tr>
<td>PCM</td>
<td>phase change memory</td>
</tr>
<tr>
<td>PLD</td>
<td>pulsed laser deposition</td>
</tr>
<tr>
<td>RAM</td>
<td>random access memory</td>
</tr>
<tr>
<td>RO</td>
<td>ring oscillator</td>
</tr>
<tr>
<td>RRAM</td>
<td>resistive random access memory</td>
</tr>
<tr>
<td>R-T</td>
<td>resistance temperature characteristic</td>
</tr>
<tr>
<td>SEM</td>
<td>scanning electron microscopy</td>
</tr>
<tr>
<td>SNN</td>
<td>spiking neural network</td>
</tr>
<tr>
<td>STDP</td>
<td>spike-timing-dependent plasticity</td>
</tr>
<tr>
<td>STEM</td>
<td>scanning transmission electron microscopy</td>
</tr>
<tr>
<td>SThM</td>
<td>scanning thermal microscopy</td>
</tr>
<tr>
<td>STO</td>
<td>spin torque oscillator</td>
</tr>
<tr>
<td>TDC</td>
<td>time to digital converter</td>
</tr>
<tr>
<td>TEM</td>
<td>transmission electron microscopy</td>
</tr>
</tbody>
</table>
### List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEMA V</td>
<td>tetrakis(ethylmethylamino)-vanadium</td>
</tr>
<tr>
<td>VT OP</td>
<td>vanadium triisopropoxide</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 The Computational Challenge of Deep Learning

The 21st century, with the development of advanced electronic technologies, has seen an exponential increase in the amount of data that is daily created and processed worldwide. A recent report from the International Data Corporation [1] predicts that 59 zettabytes of data will be produced, copied or consumed only this year. With the growth of edge computing, it is estimated that in the next three years we will further create more data than what has been generated in the last thirty years worldwide [1]. In parallel to the extended availability of information, machine learning algorithms have arisen with the aim to extract relevant knowledge from diverse and unstructured data. These algorithms, which rely on multi-level layered networks and back propagation training, have been devised theoretically in the 80's. However, as they require high computational power and a large amount of data for training, they have been only recently brought to commercial use. The advances in hardware technologies, which enabled to have faster, more powerful computing machines, together with the increased data-availability, allowed the exploitation of machine learning, and in particular, deep neural networks to perform data analysis and information extraction that was not easily implementable in the past. Therefore, the need of processing the spectacular amount of data available today required the employment of deep-learning techniques. At the same time, deep learning would have never been possible without the availability of large datasets [2].

Deep learning is a branch of machine learning which uses brain-inspired concepts to perform learning of a task on a machine. Deep learning nowadays powers many aspects of our society: it is used in web searches to select relevant results, content filtering and personalized advertisements in social networks, facial recognition and speech recognition. Moreover, it is successfully employed also for aiding the synthesis of new drugs [3], for diagnostics [4] and cybersecurity [5]. These algorithms are based on a representation-learning method for which raw data is fed into a network. The network automatically computes and discovers the set of features needed to perform a classification or a detection task [6]. The deep neural network algorithms are constructed on a set of subsequent layers, which are composed by
simple, non-linear modules. Each module transforms the raw data from the previous level, performing a weighted sum of the various inputs, which on a high level resembles the accumulation of spikes from the incoming synapses performed by neurons in the human brain. After undergoing a non-linear activation function, the output of each layer results in a more abstract representation of the information; the last module structures the output information as, for example, a classification vector for an image classification application. The training of the network consists in the tuning of the weights between the network nodes, usually performed by a back-propagation algorithm. Per definition, the higher the number of layers in the network, the deeper the network is. It has been shown that deeper networks perform with better accuracy than shallow networks [7]. Naturally, the bigger the network, the higher is the number of weights to be trained. State-of-the-art deep neural networks, such as the VGG [8] or ResNet [9] architectures, comprise hundreds of layers and millions of weights [10]. The training of such networks becomes extremely energy and time consuming and requires to run on powerful or specialized hardware, such as graphic unit cards (GPUs) or field programmable gate arrays (FPGA). The success of GPUs as a platform to run neural network algorithms is found in the parallelization of the basic operation behind neural networks computation, the vector-matrix multiplication [11]. The best performing NVIDIA GPU, the Tesla A100, released in May 2020 with a 7 nm transistor technology, can operate at 312 GFLOPS/W for single-precision floating-point data format. It is reported that a ResNet-101 (where 101 refers to the number of layers in the network) can be trained on the ImageNet dataset [12] in about 5 hours using 8 of such GPUs [13], for a total power consumption of 16 kWh. This power consumption is almost four times what an European household needs in average for one day [14] and can increase steeply when the training is performed on older generation hardware platforms [15].

One of the major limitations of standard computer architectures in performing deep learning computation is the so-called von Neumann bottleneck, that results from the physical separation between the memory and the computation units. The imbalance arising from the difference between the speed of the computation and the speed of data retrieval from the Random Access Memory (RAM), causes the processor to remain idle during the time needed to access the RAM or the cache memories [16, 17, 18, 19]. Moreover, the access to the memory is also responsible for a high power consumption [20]. For each operation in a deep neural network, the data to be processed, together with the connection weights of the layers, need to be transferred from the memory to the computational unit. The result of this operation is then moved back to the memory. These algorithms require therefore an extensive usage of memory resources, which contributes to increase the time needed for performing the computation.

In comparison, the human brain can perform tasks as recognition and classification much more efficiently than modern processors. Counting around $10^{10}$ neurons and $10^{15}$ synapses, the massively parallel, plastic structure of the brain serves as an inspiration to design novel hardware architectures, or neuromorphic chips, which are devised to bring the memory and the computation unit close together, with the aim of increasing the computing performance.
1.2 Neuromorphic Computing: an Overview of the State of the Art

The concept of neuromorphic computing was first proposed by Carver Mead in 1990 [21] and relies on using analog electronic circuits and systems to mimic the biological architecture of the human brain. Multiple approaches to realize specialized hardware for neuromorphic computing applications are currently under investigation, with the aim of designing a highly parallel, interconnected and reconfigurable system which does not suffer from the separation between computational and memory unity. In [11], Kendall et al. have identified 10 characteristics for a successful neuromorphic hardware: the neuromorphic system should be highly parallel (1) and perform in-memory computing (2), therefore storing the information next to the computational units; it should perform analog, low precision but noise resilient computation similarly to the human brain (3); therefore, it should be accepted that its output will be a probabilistic solution (4) and should allow for mistakes produced by causality (5); plasticity (6) is needed to reconfigure the state of the system and enable learning (7); non-linearity (8) is essential to reproduce neural network behaviors; finally the system should be highly scalable (9) and implement sparsity (10) of the neuron-to-neuron connections. Building a system that respects all these characteristics is very challenging and there is no clear direction on which architecture is ultimately more apt to embed the expectation of neuromorphic hardware. Neuromorphic engineering is a very wide research field, featuring thousands of contributions which often follow diverging paths. The research attention nowadays spans from new systems and architecture designs, to the re-thinking of silicon based technology, in favor of novel materials and devices which offer more flexibility for an analog approach to computing [22, 23]. One common ground among neural network algorithms and the various computing platforms, including the oscillating neural network concept discussed in this thesis, is the original inspiration to the massively-parallel architecture of the brain. Following the steps of [24, 11], we distinguish between research projects that aim to build hardware accelerators for deep neural networks and those more directly inspired by neuroscience.

1.2.1 Biologically-Inspired Platforms

The primary goal of biologically-inspired platforms is the emulation of large-scale biological neural networks. These platforms often rely on the hardware implementation of spiking neural networks (SNNs), in which the information is communicated through asynchronous and sparse binary events, or spikes, between neurons [25]. The neurons are connected via synaptic weights, whose value can be tuned during the training of the SNN through the spike-timing dependent plasticity (STDP) concept [26, 27]: a synaptic connection between two neurons is strengthened when the post-synaptic neuron spikes after the pre-synaptic neuron; vice versa, it is weakened when the post-synaptic neuron spikes before the pre-synaptic one. With STDP, contrary to what happens in deep neural networks (DNNs), it is possible to implement unsupervised learning; however, the STDP-based learning architectures so far devised are not as reliable as the supervised learning implemented with the backpropagation algorithm in DNNs [28].
Many realizations of SNNs have been developed with standard CMOS technology, offering prototypes containing millions of neurons and synapses. Between the most complex realizations, we mention IBM’s True North chip in 2014, counting one million neurons and 256 million synapses [29]; the SpiNNaker system, which counts low power ARM cores to perform real time simulations of SNNs [30, 31]; Loihi from Intel [32, 33], which counts 130’000 neurons and 130 million synapses. Even though the primary purpose of these implementations is the emulation of biological networks, they are also increasingly considered as accelerators for deep neural networks algorithms, even though so far they demonstrated lower accuracy than DNNs running on standard GPU hardware [24, 34, 35, 36, 37].

Special attention is also given to the realization of the single neuron or synapse unit [38, 27]. Many efforts have been carried out to design neuron models in CMOS technology, spanning from biologically-plausible designs [39, 40] to simpler compact models [41, 42]. With CMOS technology, multiple transistors are usually required to implement a spiking element. As an alternative, emerging technologies have been proposed, which can reproduce the spiking behavior with single compact components, from electrochemical metallization neurons [43, 44], to resistive random access memory (RRAM) technologies [45, 46], to volatile phase-transition oxides neurons [47, 48]. Similarly, nanoscale emerging memory technologies which can encode multiple memory states have been studied to realize the synapses. Resistive RAMs [49, 50], phase change memories (PCM) [51], and magneto-resistive random-access memories (MRAM) [52] have the potential to improve the circuit integration density and to greatly reduce the power dissipation in neuromorphic systems [53, 54].

1.2.2 Neuromorphic Accelerators for Neural Networks

As mentioned before, in contrast to SNNs, DNNs lack of biological realism, favouring linear algebra techniques of vector-matrix multiplication combined with non-linear activation functions to compute. They are based on supervised learning through back-propagation techniques, which require large, labelled datasets and benefit from the highly parallel matrix multiplication techniques offered by modern GPUs. The second branch of neuromorphic engineering is devoted to the design of neuromorphic accelerators, which are fast, energy efficient platforms specialized for DNN algorithms and able to bridge the physical separation between the memory and the processor units.

From an architecture point of view, many accelerators have been produced bringing the computational units next to the memory, therefore increasing the bandwidth between CPU and the storage units [55]. In the past, the efforts of realizing in-processing memory by positioning dynamic RAMs and CPU close together have been hindered by the technological challenge of bridging the manufacturing differences between the processes utilized to realize them [56]. Recent advances in 3-D memory stacking pose a promising route in this direction [57, 58].

One of the most encouraging emerging concepts for neural networks hardware accelerators consists in implementing in-memory computing with memristive devices (RRAM, PCM,
1.2. Neuromorphic Computing: an Overview of the State of the Art

The idea comes from the consideration that analog resistive memory elements can perform the multiply-accumulate operations, which are at the heart of DNNs, exploiting the physical attributes of nanoscale devices. By building a crossbar array of memristive devices, the multiply operation can be performed at the crosspoint of each memory element by Ohm's law; similarly, the summation can be computed through the sum of currents following Kirchhoff's first law, as explained in [28]. Therefore, the memory itself, while storing the information, can also perform the computation. By avoiding the movement of data, it is expected to allow for a significant reduction of the power consumption and an increased computation speed [59, 60]. The entrance to the market of this technology is currently impeded by problems like non-linear conductance response, limited dynamic range, variability and drifts, which hinder the mapping of DNNs weights and the recognition performances of such platforms [59, 61, 62]. However multiple efforts are carried out to allow scaling of memristive arrays to large dimension [63] and to design more fault-tolerant approaches for mapping DNNs in these accelerators [64].

1.2.3 Beyond Neural and Synaptic Behaviors

Parallelism and in-memory computation are key ingredients in neuromorphic computing. To build a memory platform capable to perform also computation, it is of interest to understand how the biological memory works. In particular, it has been demonstrated that rhythmic firing of neurons is connected to long-term information processing in the memory [65, 66]. Following this observation, algorithms have been designed that use the evolving dynamics of non-linear systems to perform computation. In particular, Hopfield in 1980s proposed a recurrent, fully-connected network based on the collective behavior of simple non-linear elements (neurons) to perform several tasks, including image recognition [67]. This network has the unique function of possessing an associative memory, that can be programmed by tuning the connection weights between the neurons. The associative memory is able to identify accurately objects (or input electrical signals) even if they are contaminated by noise [68]. One of the possible realization of such a memory system is through the exploitation of the frequency and phase synchronization of coupled oscillators [69, 70]. Designing an Oscillatory Neural Network (ONN) in hardware is interesting as it yields the advantages of utilizing the same devices that are storing the patterns to perform the computing, therefore representing another kind of in-memory computing platform. Second, the nature of the associative memory provides a more robust way of pattern recognition compared to algorithms based on content-addressable memories, as it is highly resilient to input pattern distortion and noise [71]. Figure 1.1 attempts to put into context the research of ONN hardware with the major players in neuromorphic computing discussed above. Coupled oscillators systems are traditionally researched as hardware computing platforms for Oscillatory Neural Networks. In this work, we will bring the research a step forward, and investigate the possible exploitation of the associative memory of coupled oscillators as hardware accelerators for convolutional filters in neural networks.
Figure 1.1 | Brain-inspired computing is a wide technological field which comprises the design of neural networks algorithms and the realization of neuromorphic computing hardware. Neuromorphic computing platforms can be classified in hardware accelerators for neural network applications and biologically-inspired systems. An important example in the first category is the development of in-memory computing concepts such as the crossbar arrays for vector-matrix multiplication discussed in the previous section. The research of coupled oscillator hardware for ONNs also falls under this category. In particular, a principal investigation of this thesis concerns the exploitation of ONNs for convolutional neural networks applications. Spiking neural networks belong instead to the biologically-inspired platforms, which more closely mimic the human brain behavior. The neuromorphic computing field is however much more complex and varied than what exposed in this brief discussion and showed in this figure.

Apart from pattern retrieval, other problems can be successfully tackled by computing with Hopfield neural networks and oscillatory devices. For example, networks of chaotic oscillators can solve constrained optimization problems better than many state-of-the-art GPUs or ASIC designs [72]. In other works, it was shown that coupled oscillators can solve NP-hard combinatorial optimization problems such as vertex graph coloring [73] and the travelling salesman problem [74], or can be used as an Ising machine [75].

1.2.4 Motivation for Oscillatory Neural Networks

Compared to the more mature platforms described in the previous sections, the realization of neuromorphic computing hardware based on coupled oscillator technology is still in the first stage of development, and presents numerous challenges. In order to build an associative memory based on oscillators, it is necessary to build a network of oscillators, which can be synchronized with tunable coupling elements. Until now, image recognition with coupled oscillators requires relatively large networks, with many tunable interconnecting elements [76]. This demands for material and device research in order to fabricate reliable
oscillating devices, which can work over a high number of cycles and have low device-to-device variability. The attractiveness of oscillatory neural networks largely depends on designing a suitable oscillator as a building block. Even though they were invented more than 40 years ago, the research on ONNs has only recently gained momentum thanks to the advances in nanoscale device technology, which now allow to realize very compact, energy efficient oscillators based on the non-linearity of novel material such as phase-transition materials \cite{77,78,79} or spintronic nano-devices \cite{80,81,76}. It is argued that the physical realization of the oscillators is determinant for their success in real-life applications \cite{82}. Between the various technologies, oscillators realized with the phase-change material vanadium dioxide (VO$_2$) have been widely researched for ONN applications \cite{78,48,83,84}. VO$_2$ presents several characteristics that make it one of the top candidate for oscillatory neural network technology: it shows a phase transition with a large jump between the insulating and the metallic state \cite{85}; the phase transition happens near room-temperature, which is an important requirement for hardware applications; the phase transition can be triggered by an electrical stimuli \cite{86} and it has been proven to be ultra-fast \cite{87}; finally, VO$_2$ oscillators can be coupled with standard electrical components, such as capacitors \cite{78}.

Given the early stage of the research, it is challenging to precisely state the advantage of this technology compared to other very promising neuromorphic computing implementations. Almost no benchmarks have been drawn between oscillator-based computing versus other digital or analog solutions so far. As mentioned in \cite{82}, there are several reasons why researching this technology is interesting:

- Firstly, oscillatory neural networks are noise-tolerant: their associative memory is resilient to noisy or distorted input.

- Moreover, they allow for computing with the synchronization time or with the relative phase of the oscillators, which can be an advantage compared to systems working with a very scaled voltage power supply as they do not use the amplitude of the electrical variables of the system to process information.

- Lastly, they can be flexibly used for a wide range of possibly disruptive applications, from pattern retrieval to the solving of NP-hard problems.

Ultimately, networks of coupled oscillators require still extensive research spanning between various fields: from the materials and device, to fabricate reliable, power efficient oscillators; to the design of circuit for the interconnection and the read-out of the input and out signal; to the algorithms and computational models, to adapt them to the circuit and device technologies and to allow for practical fast, power efficient applications of these networks.
1.3 Aim of the Thesis

The scope of this thesis is to demonstrate that systems of coupled oscillators can be used as hardware accelerators for neural network algorithms. With specific attention to convolutional neural networks, we aim to demonstrate how the associative memory capabilities of an oscillatory neural network can be used to perform different filtering actions on an image, exploiting the fault-tolerant, time-encoded information processing of the oscillatory system.

This work combines the research of a novel oscillatory device, based on the phase change material vanadium dioxide (VO$_2$), with the circuit implementation of small networks of coupled oscillators. Starting from the realization of VO$_2$ devices, we explore their integration on a silicon platform and the scaling of their dimensions down to 70 nm. The phase-transition of the film in planar and crossbar devices is investigated, with special attention to the characteristic of the transition in the polycrystalline material and its impact on the device-to-device variability. We explore the frequency and phase synchronization of the compact oscillators with a coupling scheme realized with simple electrical components. In particular, we focus on the realization and encoding of the system memory with resistive coupling elements, which in perspective can be realized with the emerging memory technologies exploited for in-memory computation (RRAM, PCMs). With the aim of realizing a computing unit which works entirely through the timing of the electrical signals, we investigate the injection of the input information based on the relative delays between the voltage signals of the devices, rather than with amplitude-encoded schemes. Finally, we simulate a network design which can be integrated as an analog filter in convolutional neural networks, replacing several digital convolutional filters. The convolutional neural network accelerated by the oscillatory neural network hardware is tested on an image classification task, but can ultimately be flexibly used for other applications employing convolutional neural networks.

The thesis is structured as follows:

Chapter 1: Introduction

In the first chapter the topic has been introduced and placed in a wider scientific context.

Chapter 2: VO$_2$ Oscillators: from the Material to the Applications

The second chapter presents a theoretical overview of oscillatory neural network technology based on VO$_2$ devices. Starting from the analysis of the material, we discuss the origin of the phase transition and how to exploit it in order to realize compact relaxation oscillators, which can be coupled in frequency with simple electrical connections. We then offer a brief overview of the oscillatory neural networks, from the mathematical theory of their associative memory capabilities, to the available demonstrations of coupled oscillator computing. Finally, we introduce convolutional neural network algorithms and their application for image classification tasks.
Chapter 3: Experimental Methods

In this chapter the experimental methods which are relevant for this thesis are presented. Starting from the deposition of VO$_2$ thin films, we discuss the fabrication and characterization techniques of planar and cross-bar devices, including a scanning thermal microscopy technique for the mapping of the temperature inside the device. We briefly present the device model used for the circuit simulation.

Chapter 4: Characterization of the Phase Transition in Scaled VO$_2$ Devices

In this chapter the phase transition in planar and crossbar VO$_2$ devices is presented. Through electrical characterization and scanning thermal probe microscopy technique the impact of the polycrystallinity of the VO$_2$ film deposited on a SiO$_2$/Si substrate is analyzed. Finally, we present the realization of a sharp transition in a single-grain device.

Chapter 5: Coupled Oscillator Networks based on VO$_2$ Devices

In chapter 5 we present the experimental implementation of an Oscillatory Neural Network based on the phase-transition of VO$_2$ devices. The information is computed in the time-relations of the network signals. In particular, the storage of multiple output patterns in the relative phase of the oscillators is presented. Moreover, we offer experimental demonstrations of a computational scheme based on the encoding of the input information in the time-delays of the voltage signals in the oscillatory nodes. A demonstration of feature edge-extraction operated by a network of 4-VO$_2$ on Si oscillators is presented.

Chapter 6: VO$_2$ Coupled Oscillators as Filters in Convolutional Neural Networks

In chapter 6 the integration of oscillatory neural networks as analog filters in convolutional neural networks is discussed. Through a simulation framework, we demonstrate that a single oscillatory neural network unit is able to replace five convolutional filters. A backpropagation algorithm which can train the ONN is also presented. Lastly, envisioning the connection of the oscillatory neural unit in multi-layered networks, we examine a phase detector circuit which could buffer the information from the previous to the subsequent layer.

Chapter 7: Conclusion

The last chapter presents a conclusion of the thesis followed by an outlook for future developments.
2 VO₂ Oscillators: from the Material to the Applications

This chapter provides the fundamentals of devices based on vanadium dioxide. We discuss the origins of the phase transition of vanadium dioxide and how it can be exploited to build compact relaxation oscillators. We further present the theory on Oscillatory Neural Networks and review the state of the art of its technological applications. Finally, we briefly introduce the architecture of conventional convolutional neural networks employed for image recognition tasks.

2.1 Vanadium Dioxide

Vanadium oxides are strongly correlated materials, which have been extensively studied due to the possible applications of their insulating to metallic (IMT) transition [88]. They can be synthesized with standard deposition techniques, as chemical vapour deposition or atomic layer deposition. They present a variety of stoichiometries, which, given their different crystal structure and band diagrams, can be identified with standard Raman and X-ray spectroscopy characterization. When present, the insulator-to-metal phase transition happens at different temperature for the different oxidation states. An overview of the stable vanadium oxides that present an IMT is given in table 2.1.

Between the various stoichiometries, V₂O₃, VO₂ and V₂O₅ are the most researched for practical

<table>
<thead>
<tr>
<th>Compound</th>
<th>Crystal Structure</th>
<th>Transition Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>V₂O₃</td>
<td>monoclinic/trigonal</td>
<td>150 - 160 K</td>
</tr>
<tr>
<td>V₆O₂n-1</td>
<td>triclinic</td>
<td>250 - 70 K</td>
</tr>
<tr>
<td>VO₂</td>
<td>monoclinic-tetragonal rutile</td>
<td>340 K</td>
</tr>
<tr>
<td>V₂O₅</td>
<td>orthorhombic</td>
<td>none</td>
</tr>
</tbody>
</table>
Vanadium dioxide presents a phase transition around 340 K that comprises a change in the resistivity of the material of 3 to 5 orders of magnitude. The phase change is accompanied by a structural transition from a monoclinic to a rutile crystal structure. Adapted from [93] with permission from AAAS.

Applications. In particular, V$_2$O$_3$ is investigated for its well defined Mott phase transition characteristics [89], which happens at a temperature of around 150 K. Vanadium pentoxide (V$_2$O$_5$) is the most stable compound, that features the highest oxidation state of vanadium oxides [90]. It does not present an IMT transition, however it is extensively researched for its electrochromic applications [91, 92].

In recent years, vanadium dioxide (VO$_2$) attracted the interest of the scientific community given its near-room temperature insulator to metal (IMT) phase transition, which enables more practical applications for novel electronic devices. The phase transition in VO$_2$ happens in fact at a critical temperature $T_C = 68^\circ C$ (340 K) and it is accompanied by a change in its structural and electronic properties [94]. At room temperature, VO$_2$ presents a monoclinic (M1) crystal structure and semiconducting properties, with a resistivity of about 10 $\Omega$·cm, as depicted in figure 2.2. In this phase, the crystal structure sees a doubling of the unit cell, with the formation of V-V dimers that are responsible for the opening of the VO$_2$ band gap and consequently for the low conductance of the material. Above 340 K the material experiences an insulator to metal transition with a rise of several orders of magnitude in its electrical conductivity, accompanied by a structural transition to a tetragonal, rutile crystal structure [95, 93, 96, 97]. The phase change is volatile, meaning that when the external stimuli inducing the phase change is removed, the material goes back to its insulating state. In figure 2.3
2.1. Vanadium Dioxide

we can see a schematic representation of the VO$_2$ band diagram. The phase transition is determined by the vanadium $d$ electrons, whose bonding modifies the electronic structure near the Fermi level ($E_F$) across the phase transition. In fact, the V 3d orbitals hybridize with the O 2p orbitals, forming $\sigma$, $\pi$ and $d_{\parallel}$ orbitals. In the metallic state, the $d_{\parallel}$ orbitals are parallel to the rutile c-axis and not bonded. They partially overlap with $\pi^*$ orbitals and result not completely filled, therefore leading to the metallic properties of the material [98]. Across the phase transition, the formation of V-V dimers results in a splitting of the $d_{\parallel}$ orbitals in a bond and anti-bond configuration, in which the $d_{\parallel}^*$ is empty. In addition, the $\pi^*$ orbitals are shifted to higher energy [99]. The formed band gap is responsible for the semiconductive nature of the monoclinic VO$_2$.

Apart from the electrical conductivity, other properties of the material are affected by the phase transition, in particular optical properties [100, 101] and thermal properties like the thermal conductivity [102, 103] and the Seebeck coefficient [104]. This wide range of effects makes this material an attractive candidate for various applications, spanning from fast electrical switches [105], to optical devices [106, 107], to thermochromic coatings [108].

Researchers still debate about the nature of the phase transition of VO$_2$. The strong electron-electron interaction would classify the material along the Mott transition materials [109, 110, 111]. Following this theory, the study of the material usually proceeds from the Hubbard hamiltonian for strongly correlated fermions [112]:

$$ H = -\sum_{ij(\sigma)} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_{i} n_{i\uparrow} n_{i\downarrow}. $$

(2.1)

In this equation, $t_{ij}$ represents the probability for a non-interacting electron to hop between
sites \(i\) and \(j\). The operator \(c_{i\sigma}\) is the creation operator for the electron at site \(i\), and \(c_{i\sigma}^+\) the corresponding annihilation operator. This first summation term is a kinetic energy term that describes the energy needed for an electron to hop from site to site in the lattice, therefore from atom to atom. The second summation term takes instead into account the Coulomb repulsion between two electrons present in the same site \(i\). In particular, \(n_{i\sigma} = c_{i\sigma}^+ c_{i\sigma}\) represents the density of electrons with spin \(\sigma\) at a site \(i\), while \(U\) is the Coulomb energy cost for two electrons to occupy the same site. A qualitative explanation of the meaning of the Hubbard hamiltonian can be given considering the outermost electronic orbital in a one dimensional chain of atoms. In the non interacting limit, i.e. without considering the Coulomb repulsion (second term equal to 0), the first term would favor the sharing of the electrons among the lattice, therefore the material would be a metal. In contrast, when the second term is predominant, charge localization is favored. Consequently, for increasing values of \(U\), a material that is traditionally a metal can transition to an insulating behavior, with the opening of an energy gap in the band structure of the material proportional to the energy \(U\) [113, 114]. From this basic principle, materials which traditionally would be considered metallic present instead a transition to a so-called Mott insulator. It is to be noted that Mott transition materials are not subject to a structural phase change. However, as already briefly discussed, the VO\(_2\) IMT is accompanied by a structural transition, which suggests the presence of a Peierls instability in the material, caused by electrons-lattice interactions [115, 95]. The origins of a Peierls instability can be also qualitatively explained by considering a one-dimensional chain of atoms. Assuming the presence of one electron at each site of the chain and assuming the outermost energy band of the system half-filled, the Fermi wavevector falls at half of the Brillouin zone \(\pi/a\) (figure 2.4). If a periodic distortion is introduced in the lattice, bringing two atoms closer together of a factor \(\delta\), such that the lattice periodicity is doubled, a band gap opens exactly at the Fermi energy of the material, decreasing the electron energy. This decrease of the electron energy would compensate the increase of the lattice energy introduced by the distortions [116]. Two bands form, one filled and one empty, therefore changing the material from a metal to an insulator. In support of the Peierls transition is the fact that in the monoclinic phase the V-V bonds are dimerized and cause a doubling of the lattice constant compared to the rutile phase, therefore justifying a splitting of the bands.

Many studies have shown that the two contribution (Peierls and Mott mechanism), cannot really be separated and therefore attribute the phase change to an interplay between the two effects [118, 119]. However, more recent works have shown that the IMT in VO\(_2\) can happen without the occurrence of the structural phase transition [120, 121]. In addition, the IMT was obtained also by pure carrier injections, which would support the idea of a band splitting occurring because of the contrasting energy terms in the Hubbard hamiltonian, in the framework of the Mott transition [122]. Even though the origin of the transition is still debated, the change in conductivity of VO\(_2\) was recorded multiple times in experiments and achieved through various techniques, comprising electrical activation [123, 124, 125, 126], optical activation [127, 128], and strain [129, 130], for exploitation in various technological applications.


2.1.1 VO$_2$ Fabrication

The sharpness, as well as the width of the vanadium dioxide phase transition are greatly impacted by the quality of the material, and specifically by impurities and crystal defects [131]. It is recorded that the resistivity ratio between the insulating and the metallic state can be as high as $10^5$ in bulk [132]. However, the on/off ratio, the sharpness and the width of the hysteresis can be greatly impaired when fabricating VO$_2$ thin films [133]. VO$_2$ has been synthesized with various deposition techniques, such as pulsed layer deposition (PLD), chemical vapour deposition [134], atomic layer deposition (ALD) [135], sputtering [136] and sol-gel techniques [137]. The quality of the film and the impact on the phase transition for film deposited on different substrates has been extensively studied. For epitaxial growth, Al$_2$O$_3$ and TiO$_2$ are the most used substrates [138]. Film grown on Al$_2$O$_3$ report a very narrow hysteresis width and quite large on/off ratio. In particular, experiments have shown that the hysteresis width can be as narrow as 1°C when VO$_2$ is deposited with PLD on sapphire (1010) [132]. A narrow hysteresis is desirable for practical implementation of VO$_2$ devices, as it reduces the power needed for switching the phase. It is known that the phase transition characteristics of VO$_2$ can degrade in polycrystalline films, due to the presence of grain boundaries [139]. In particular, it has been shown that films which present smaller grains result in a widening of the hysteresis and a reduced on/off ratio, while bigger grains are associated with narrower hysteresis and larger on/off ratio [133].

In addition, stress introduced by lattice mismatch with the substrate has proven to have an effect on the temperature at which the phase transition occurs. In particular, tensile stress results in an increase of the transition temperature, while compressive strain brings to a decrease of the transition temperature. This has been demonstrated with VO$_2$ deposited on sapphire and on TiO$_2$ [140, 141]. When deposited on Si, the VO$_2$ relaxes in a polycrystalline film, where no stress in the lattice is introduced [142]. In particular, many works utilize SiO$_2$ as
a buffer layer, on top of which the VO\textsubscript{2} breaks in granular films. As in this case no tensile stress is present, the phase transition is reported to happen at 68° C, however, it is affected by the widening of the hysteresis and the reduction of the on/off ration mentioned above [143].

Apart from stress, also doping can change the temperature transition of VO\textsubscript{2}. Introduction of dopants like Fe, Co, Ni lower the transition temperature [144], while it is experimentally demonstrated that Ge can bring the transition temperature up to 90° C [145, 146, 147]. For technological reasons, it would be advantageous to control the transition temperature and engineer it to higher values. In fact, envisioning commercial electronics applications, any VO\textsubscript{2}-based technology should guarantee its functionality up to temperatures higher then 80° C, which is the temperature commonly reached by a computer processor.

### 2.1.2 VO\textsubscript{2} Electrically-Triggered Transition

Already in 1959 it was shown that the insulator to metal transition of VO\textsubscript{2} can be electrically triggered in planar devices [104]. In these early works, it was possible to image with an optical microscope the portion of the film that was undergoing the IMT transition, therefore highlighting the filamentary character of the transition. The transition was attributed to the reaching of the temperature threshold by Joule heating in the device [148]. However, recent investigations of the IMT on scaled devices, mostly conducted through simulation efforts, pointed out that the Joule-heating effect could be insufficient to justify the transition and consequently opened a debate on the concurrence of non-thermal effects in the phase change. In this section the origin of the electrically-driven transition is briefly discussed, with a focus on the hypothesis regarding the nucleation of the phase transition and the evolution of the metallic filament in electrical devices.

A typical current versus voltage (I-V) curve of a vanadium dioxide device is shown in figure 2.5. In the first part of the curve, the VO\textsubscript{2} device is in its insulating state and acts as as a high impedance resistance. The power dissipated in form of Joule heat increases the local temperature of the device. When hitting a threshold voltage value $V_{TH}$, the device undergoes the insulator to metal transition. During this process, the voltage measured on the device decreases even though the current increases, which is reported as a feature of negative differential resistances [149]. After the transition, the device stabilizes to a low resistance value. Upon lowering of the voltage applied to the device, the voltage threshold $V_{TL}$ for the metal to insulator phase transition is reached. The device therefore switches back to its high impedance state.

As the phase transition in vanadium dioxide happens as a function of temperature, it would be natural to assume that Joule heating is responsible for the transition to the metallic state in an electrically activated device. However, the nature of the electrically driven transition, similarly to the nature of the transition itself, is greatly disputed in literature. In fact, many studies suggest that the Joule heating generated in the devices is insufficient to rise the temperature of the device above the phase transition. In addition, it has been shown that the phase transition
2.1. Vanadium Dioxide

Figure 2.5 | The current vs. voltage characteristic of a VO$_2$ device presents three regions: at first, the device is in its insulating state. As the voltage drop across the device is high enough to trigger the phase transition, the material undergoes the phase change and a negative differential resistance regime is formed. Lastly, the device stabilizes in its metallic state. The measurement was conducted sourcing a current and measuring the voltage across a VO$_2$ device.

can be triggered by carrier injection with the use of electrolyte gating [124] and it is predicted to occur upon the application of high electric fields, on the order of 1-10 MV/cm$^2$ [150], which might be reached in very scaled devices.

Multiple experiments conducted with a DC bias or with low frequency input signals, combined with measurements of the temperature of the device, sustain the theory that the electrically-driven phase transition is triggered by Joule heating. To cite just a few works, Mun et al. [151] showed an agreement between the Joule heating models of the device and measured voltage thresholds when sweeping the ambient temperature of the measurement setup. Similarly, Radu et al. [152] studied the threshold voltage dependence on the device dimensions, specifically the electrode separation, and on the power needed to trigger the IMT in the device. They find that when the stage temperature is increased, the power needed to switch the device diminishes, which is an indication of a temperature-triggered transition. In addition, the time delay for the metal to insulator transition (MIT) when the external voltage stimulus is removed is in agreement with the time needed to dissipate the heat of the device. Lee et al. [153] performed the same analysis utilizing a low frequency input signal, integrating also the impact of load resistor, temperature stage and frequency on the threshold voltage value, and found the experimental results in agreement with a numerical Joule heating model. In other works, the local temperature of VO$_2$ devices is measured with optical techniques during the electrical activation of the phase transition. In particular, Zimmers et al. in [154] integrated fluorescent particles on the surface of lateral devices and used them to monitor the temperature of the material. They observed that in all the cases the temperature of the device at the IMT corresponded to the temperature $T_c$ expected at the phase transition ($T_c = 68^\circ$ C).
However, other evidences support the hypothesis of a non-thermally induced transition. In particular, in contrast to what stated in [151, 152, 153], many simulation works show that temperature rise due to Joule heating of the device is not enough to reach the transition temperature. In particular, in [155], Gopalakrishan et al. suggest that the phase transition happens at a temperature lower than the phase transition temperature; however, upon switching to the metallic state, the increase in current would develop enough Joule heat to increase the temperature above $T_C$, therefore explaining why temperature measurement techniques have found in plural occasions the local temperature of the device higher than $T_C$. In addition, experimental evidences suggest that the transition in nanoscale device can arise from electrical activation. Fast voltage or light pulse measurements were used to monitor the delay time for the transition, and found that this time was too short to justify a thermal process [156, 157].

To complicate the picture, is the fact that the phase transition doesn't occur uniformly in the device but rather proceeds from a filament formation [158]. Li et al. have argued that the discrepancy between the results supporting the Joule heating and the field induced transitions can be overcome when including filament formation models in simulations, showing that the temperature of the filament can be as high as the $T_C$ [159]. Moreover, Aetukuri reported that the fast transition times registered for VO$_2$ would also agree with a Joule heating driven transition when the dependence of the insulating resistance with temperature is taken into account [114]. All these new studies would point to the conclusion that the phase transition in VO$_2$ scaled devices is temperature driven. However, a more recent, experimental study from Kalcheim et al., suggests that defects in VO$_2$ can also play an important role for the phase transition [122]. The authors fabricated high quality VO$_2$ nanowires with scaled widths to obtain single-domain crystals. With careful temperature calibration, they studied the Joule-heating induced transition in the nanowire and concluded that for high quality material the transition results to be temperature driven. In addition, they systematically created local defects damaging the wires with an ion beam and showed experimental proof that this time the IMT was occurring before the device could reach $T_C$. They explained the change in behavior of the nanowires with the Poole-Frenkel effect, in which an applied electric field reduces the energy barrier for excitation of carriers trapped in defects. The excited carriers can occupy the upper Hubbard band, therefore destabilizing the insulating phase and provoking the IMT similarly to a carrier injection technique. The mechanism responsible for the phase transition of electrically-driven devices remains therefore unclear, specially for devices fabricated on lattice-mismatched substrates, where the reduced quality of the material and the higher number of defects can play an important role in defining the IMT.

### 2.2 VO$_2$ Relaxation Oscillators

Vanadium dioxide can be used to build compact relaxation oscillators, exploiting its volatile phase transition. The first oscillations produced by VO$_2$ were already observed in 1975 by Taketa et al. [160]. However, it was not until 2008 that this phenomenon was systematically investigated [161, 162]. VO$_2$ two-terminal devices are fabricated and connected in series with
2.2. VO₂ Relaxation Oscillators

Figure 2.6 | a) Schematic of the circuit implementation of a VO₂-based oscillator. b) The series resistance biases the VO₂ device in the negative differential resistance regime. As the VO₂ device cannot stabilize either in the metallic or in the insulating states, the phase transition is self-sustained and relaxation oscillations form at the voltage output. c) and d) Voltage and current waveforms of the VO₂ device.

a standard resistor to generate the oscillating circuit. No inductance or transistors connected in a positive feedback loop are required. The circuit configuration used to build an oscillator is shown in figure 2.6. The oscillations are present in the system when the VO₂ is biased through the series resistance $R_s$ in its negative differential resistance regime (figure 2.6 (b)). With this biasing conditions, the VO₂ resistor cannot settle either in the metallic or the insulating state, resulting in relaxation oscillation at the output voltage $V_{out}$. The output waveform of such a system is a periodic, non-linear oscillation that can be classified as a relaxation oscillator, and it is shown in figure 2.6 (c) and (d).

A mathematical analysis of the output waveform can be conducted and the oscillation conditions can be found knowing the voltage thresholds $V_{TL}$ and $V_{TH}$ [163, 164, 161]. At the beginning, the device is in its insulating condition and the output voltage $V_{out}$ charges with the exponential dependence:

$$V_{out} = V_{ss} + (V_{st} - V_{ss})e^{-\frac{t}{\tau}},$$

(2.2)
Chapter 2. VO₂ Oscillators: from the Material to the Applications

where \( V_{ss} \) and \( V_{st} \) are respectively the steady state condition and the starting value of \( V_{osc} \) and \( \tau \) is the time constant of the rising exponential. In particular, defining \( R_{ins} \) the insulating resistance of the VO₂ device, upon the first rising edge we can calculate:

\[
V_{ss|ins} = \frac{R_{ins}}{R_{ins} + R_s} V_{in} \tag{2.3}
\]

\[
V_{st} = 0 \tag{2.4}
\]

\[
\tau_{ins} = \frac{R_{ins} R_s}{R_{ins} + R_s} C \tag{2.5}
\]

The condition for the device to undergo the IMT is:

\[
V_{ss|ins} > V_{TH} \tag{2.6}
\]

When reaching the voltage \( V_{TH} \), the device undergoes the IMT and becomes metallic, and the voltage \( V_{osc} \) starts to discharge following equation 2.2 with:

\[
V_{ss|met} = \frac{R_{met}}{R_{met} + R_s} V_{in} \tag{2.7}
\]

\[
V_{st} = V_{TH} \tag{2.8}
\]

\[
\tau_{met} = \frac{R_{met} R_s}{R_{met} + R_s} C, \tag{2.9}
\]

where \( R_{met} \) is the valued of the resistance fo the VO₂ device in its metallic state. For the circuit to oscillate, the device needs to undergo the MIT, therefore it needs to reach the voltage \( V_{TL} \):

\[
V_{ss|met} < V_{TL} \tag{2.10}
\]

Putting together equation 2.6 and 2.10, we can conclude that the conditions on \( V_{IN} \) and \( R_s \) for obtaining oscillations in the circuit are:

\[
\frac{R_{ins} + R_s}{R_{ins}} V_{TH} < V_{in} < \frac{R_{met} + R_s}{R_{met}} V_{TL} \tag{2.11}
\]

\[
\frac{V_{in} - V_{TL}}{V_{TL}} R_{met} < R_s < \frac{V_{in} - V_{TH}}{V_{TH}} R_{ins} \tag{2.12}
\]

From equation 2.2, solving for the rising and falling edges of the oscillations and for the limits
2.2. VO$_2$ Relaxation Oscillators

$V_{TL}$ and $V_{TH}$, the oscillation frequency can be calculated:

$$\frac{1}{f} = t_{rise} + t_{fall} = \tau_{ins} \ln\left(\frac{R_{ins}}{R_{ins}+R_s} V_{IN} - V_{TL}\right) + \tau_{met} \ln\left(\frac{R_{met}}{R_{met}+R_s} V_{IN} - V_{TH}\right)$$  \hspace{1cm} (2.13)

The equations so far presented show the dependency of the frequency of the VO$_2$ oscillators from the electrical parameters. The oscillators normally report a maximum operating frequency in the order of 100 kHz [165], with a few exception that reported frequencies in the order of MHz [166]. The maximum frequency obtained by the oscillator depends on the various circuit parameters. The simplest way for achieving high frequency oscillations is a reduction of the capacitance $C$, that in the experimental realizations is mostly represented by parasitics in the connections with the circuit elements and in the set-up used to measure the devices. However, also the device parameters, and in particular the hysteresis width $(V_{TH} - V_{TL})$ influence the oscillation frequency. From the equations presented above it is clear that a narrower hysteresis is desirable for faster oscillations. However, not only the electrical parameters, but also the thermal parameters influence the device frequency. In particular, for very small electric constants, the thermal constant can become a decisive limiting factor of the oscillation frequency [159, 167]. In addition, Driscoll in [168] also discusses the role of domains in the device in influencing the oscillating behavior.

2.2.1 Coupled Oscillators

Frequency locking of coupled oscillators is a well known phenomenon, comprising effects like frequency injection locking, frequency pulling and frequency and phase synchronizations, widely used for application in synthesizers, transmitters and receivers [169, 170]. Oscillators that present a weak coupling can interfere with each other and eventually lock in frequency and phase. In [163] Shukla et al. have demonstrated the frequency locking of two VO$_2$ oscillators using a capacitor as a coupling element. Figure 2.7 (b) shows the different, natural frequencies of two uncoupled VO$_2$ oscillators. When the oscillators are connected with a capacitive element, the two oscillators lock at a frequency lower respect to their natural ones. The locking to a lower frequency can be easily explained taking into account the additional capacitive element, i.e. the coupling capacitance, which increases the electrical time constant of the circuit. A model describing the coupling dynamics of the oscillators is presented in [83], and applied to different coupling schemes. In this work, it is simulated that purely capacitive coupling between two identical VO$_2$ oscillators leads to out-of-phase synchronization of the waveforms, while in-phase synchronization could be obtained with purely resistive coupling. An R-C coupling scheme would in addition allow for more complex phase relations. After this pioneering work, other implementation of VO$_2$ coupled oscillators were proposed. For example, Velichko et al. in [171] achieved frequency synchronization exploiting the thermal coupling between two closely-fabricated VO$_2$ devices.
Chapter 2. VO₂ Oscillators: from the Material to the Applications

Figure 2.7 | The figure shows the circuit schematic (a) and the frequency locking (b) of two coupled oscillators based on VO₂. The oscillators are coupled through a capacitive element. Even though the natural frequency of the oscillators is not the same, with the introduction of the coupling element, the oscillators lock in frequency. Reproduced with permission from [163], licensed under CC BY-NC-ND 3.0.

2.3 Oscillatory Neural Networks

Synchronization of large systems of oscillators is ubiquitous in nature and dominates many of our essential living processes. For example, pacemaker cells in the heart produce synchronized periodical electrical impulses that allow our hearts to beat; insulin cells in the pancreas release insulin in a synchronized, oscillating mechanism. Other examples include synchronized behavior of different organisms, such as the cricketing of crickets at night [172], and more impressively the synchronization of the light flickering in fireflies [173]. Coupled oscillators systems have been studied mostly in biology, specially exploring the synchronous behavior in the brain activity and its role in our cognitive process. It has been theorized that different neural groups in the brain can communicate only through synchronization of their rhythmic behavior [174, 175]. In particular, oscillatory fluctuations in the brain have been associated with our capability to memorize sequence of events, to recognize novelty, process sensorial signals and the retrieval of stored memories [65, 66, 176]. In this section we discuss the computation capabilities of oscillators. In particular we focus on the definition of associative memory in Hopfield Neural Networks and we describe in detail the Hoppensteadt and Izhikevich's model for coupled oscillators computation. Finally, we give a brief overview of the oscillators’ technologies and the computing capabilities that were demonstrated in literature.
2.3. Oscillatory Neural Networks

2.3.1 Hopfield Neural Network

The Hopfield network is one of the most studied neural networks. It is a form of recurrent neural network, meaning that the timing evolution of the network has an influence on the network output, or equivalently, that the network has a memory. The Hopfield network is organized by neurons that are fully-connected to each other through synaptic weights. In the Hopfield network, each neuron \( x_i \) can assume only two predefined states, namely \( x_i = +1 \) or \( x_i = -1 \) (other combinations are possible, such as \( x_i = +1 \) or \( x_i = 0 \)). The connection between a neuron \( i \) and a neuron \( j \) is represented by the weight \( w_{ij} \), that can be either positive or negative. We define as a state \( X \) of the network the collective values of the neurons in a certain point in time; the network state is progressively updated over time:

\[
X = [x_1, x_2, \ldots, x_N]^T.
\] (2.14)

At each point in time, the state, or output, of the neuron \( i \) is calculated as the sum of the inputs coming from the other neurons, processed by the weights \( w_{ij} \) and the biases \( b_j \), activated by a non-linear function \( f \):

\[
x_i = f(\sum_{j=1}^{N} w_{ij} x_j + b_j).
\] (2.15)

The non linear function can be represented by a sign function, that would automatically set the state value to \( x_i = \pm 1 \). As Hopfield himself describes in his paper [67], such a system, that derives from a multitude of simple components, presents spontaneous computing capabilities when its collective behavior is considered. In particular, given an ensemble of connections \( w_{ij} \), the system has only a certain number of states that are stable [82]. Defining as \( X_a, X_b, X_c \ldots \) the stable states in the system, if in a certain point in time the system state is \( X = X_a + \Delta \), it will relax in time to the state \( X_a \). The stable states of the system are called attractors, as they attract each unstable state in the network to converge to them. Alternatively, the attractors can be seen as local minima in the energy landscape of the system:

\[
E = -\frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} w_{ij} x_i x_j.
\] (2.16)

The approaching of a given state to the nearest attractor can be seen as a relaxation of the system to a lower energy configuration [177]. The system can therefore be seen as a general content addressable memory, also called associative memory. The memory of the circuit can be controlled to memorize specific patterns, by careful choice of the coupling weights, for
example with the Hebbian learning rule:

\[ w_{ij} = \frac{1}{N} \sum_{s=1}^{M} x_i^s x_j^s \]  

(2.17)

where \( s \) is the state to be memorized, \( x_i^s \) is the value of the neuron \( i \) in the state \( s \), and the sum is done over the total number \( M \) of states that need to be memorized. The Hopfield model also wants the connection \( w_{ii} = 0 \) and \( w_{ij} = w_{ji} \). It has to be noted that, as Hopfield explains in his paper, the state of “all equal neurons”, i.e. the state in which all the neurons have the same value, is always a stable state [67]. A network with \( N \) neurons has however just limited storage capabilities. In particular, it was found that the maximum number \( M \) of memorizable stable states scales linearly with \( N \) and is limited to \( M < aN \), with \( a = 0.14 \) [178]. When this limit is violated, spurious attractors are formed in the system, and the pattern retrieval is impaired.

By relaxing the condition \( w_{ij} = w_{ji} \), it is however possible to increase the maximum number of pattern to \( M = N \) [179]. The linear scaling of the memory storage’s limit is a problem for the practical implementation of such a network. In fact, as the network is fully connected (FC), the number of weights scales exponentially with the number of neurons. Therefore for going to higher capacity of the memory, more neurons are needed, but we encounter the problem of providing the connections for such large networks. More recent studies focused on how to increase the memory capacity of the Hopfield network. In particular, Folli et al. found that the storage limitation can be extended without loosing in retrieval accuracy for \( M >> N \), with imposing \( w_{ii} \neq 0 \) [180].

The Hopfield network has proven to be robust to deviations in the model, such as the definition of the neuron behavior or the implemented coupling scheme. Hopfield himself in his manuscript predicted that with adding details or implementing small parameters changes, the working principle of the network will not be modified. This robustness to deviation from the canonical model contributes to practical implementation of the neuron and synapses with a variety of electronic components representing the neurons and the synapses in the Hopfield network, as documented in the following session.

### 2.3.2 Models of Oscillatory Neural Networks

One variation of the Hopfield model was proposed in [69, 181], where the possibility of using more complex neuron models is explored. In particular, the Hopfield model was modified to introduce oscillatory states instead of the bistable states of the neurons; the information of the network would therefore be encoded in the relative phase of the oscillators, and the computing is performed by the complex synchronization dynamics of the network. The oscillatory neural network (ONN) retains the main properties of the Hopfield neural network: it is in fact an attractor network, meaning that only specific dynamics between the oscillators are stable (attractors), and any deviation from the stable state will naturally relax to one of the attractors. In the case of ONNs, the attractors are not represented as a single point in state values as in
The Hopfield network, but rather as the relative phase of the oscillating neurons. This means that the attractor is a periodic attractor, that can be represented as a periodic orbit, or limit cycle (figure 2.8). In this case, the analysis of the network is done utilizing a phase model, i.e. a mathematical model that describes the phase relations between the oscillators. In this section the phase model of an oscillatory neural network will be discussed following the formalism described in [182].

The phase model to describe the oscillator system is based on the observation that, once the oscillator relaxes to its limit cycle, it can be described by only one variable, its phase $\theta$ [183]:

$$\dot{\theta}(t) = \Omega,$$  \hspace{1cm} (2.18)

where $t$ is the time, $\Omega$ is the frequency of the motion on the attractor. As depicted in figure 2.8, every point on the attractor corresponds to a specific value of the variable $\theta$. Any oscillating configuration that does not start on the attractor, but will relax on the attractor, can be equivalently described by the phase model [184]. In addition, if a neuron $x$ evolves in time with a dynamic activity represented by a function $f$:

$$\dot{x}(t) = f(x),$$  \hspace{1cm} (2.19)

it has been demonstrated that if this dynamic is periodic, any solution $x(t)$ can be projected near an attractor $\theta(t)$ in the canonical phase model of equation 2.18. In other words, the phase model of equation 2.18 can describe the periodic activity of an oscillatory neuron expressed by any state equation like equation 2.19. Independently on the form of the oscillatory neuron,
the phase model can describe the periodic behavior of its attractors and of the unstable states tending to the attractors. Similarly, in a network of coupled oscillatory neurons, the dynamics of the network can be described with the phase model of the collective periodic activity of the neuron, independently from the specific neural model used to describe the neuron activity [182]. The phase equation of such a system has the form of:

\[ \dot{\theta}_i = \Omega_i + \varepsilon \sum_{j=1}^{N} h_{ij}(\theta_i, \theta_j, \varepsilon), \quad (2.20) \]

where \( \varepsilon \) is a term that represents the collective strength of the coupling in the network, \( N \) is the number of the oscillators present in the network and \( h_{ij} \) is a function that represents the coupling between an oscillator \( i \) and an oscillator \( j \). This phase model proceeds from Kuramoto model on a network of coupled oscillators and it models phenomena as frequency and phase locking of the oscillators. If the natural frequency of the oscillators is similar, we can describe the frequency of each oscillator as a deviation \( \omega \) from the mean frequency \( \Omega_0 \) of the oscillators:

\[ \Omega_i = \Omega_0 + \varepsilon \omega_i. \quad (2.21) \]

Subsequently, the phase of the oscillators can be represented as:

\[ \theta_i = \Omega_0 t + \varphi_i. \quad (2.22) \]

Substituting equation 2.22 in equation 2.20, we obtain:

\[ \frac{d\varphi_i}{d(\varepsilon t)} = \omega_i t + \sum_{j=1}^{N} H_{ij}(\varphi_i - \varphi_j) + o(\varepsilon). \quad (2.23) \]

The term \( o(\varepsilon) \) collects the higher order terms of \( \varepsilon \), and can be discarded in the approximation of weakly coupled oscillators. Under the assumption of symmetry of a pairwise odd form of the term \( H_{ij} \), the energy landscape of the system can be described by the formula:

\[ E = \frac{1}{2} \sum_{j=1}^{N} \sum_{i=1}^{N} R_{ij}(\varphi_i - \varphi_j), \quad (2.24) \]

where \( \frac{dR_{ij}}{d\theta} = H_{ij} \). Similarly to what happened in the Hopfield model, the energy function of the oscillatory neural network presents local minima that act as attractors for the dynamic system. This assures the stabilization of all the frequency-locked oscillators to determined phase-relations, that represent the associative memory of the system. This memory can be
2.3. Oscillatory Neural Networks

![Memorized Patterns](image)

**Figure 2.9** Example of image recognition with a system of coupled oscillators. In this figure, each white (black) pixel is an oscillator in phase (out-of-phase) with a reference. Three patterns are memorized in the oscillators’ net. When a search pattern is presented to the network, the network relaxes to the nearest attractor, therefore recognizing the digit “1” [70].

controlled with the Hebbian rule as in equation 2.17. It has to be noted that this model does not make any assumption on the nature of the oscillators, therefore any type of weakly coupled oscillator can be used to build such a system. As Izhikevich mentions in [182], “everything that can oscillate, can also compute”.

To give a schematic representation on how computing with oscillators works in practice, we refer to figure 2.9. In the digit images presented in this figure, each pixel is an oscillator. All the oscillators are locked in frequency and the information is carried in their relative phase relation. White pixels are oscillators that oscillate in phase with a reference, black pixels are oscillators that oscillate in out-of-phase with a reference. The connections between the oscillators are programmed so that the memory of the network stores the digits “1”, “2”, “3”. To recognize a distorted pattern (also referred as test pattern), the oscillators are initialized to a phase condition which encodes the value of the pixels of the search pattern. As we will discuss later, in various experimental implementation of such networks this initialization is carried out with different techniques. After the initialization, the oscillators are left to relax to the nearest collective attractor, therefore recognizing the correct digit.

As in the Hopfield model, the oscillatory neural networks are limited in their storing capabilities to a number of patterns that scales linearly with the number of oscillators, or neurons, that are present in the network. Therefore, to store a high number of patterns, large networks are required. The network is fully-connected, meaning that each oscillator \( i \) is connected to each oscillator \( j \). This becomes an important limitation in practical implementations, as it is difficult to realize such heavily interconnected networks in large sizes. A different design, that tries to overcome these limitations, has been proposed by Izhikevich and Hoppensteadt in 1998 and has featured since then many experimental implementations [70]. In this alternative
approach, instead of using separate connections for coupling all the oscillators, the oscillators are connected to a single node, that virtually provides all the coupling through an externally injected signal (figure 2.10). This configuration takes the name of Frequency Modulated (FM) oscillatory neural network [185].

The idea of this network is to implement the coupling of the oscillators with a input time-dependant signal $a(t)$, rather then with physical connections. This has been demonstrated with a mathematical model that envisions sinusoidal oscillators. In this case the phase shift dynamics of equation 2.20 reduce to:

$$\dot{\vartheta}_i = \Omega_i + \varepsilon a(t) \sum_{j=1}^{n} \sin(\vartheta_i - \vartheta_j).$$ (2.25)

By choosing a coupling matrix $w_{ij}$ and a periodic input signal $a(t)$ of the form:

$$a(t) = a_0 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \cos((\Omega_i - \Omega_j) t),$$ (2.26)

and making similar assumptions on $\Omega$ and $\varepsilon$ as for the fully connected network, the phase shift dynamics reduce once again to:

$$\frac{d\phi_i}{d(\varepsilon t)} = \sum_{j=1}^{n} \frac{w_{ij} + w_{ji}}{2} \sin(\phi_i - \phi_j).$$ (2.27)

The associative memory properties of the network are retained. The coupling coefficients $w_{ij}$
2.3. Oscillatory Neural Networks

can be calculated again with the Hebbian learning rule (equation 2.17), therefore encoding the patterns to memorize in the network. The retrieval process is however in this case more complicated. In fact, \( a(t) \) is at the same time the signal used to input the search pattern and the memory of the system. Therefore a pattern search needs to proceed in two steps:

1. Defining a search pattern \( \xi = [\xi_i, \xi_j, ... \xi_N] \), the black pixels have \( \xi = -1 \) and the white pixels a value \( \xi = +1 \). The input signal \( a(t) \) is initialized by choosing \( w_{ij} = \xi_i \xi_j \). In this way all the oscillators corresponding to a black pixel in the search pattern will stabilize to a phase value \( \varphi_- \), while the ones corresponding to a white pixel will stabilize to \( \varphi_+ \). The oscillatory network therefore oscillates stably to phase differences corresponding to the search pattern.

2. At this point the recognition process begins. The search pattern should be recognized as one of the memorized patterns. Therefore the input signal \( a(t) \) needs to be changed to provide the memory of the system: \( a(t) \) is therefore altered according to the Hebbian learning rule corresponding to the memorized patterns. At this point the oscillator network evolves relaxing to the newly encoded attractors of the system, and therefore finalizing the recognition process.

Even though this second technique allows for realization of large nets with limited interconnections, the practical implementation in compact electronic circuits is hindered by the complex waveform generator needed to synthesize the signal \( a(t) \). In the following section, an overview of the oscillatory neural networks implemented with these two techniques, FC and FM, will be given. However, for the scope of this thesis only the FC implementation will be taken in consideration. As explained in Chapter 5 and 6, the disadvantage of the high interconnectivity of the FC-ONN can be overcome by utilizing the ONN in combination with other neural network concepts, like convolutional neural networks, that require very small matrices of ONN to compute.

To summarize, an oscillatory neural network is a type of network that computes with the phase of periodic, dynamical systems. The system relies on the following properties to perform computation [84]:

1. The oscillatory network system has associative memory properties.
2. Under the assumption of weak coupling between the oscillatory neurons, the phase dynamics model can be linearised and expressed with a set of linear equations. The model is independent from the actual waveform of the oscillator.
3. The information of the system is encoded in the relative phase of the oscillators and/or in the frequency coupling of the system.
4. The periodicity of the system makes it possible to read the phase difference over multiple cycles, which allows to suppress jitter and noise.
Chapter 2. VO₂ Oscillators: from the Material to the Applications

2.4 Technologies that Compute with Oscillators: an Overview

In this section we review some experimental demonstrations of oscillatory neural networks, that use different materials and devices to build the oscillators as well as different system architectures to implement the computation. We will also talk briefly about different computation schemes realized with oscillators, that do not exploit a neural network concept, but are still important in their signal processing functionalities to understand the ideas of this work.

2.4.1 Materials and Devices for Oscillatory Neural Networks

Many types of oscillators have been proposed as a building block for ONN. The most popular are electrical oscillators, i.e. oscillators that rely on the exchange of charges. Given that an electrical signal originates the periodic behavior, the coupling elements are usually implemented with simple circuit components.

The most straightforward way to realize an electrical oscillator is to build an LC oscillator. However, as inductors are notoriously difficult to scale, the practical implementation of ONNs with LC components cannot easily meet the industry standards for a competitive technology. One of the most successful experimental demonstrations of an ONN in hardware was however realized with 8 LC-oscillators coupled in the FM configuration in [186]. Other theoretical works are based on LC oscillators as their sinusoidal voltage oscillation allows for a simpler mathematical model. For example, in [187], a simulation work comprising LC oscillators coupled with a differential amplifying stage is suggested. Alongside analog electrical oscillators, also digital oscillators, such as ring oscillators (RO) have been proposed as building blocks for ONNs [188, 189]. As an ONN is an analog circuit, Hopfield networks realized with RO also behave as analog circuits, despite the digital nature of the individual components. The power consumption of these devices is quite limited, with the best performing ones achieving a dissipation of only 24 nW @ 5.24 MHz [190].

Metal-insulator transition materials, such as VO₂, have been explored for ONN technology given their compact structure and the low power consumption of the devices. In [78] it is envisioned that VO₂ oscillators, when scaled to 20 nm dimensions can consume 0.5 µW @ 1.6 GHz, for a total energy consumption per cycle of \( E \approx 10^{-16} \) J. However, these numbers have not yet been confirmed by experimental results. The best performing devices so far operate at a maximum frequency of 1 MHz [163] and a minimum power consumption of 8 µW @ 10 Hz [191]. Other phase change materials have been explored to build compact oscillators. For example, in [192], electrical relaxation oscillators are built with amorphous GST (GeSbTe), but the progressive decrease of amorphous volume in the chalcogenide layer causes a damping behavior in the oscillations up to their complete disappearance. This opens the discussion to another important figure of merit for phase-change material based relaxation oscillators: the endurance. Implementation of relaxation oscillators based on other transition metal oxides materials, such as niobium oxide [77, 193] tantalum oxide [194, 195, 79], have been more successful in demonstrating MHz frequency operation and endurance respectively up to 10¹⁰
2.4. Technologies that Compute with Oscillators: an Overview

and $10^6$ cycles. Vanadium oxide oscillators have demonstrated endurance without fatigue up to $10^9$ cycles [196].

A large class of oscillators relies on the spin-procession in ferromagnetic materials, like spin torque oscillators (STOs), which can perform high frequency with limited power consumption, around $10^{-15}$ J/cycle [80, 81, 82]. In STOs the oscillations are given by the procession of the magnetic moment in a magnetic thin film and are excited by a spin-polarized current or by the spin Hall effect. The magnetic moment of the magnetic film precesses and alongside this precession the resistance of the device changes. The precession frequency can be locked and controlled by an ac driving of the input current. The main challenge faced in this technology is found in the design of a simple and efficient interconnection scheme, as it either requires the magnetic signal to be converted in the electrical domain and back, or it needs to rely on other techniques as dipolar or spin wave coupling [197]. The design of an electrical coupling was presented in [197], where the STOs signal is picked up by an RC filter, processed through a summing node and then reconverted in magnetic signal by magnetic-field coupling to an electrical wire. However, upon the magnetic-to-electrical conversion, the amplitude of the voltage signal is quite small, usually on the mV range. Demonstrated coupling schemes also comprehend interaction of closely-spaced devices [198], even though this does not allow for the reconfigurability of the system.

Nano electromechanical systems (NEMS) and micro electromechanical systems (MEMS) have also been proposed for coupled oscillators’ technology [199, 200, 201, 202], as well as phase locked loops implementations [203]. In table 2.11 we summarize the types of oscillators discussed and their figures of merit. The motivation behind the choice of this work to use VO$_2$ to build electrical oscillations is given by its competitive performances in terms of power consumption, endurance and simple electronic coupling schemes.

<table>
<thead>
<tr>
<th>Oscillator technology</th>
<th>Area (µm)</th>
<th>Frequency</th>
<th>Energy (J) /cycle</th>
<th>Coupling mechanism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC oscillator</td>
<td>Up to 100 GHz</td>
<td>$10^{-17}$</td>
<td>Electrical</td>
<td>[82]</td>
<td></td>
</tr>
<tr>
<td>Ring oscillator</td>
<td>Up to 20 GHz</td>
<td>$10^{-12}$</td>
<td>Electrical</td>
<td>[188]</td>
<td></td>
</tr>
<tr>
<td>TaO$_x$</td>
<td>5</td>
<td>250 MHz</td>
<td>$10^{-12}$</td>
<td>Electrical</td>
<td>[194]</td>
</tr>
<tr>
<td>NbO$_x$</td>
<td>150</td>
<td>20 MHz</td>
<td>$10^{-12}$</td>
<td>Electrical</td>
<td>[193, 77]</td>
</tr>
<tr>
<td>VO$_2$</td>
<td>100</td>
<td>10 MHz</td>
<td>$10^{-17}$</td>
<td>Electrical</td>
<td>[78]</td>
</tr>
<tr>
<td>STO</td>
<td>375</td>
<td>50 GHz</td>
<td>$10^{-15}$</td>
<td>Electrical and magnetic</td>
<td>[81, 82]</td>
</tr>
<tr>
<td>MEMS/NEMS</td>
<td>200/20 µm</td>
<td>20 GHz</td>
<td>$10^{-14}$</td>
<td>Electrical and mechanical</td>
<td>[82]</td>
</tr>
<tr>
<td>PLLs</td>
<td>275</td>
<td>250 MHz</td>
<td>$10^{-12}$</td>
<td>Electrical</td>
<td>[203]</td>
</tr>
</tbody>
</table>

Table 2.11 | Comparison with different oscillators technology. Many novel devices allow to realize scaled, power efficient oscillators. VO$_2$ oscillators are between the best performing devices. Adapted from [82] with the permission of AIP publishing.
2.4.2 Image Processing with Oscillators

Image processing can be computing intensive, specially when a large number of convolution and filtering actions need to be performed on high volumes of data, as for example happens with image classification tasks.

Many simulation works, as well as experimental approaches, have explored the capabilities of ONNs to retrieve information from distorted patterns. As an example, as already shown in figure 2.9, an ONN can be trained to recognize handwritten digits. When an input, distorted handwritten image is fed into the network, the network will relax to the nearest pattern in its memory, therefore recognizing the image. This recognition process doesn't need training over a high number of images of the digits to perform the recognition: the training is in fact a one step process implemented with the weight tuning according to the Hebbian learning rule, as explained in section 2.3.2. Many demonstrations of this concept have been provided with different oscillators technologies. For example, Maffezzoni et al. in [187] simulate handwritten digit recognition with a 60-neurons ONN based on LC-coupled oscillators. In this work, a FC network is initialized to the right phase with setting the circuit weights for storing the search pattern attractor, as more commonly done in the FM configuration. Jackson et al. in [79] have simulated digit recognition in a 20-neurons FC ONN with tantalum oxide oscillators; the initialization of the network to the search pattern is done by starting the oscillations at the phase corresponding to the search pattern through different frequency multipliers driven by a global clock. Holzel et al., in [186], present instead a experimental coupling of 8 Van der Pol oscillators in the FM configuration and initialization scheme. One of the most notable works was instead presented by Jackson et al. in [203, 204, 205] with 28-nm CMOS technology and phase locked loops. The authors here discuss an experimental design of a whole ONN systems based on phase-locked-loops oscillators, counting 100 neurons connected in a FC configuration with programmable synapses, representing the biggest experimental demonstration of an ONN.

Although these implementations are very promising and illustrate the pattern retrieval capabilities of the ONNs, they come with certain challenges which makes it difficult for them to compete with already established computer technologies in practical applications. As mentioned before, the storing capability of Hopfield networks is greatly limited to the number of neural nodes. Increasing the memory of the system would imply increasing the number of neurons, that would make the connectivity of FC networks too cumbersome for practical implementation. The networks also suffer from arising of spurious patterns that can spoil recognition.

Alternative demonstrations of pattern retrieval systems are described for STO technology by Nikonov et al. [206] using a quasi-synchronization concept called degree of match (DOM), which was first introduced in [207]. In this case, instead of a normal phase-encoding of the information, the authors use the frequency as a state variable of an FM system. They control the frequency of the STOs from the input current, and they program each neuron at a frequency
deviation from the central frequency depending on the difference in value between the pixel of the stored pattern and the pixel of the search pattern. If the collective frequency deviation between search and stored pattern is high, the ONN will fail to synchronize. Viceversa, when the search pattern is close enough to one of the stored pattern, synchronization will be achieved. The deviation of the synchronization frequency from the natural frequency of the oscillator gives a probability of right recognition for this pattern. This frequency-encoding technique is also used in many other works [208, 209, 210, 211], and it solves the problem of the low capacity of the Hopfield network associative memory, however introducing the disadvantage of having to compare the search pattern with each of the stored patterns.

Other computing schemes have been investigated harvesting the dynamic connectivity of coupled oscillators. The idea is to focus on different image processing tasks, and not specifically on pattern retrieval. Shukla et al. in [196] demonstrated the possibility of using two-coupled oscillators for high-saliency problems. Two neighbouring pixels are confronted via two capacitively coupled VO$_2$ oscillators. The pixel value is encoded in the gate voltage of a series transistor used to drive the oscillators, as depicted in figure 2.12 (a). The output phase (in-phase/out-of-phase), gives a measure of the similarity of the pixels. When each pixel is confronted with the 8 neighbouring ones, the edges of the images can be highlighted. Other works present the same results exploiting the DOM technique [212, 213, 214, 215]. Moreover, Cotter et al. in [216] perform edge detection and high-saliency detection with a simulation effort of a 9-coupled oscillators system where the relevant information is encoded in the time needed for the oscillators to achieve phase synchronization. The output is then converted to a black or white value of the central pixel of the sliding window (figure 2.12 (b)). Similar operations are also shown by Tsai et al. in [217] with VO$_2$ coupled oscillator. In this case edge detection, erosion, dilation and color detection are simulated for a 9-coupled oscillators system connected through a coupling capacitance to an output node (figure 2.12 (c)). The input of the image is given as a voltage on a transistor controlling each oscillator; a second voltage-controlled gate is used to set the function to be recognized, respectively in terms of horizontal, vertical, diagonal edge gradient, dilation, erosion, or for values corresponding to the detection of different colours. The information calculated from the 9-pixel filter processing operation is contained in the voltage at the output node of the system. The output voltage is a time-varying signal, whose shape depends on the synchronization state of the 9-coupled oscillators. First, if the input voltage values of the oscillators (corresponding to the different pixels of the image) are close in value, synchronization will be achieved and the output voltage signal will be a periodic voltage with a fixed peak-to-peak amplitude which will reflect the amount of deviation between the oscillators; when the input signals are not close together, synchronization is broken and the output voltage will have a drifting amplitude value; if the input voltages are out of certain boundaries, the system will cease to oscillate. A read out scheme comprising the value of the output voltage amplitude as well as its dynamics (regularly oscillating, time-varying or not oscillating) can convert the information corresponding to the filtering action performed. The filtering operations are not performed in parallel, but for each operation the system needs to be reconfigured, similarly to what happens in a DOM system. For directional edge detection,
it is mentioned that the degree of similarity needs to be calculated from the relative deviation of 3 neighboring pixels oriented in the same direction. A summary of the three techniques here presented in table 2.13.

Figure 2.12 | Proposed schemes of edge detection and high saliency detection with coupled oscillators circuits. a) The system described in [196] is able to recognize the edge of an image comparing each pixel of the image to its 8 neighboring pixels. The pixel color difference is encoded in the input gate voltages of the oscillators. The output is read on the coupling capacitance and converted in a string of 0 and 1 with a XOR gate. The average represents the relative matching of the two pixels. Reproduced with permission from [196], © 2014 IEEE. b) A DOM configuration is simulated for different oscillator technologies in [216]. For the edge detection a window of 3x3 oscillators receive as an input a frequency shift proportional to the difference between the input pixels and the edge to be recognized. The time to synchronization is an evaluation of the degree of matching. A comparison needs to be done for each edge to be recognized. DOM schematic adapted with permission from [206], © 2015 IEEE; image processing reproduced with permission from [216], © 2014 IEEE. c) The complex system of VO_2 coupled oscillators described in [217] is able to perform a wide range of function calculating the amplitude of the output voltage V_{out}. The function can be selected tuning the input gate voltage of the transistor V_y and a variable series resistor. Between the various computations, the detection of edges is here depicted. Reproduced with permission from [217], © 2016 IEEE.

As a conclusive remark, the aim of this work is to demonstrate image classification using ONNs. A Hopfield network working with an associative memory is not able to perform a classification task. In fact, in a Hopfield network an n \times m pixel image can be stored in an n \times m network of fully-coupled oscillators. A distorted pattern of the same n \times m dimension can be recognized by the network, with the output of the network being an n \times m image in which the information is stored in the relative phase of the oscillators, that converge to one of the stored
2.5. Convolutional Neural Networks

A convolutional neural network (CNN) is a well-known deep learning algorithm architecture that is inspired by our visual perception mechanism. A first discussion of a neural network model for visual perception can be found in [218], were the authors present "neocognitron", a network consisting in multiple layers of cascaded cells, connected together by trainable weights. In 1997, LeCun et al. laid the basis of the modern framework of CNNs, designing a neural network that will become known as LeNet [219]. As other neural networks, LeNet was composed by multiple layers, connected by weights that could be trained by a backpropagation algorithm. It was able to perform recognition of handwritten digits with good precision. In this section we briefly introduce the basics of convolutional neural networks, following the formalism described by Gu et al. in [220].
In literature many variants of convolutional neural network architectures can be found, however the building blocks are the same for every implementation. Convolutional neural networks are usually built with a combination of three different layers: convolutional, pooling and fully connected layers. The convolutional layers represent the first layers of the network and perform a series of convolutions on an input image to learn the characteristic features of locally correlated data points. Each layer is composed by numerous kernels (or filters) that select and compute different feature maps. We can see a kernel as a set of weights, which connect neighbouring pixels (or input neurons) of an image and perform a computation whose result, activated by a non linear function, will be the value of one neuron of the following layer. The entire feature map of an image is therefore calculated with a convolution operation of the input image, in which the input values are multiplied element-wise with the kernel weights. The kernel slides on subsequent subsets of pixels of the input image, like a sliding filter, and computes the output. Each convolutional layer in CNNs comprehends several kernels, which calculate multiple feature maps of the input image. Mathematically, this operation is represented by the equation:

$$z_{l,i,j,k} = w_k^T x_{l-1,i,j} + b_k$$

(2.28)

where $z$ is the neuron at layer $l$ in position $i, j$, $w$ is the weight vector and $b$ is the bias vector representing kernel $k$ of the $l$-th layer, $x_{i,j}$ is the input patch centered in $i, j$ of the $l-1$ layer. A single kernel is applied element-wise to the entire input image, with a sliding stride that can be varied. The output $z$ of a kernel is then activated by a function that introduces the non-linearity in the network, necessary for having a successful network training:

$$a_{l,i,j,k} = a(z_{l,i,j,k})$$

(2.29)

Typical activation functions are sigmoid, $tanh$ and ReLU [221, 222]. The size of a convolved layer, or feature maps, is determined by:

1. the size $m$ of the kernel used;
2. the depth of the layer, i.e. the number of kernels $K$ used in the convolution operation;
3. the stride $s$ of the convolution operation, meaning the number of pixels by which the kernel is shifted over the input vector. For example, when the stride is equal to 1, then the kernel is moved on the input image one pixel at a time.
4. the padding $p$ of the input image. Many times a contour of zeros is added to the input vector of the convolution filter, in order to convolve also the border of the filter with the kernel.

The size of the output convolution is a layer of dimensions $(W, H, K)$, where $K$ is the number
of feature maps and $W$ and $H$ can be calculated as follow:

$$W, H_l = \left\lfloor \frac{W_{l-1} + 2p - m}{s} \right\rfloor + 1$$  \hspace{1cm} (2.30)$$

Following a convolutional layer it is usually a pooling layer, or subsampling layer. The aim of this layer is to reduce the dimensionality of the feature map while retaining the important information. For example, we could define that a pooling layer would take a spatial neighbourhood, for example of $2 \times 2$ pixels, and select the largest element in this neighbourhood as the only information passed by to the next layer. This operation is called Max Pooling. Other types of poolings are commonly used, as average pooling, in which the selected information is the average of the pixels in the neighbourhood, sum pooling etc [223, 224]. After more convolutional layer, the last layers of the network are fully-connected layers that result in the classification of the input image. Fully-connected layers take all the neurons in the previous layer and connect them each to all neurons of the following layer. Usually, to go from a convolution layer of $(W, H, K)$ dimensions to a fully-connected layer, a flattening operation is performed to transform the output of the convolutional layer in a one dimensional vector of $(1, W \times H \times K)$ dimension. The last layer of the network is the classification layer. Every neuron of this layer represents a class, and the value of the neuron represents the probability for the input image to belong to this class. Usually the last layer goes through a softmax function.

Convolutional neural networks are usually trained with a backpropagation algorithm and a form of supervised learning. The aim of the learning is to find the optimum ensemble of parameter $\gamma$, where $\gamma$ includes all the weights and biases of each layer of the network, for which a certain loss function is minimized. The loss function is usually defined as a distance of the output of the last layer of the CNN to the label that was originally assigned to the input. If $x_n$ is the input of the network, $y_n$ the label assigned to this input, and $o_n$ the output of the network, the loss function can be calculated as follows:

$$L = \frac{1}{N} \sum_{1}^{N} l(\gamma, y_n, o_n)$$  \hspace{1cm} (2.31)$$

Stochastic gradient descent is the most common technique used to train the networks; the network training is therefore tackled as a global optimization problem that aims to the reduction of a loss function [225, 226].

Even though CNN algorithms were first introduced in 1990 by LeCun, they only became widespread after 2010, when the increased computation power of modern hardware made possible to build deep networks. Deep networks generally refer to neural networks with a large number of layers, and therefore a large number of parameters to train. The higher the number of parameters, the slower the training of the network becomes [227]. The key of the success of CNNs is its ability to extract low, mid and high level features along its multilayered structure.
Chapter 2. VO₂ Oscillators: from the Material to the Applications

The stacking of several processing layers gives the network the ability of extracting complex features working with progressive levels of abstractions. Many works have investigated the feature extraction of CNN across different layers of the network [228, 229, 230]. These studies revealed that the first layers of the network focus on extracting low level features, such as horizontal, vertical, and diagonal edges or differentiating colors. Moving up on the layers hierarchy, more complex features are recognized, such as circular shapes, or similar textures. In the last layers the different feature maps are able to select and group different categories, such as animals, faces, flowers. The first breakthrough of CNN was given in 2012, when Krizhevsky et al. demonstrated a 10% improvement in recognition probability of a CNN called AlexNet compared to computer vision techniques [231]. The AlexNet architecture comprised 5 convolutional layers and 3 fully-connected-layers, for a total of 60 million trainable parameters [227]. The filter size of the convolutional layers were between 11×11 and 3×3 dimensions. More recent architectures have proven that the CNNs perform better when the kernel size is reduced to 3×3 dimensions, which is the chosen dimension for most of the famous deep CNN architectures, like VGG and ResNet. The VGG architecture [8] is a deep network, with more than 10 layers (16 in the VGG-16 architecture and 19 in the VGG-19 architecture) that utilizes exclusively 3×3 convolutional filter sizes and introduces the max pool layers only every few convolutional layers. Deeper networks, despite being more resource hungry, have in general demonstrated increasingly better recognition performances. However, when going to very deep networks the recognition accuracy of the system can saturate. The saturation occurs because of the vanishing gradient, a phenomenon for which the gradients calculated from the loss function and used to update the weights of the network become close to zero, practically impeding further weight update. This problem cannot be easily resolved with increasing the training dataset, as it is related to loss of information between the layers of the networks [232, 233, 234]. The ResNet alleviated this problem introducing the concept of skip-connections [9], where the output of the current layer is summed to the output of the previous layer without undergoing further filtering. The ResNet architecture counts 50 layers and over 23 million parameters to train and has been very successful in achieving record recognition performances on the ImageNet dataset.

The performance of a network of recognizing and extracting the right features that would allow for a successful recognition greatly depends not only on the network implementation, but also on the availability of large training datasets. In many cases, the performances of a CNN on a test set can be increased increasing the number of training images. However, the larger the training dataset, the longer the time the network needs to be trained. Recent studies that focus on the reduction of the training time of CNNs, demonstrate that the feature extracted in a CNN to recognize an image can be transferred from a pretrained, specialized network to a generic recognition network with a process called a transfer learning algorithm [235, 236]. This technique offers the advantage of not having to train from scratch a network with a specialized architecture, but rather to re-use the weights already trained on another network or dataset.

In general, CNNs can be very energy and time demanding. The time and energy consumption of a CNN increases with the number of parameters and images that need to be trained. The
2.5. Convolutional Neural Networks

training algorithm is nowadays run mainly of powerful hardware such as very fast GPUs and FPGA platforms. It is known that the major cause of energy and time usage of these algorithms is given by the overhead on memory access, needed to retrieve at each step the input and training parameters of the network to perform the operations of vector matrix multiplications [237, 2, 238, 239]. Therefore, massive parallelization and introduction of memory buffers are commonly used in specialized hardware to achieve better performances in such algorithms [2]. Novel architectures based on neuromorphic computing concepts, that remove the classical separation between memory and CPU and perform the so-called "in-memory computing" can become disruptive to accelerate the deep neural network implementation in fast, energy efficient platform, enabling new functionalities in the cloud and at the edge [240, 241, 242]. In the next chapters we will discuss how oscillatory neural networks can be used as hardware accelerators for the convolutional operations in convolutional neural network, therefore bridging the gap between image retrieval capabilities of Hopfield nets and modern algorithms used for image classifications.
3 Experimental Methods

In this chapter, the experimental methods employed in this thesis are reviewed. Starting from the VO$_2$ thin film preparation, we discuss deposition techniques and annealing of the material. We present Raman spectroscopy and resistivity measurements to characterize the properties of the films. In addition, we present a scanning thermal probe microscopy technique that is here used to explore the temperature distribution upon the phase transition in the fabricated scaled devices. Finally, we briefly introduce a behavioral model of the VO$_2$ resistors employed for conducting circuit simulations of the coupled oscillatory networks.

3.1 Device Fabrication

3.1.1 Deposition Techniques

As already discussed in section 2.1.1, multiple approaches can be followed to grow VO$_2$ films. In the course of this thesis, two techniques were explored in the frame of the Horizon 2020 European Phase-Change Switch project, to obtain high-quality thin films of vanadium dioxide on silicon: pulsed laser deposition and atomic layer deposition. The vanadium dioxide films were deposited in both cases on a 4 inches Si wafer on top of 1 $\mu$m of thermal oxide, to provide electrical insulation between the VO$_2$ and the silicon. The investigation of different deposition techniques was carried out with the aim of identifying the procedure which would retain the better film quality in terms of thickness uniformity, crystallite size, roughness, alongside better electrical quality, measured on the steepness and width of the phase transition and its reproducibility in electraly-activated devices.

Pulsed laser deposition (PLD) is a deposition technique that is conducted in a vacuum chamber, where a target of the material to be deposited is vaporized through a high power, pulsed laser beam (figure 3.1). The evaporated material forms a plasma plume, which results in the deposition of the material on a wafer surface. In case of deposition of oxides, the stoichiometry of the deposited material can be adjusted with the insertion of other reactants, typically oxygen. Compared to other deposition techniques, in PLD the energy of deposited particles is
Figure 3.1 | Illustration of a PLD system. The target is vaporized through a high power laser beam. The material deposits on a sample, which is kept at a high temperature by an external heater, to favor the reaction with the sample surface.

typically an order of magnitude lower, which makes this technique more suitable for epitaxial deposition or for ultra-thin films where low roughness is essential. Typically, PLD systems use small targets of 1-2 inches, therefore the layers with uniform thickness are limited to sizes of a few cm$^2$.

PLD-deposited VO$_2$ films have been reported to have steep and narrow hysteresis when deposited on sapphire [132, 243, 244]. Other works have extensively investigated the formation of polycrystalline films on Si/SiO$_2$ substrates [245, 246, 247]. In the scope of this project, the PLD VO$_2$ films were deposited with a Solmates PLD tool in EPFL nanolab by M. Cavalieri. The system uses a V$_2$O$_5$ target, evaporated through a KrF laser with 248 nm of wavelength, a frequency of 20 Hz and a laser energy density of 10 J/cm$^2$. The chamber pressure was kept at 0.01 mbar with an oxygen flow of 5 sccm, corresponding roughly to an oxygen partial pressure of 10$^{-4}$ mbar. The deposition was conducted at a temperature of 400°C and was followed by an anneal process of 450°C for 10 minutes. The deposition rate was calculated to be around 10 nm/min. The resulting films deposited on SiO$_2$/Si substrates showed dense polycrystalline structure with the grain sizes between 50 and 100 nm. The film uniformity is limited to a 2 cm × 2 cm area in the middle of the wafer, as revealed from the ellipsometry thickness measurements presented in figure 3.2. The porosity of the film increases with the distance from the center to edges, while the thickness decreases by 25-30% near the edges compared to the middle of the wafer (figure 3.2 A and B).

Atomic layer deposition (ALD) is a vapor phase technique which deposits a material onto a substrate utilizing alternating precursors, which are introduced sequentially. At each deposition pulse, one of the precursor gaseous molecules reacts with the surface in a self-limiting way. The reaction of the precursor with the substrate surface stops once all of the reactive sites
3.1. Device Fabrication

Figure 3.2 | Left: ellipsometry measurement of the thickness of the VO$_2$ film deposited with PLD. The thickness is measured in a 6 cm $\times$ 6 cm area in the middle of the wafer. At the center of the wafer, the film presents a uniform thickness that results in a continuous material, as shown from the scanning electron microscopy image taken in the area A of the wafer. The thickness of the film decreases greatly from the center towards the edges, resulting in the breaking of the film in discontinuous, separated grains, as underlined in the scanning electron microscopy picture of the area B of the wafer.

The process of ALD can be performed with moderate heating of the substrate, which makes it a favorable technique for silicon back end of the line compatibility. In literature, different precursors were explored for the deposition of vanadium dioxide on a silicon substrate through ALD. The principal precursors utilized for vanadium were vanadyltriisopropoxide (VTOP) [248], VCl$_4$ [249], and tetrakis(ethylmethylamino)vanadium (TEMAV) [135, 250, 251, 252]. Between the various precursors, TEMAV was particularly successful for the deposition of VO$_2$ as in TEMAV the vanadium atom has an oxidation state of V$^{4+}$ [248]. This facilitates the synthesis of VO$_2$, which presents in its molecular structure V$^{4+}$ ions, compared to V$_2$O$_5$ which features V$^{5+}$ ions [250]. As a second precursor, in combination with TEMAV, water (H$_2$O) [251, 250] and ozone (O$_3$) [135, 252] have been successfully employed. The ALD as-deposited films result to be oxygen poor (VO$_x$), amorphous films. The right stoichiometry of VO$_2$ can be obtained by a post-anneal process under controlled oxygen partial pressure, which also causes the material to crystallize in a rough, granular film.

In the scope of this thesis, we investigated ALD deposition of VO$_2$ with TEMAV and both ozone and water as precursors. The ozone process, even though it demonstrated successful deposition of VO$_2$ at first, was not reliable in its reproducibility. Therefore, the devices presented
in the following chapters were fabricated with a water based process, in the Oxford ALD tool of IBM Research Laboratory in Zurich and in a Savannah ALD-100 of Cambridge University by our partners K. Niang and G. Bai. The two systems utilized different successful recipes for depositing the VO$_2$. In the Oxford ALD tool the deposition was conducted at a controlled pressure of 0.01 mbar with a 3 s TEMAV pulse followed by a 10 s purge time with N$_2$, and with a 5 s H$_2$O pulse time followed by a 10 s purge with N$_2$. The chamber temperature was kept at 150° C. In the Savannah tool, the TEMAV pulse was kept at 0.02 s, followed by a N$_2$ pulse of 3 s. This was repeated 8 times before introducing the water precursor, with a 0.02 s pulse and 5 s N$_2$ purge cycle. The chamber pressure was kept at 0.13 mbar and the temperature at 150°C [253, 254]. For both the processes, a post-anneal process was necessary to stabilize the film in the right stoichiometry. We explored two different annealing processes, discussed in Section 3.1.3, which resulted in different roughness and grain properties of the material.

3.1.2 Raman Spectroscopy

Before discussing about the annealing techniques employed to oxidise the VO$_x$ films after the ALD deposition process, we need to briefly introduce the Raman spectroscopy technique applied to vanadium oxides films. Raman is a spectroscopy technique which relies on inelastic scattering of a light beam impinging upon a liquid or solid material. It is able to give clear indication about the composition, molecular structure, phase and polymorphism, stress and contamination of a material. This spectroscopy technique relies on a laser light source, which interacts with the material, and in particular with the molecular vibrations or phonons. The elastic scattered radiation (Rayleigh scattering) presents naturally the same wavelength of the laser source, and is filtered out by a notch filter. The inelastically scattered light is collected with a detector. A typical Raman spectra presents the the intensity of the collected scattered radiation for each frequency. The photons that lost energy during the scattering process represent the so-called Stokes radiation, while those which acquired energy represent the
3.1. Device Fabrication

Figure 3.4 | Raman spectra of different vanadium oxide stoichiometries deposited on silicon by reactive sputtering. VO₂ is characterized by the distinctive double peak at the Raman shift numbers of 193 and 223 cm⁻¹, while V₂O₅ can be identified by the 142 cm⁻¹ peak. Reprinted from [255] with permission from Elsevier.

Anti-Stokes radiation. Usually, the Stokes lines are the ones used to identify the chemical composition of the material, through a measurement of the wavelength shift of the collected light.

Vanadium oxide Raman spectra allow for recognition of the different stoichiometries of the material. This type of characterization was fundamental to characterize the thin films after deposition and annealing to check that the right material phase was achieved. Figure 3.4 shows the Raman peaks of different vanadium oxide phases. In particular, for this work it was important to distinguish successful deposition of VO₂ with unsuccessful deposition of its more stable compound, V₂O₅. As it is highlighted in figure 3.4, VO₂ is easily recognized by the distinctive double-peak around 200 m⁻¹ Raman shift, while V₂O₅ presents a prominent peak around 142 cm⁻¹ and other higher Raman shift peaks. In this work, the samples have been investigated with a laser of 561 nm of wavelength. In general, for very uniform materials, the height as well as the width of the Raman peaks can give an indication about the crystal quality of the films. However, given the high polycrystallinity of the VO₂ films on SiO₂, this quantification can be difficult to make and therefore was not pursued in the scope of this thesis.
Chapter 3. Experimental Methods

3.1.3 Annealing Techniques

As mentioned in section 3.1.1, the vanadium oxides films deposited with ALD are originally amorphous, and present an oxygen-poor stoichiometry. Post-annealing deposition under controlled oxygen partial pressure is fundamental to crystallize the film in the right stoichiometry.

In literature, as well as in the scope of this thesis, a standard anneal furnace was used to perform the annealing of the device under oxygen flow. In this respect, a small window of parameters ensures the oxidation to VO$_2$, with controlled oxygen partial pressure in the order of magnitude of 200 \text{ Pa}.

Figure 3.5 | The annealing time has an effect on the morphology of the VO$_2$ film. a) A scanning electron microscopy image of a 50 nm thick film, deposited at 150°C and annealed for 20 minutes at 400 °C, is shown. The film presents grains of average dimension of 50-80 nm and appears continuous. b) SEM image of a 35 nm film, which was deposited in the ALD tool at a higher deposition temperature (200°C). The film already appears disconnected and with grains as big as 250 nm. c) As the film is annealed for longer time, the grain size increases and pronounced voids appear between the grains. d) When annealed for two hours at the same temperature (400° C), the film completely disaggregates in big, isolated grains, in the fashion of droplets. On the bottom, the Raman characterization of the 35 nm film annealed for 20 minutes and for 2 hours show that in both cases the material is VO$_2$. In the two measurements, in fact, the distinctive double peak around the Raman shift of 200 cm$^{-1}$ can be recognized.
of $10^{-2}$ mbar up to 1 mbar and an annealing temperature between 400 and 500°C [135, 254]. At higher pressures as well as at higher temperatures, the films oxidize to the V$_2$O$_5$ stoichiometry. After the annealing, the film crystallizes in grains and reports VO$_2$ peaks when investigated with Raman spectroscopy. The annealing time also has a major influence on the crystallinity properties of the film, and in particular on the size of the grains and the uniformity of the film, as can be seen in figure 3.5. A 53 nm ALD film annealed at 450°C for 20 minutes was investigated with scanning electron microscopy (SEM), revealing a continuous film of granular nature, with average grain size between 50 nm and 80 nm (figure 3.5 (a)). The film prepared at the same condition but annealed for 2 hours presents instead bigger grains of 100 nm average size, which are not all connected but are intertwined with voids (figure 3.5 (c)). Moreover, when going to thinner films, a longer annealing time results in big, disconnected VO$_2$ nanoparticles, as shown for a 35 nm film in the insets (b) and (d). The process of forming nanoparticles upon heating of a thin film is known as solid state dewetting [256] and was already observed for VO$_2$ [139, 257]. Independently from the annealing time, the film resulted in VO$_2$, as revealed in the Raman spectra, which present the typical VO$_2$ double peak around 200−1 cm.

A second annealing technique, the flash lamp anneal (FA), was investigated for its properties of annealing thin films at reduced temperatures. FA has demonstrated to be useful to stabilize other types of phase change materials [258]. The aim for the exploration of the FA is to investigate the possibility of having a more compact film, with reduced roughness and without the presence of voids between grains. The FA technique is a thermal treatment generally used for semiconductor processing. It is characterized by a strong temperature gradient along the thickness of the sample during the annealing process. Thus, material characteristics can be altered with minimal thermal loading of the substrate below. To achieve this effect, the surface of the sample is heated through a short but powerful flash, obtained by charging a capacitor and inductor system and discharging it over a Xenon Flash lamp. The sample can be pre-heated through a local heater in the sample holder. The capacitor/inductor system pair determines the time of the anneal (20 ms for the experiments conducted in this work). The anneal can be performed in vacuum, or reactive gases can be introduced in the chamber. In the case of VO$_2$, oxygen is introduced in the chamber. Different conditions for the anneal were explored, varying three parameters: the substrate heating between 200°C and 400°C, the power of the lamp and the oxygen pressure. Through Raman investigation of the samples after anneal, we could discern the annealing condition which stabilized V$_2$O$_5$ and the ones which resulted in the oxidation of the VO$_2$ stoichiometry. A combination of too high temperature, annealing power, or oxygen pressure results in the formation of the V$_2$O$_5$ stoichiometry. Vice versa, a combination of low annealing power, temperature or oxygen pressure would fail to anneal the sample, which resulted still amorphous. In table 3.6 the annealing conditions that successfully brought to the stabilization of VO$_2$ are summarized.

Interestingly, for certain combinations of lower power and substrate temperature, it was possible to obtain smoother films with smaller grain size. This result could be of great aid in obtaining connected film (without voids) also when reducing the film thickness.
Table 3.6 | Different flash annealing conditions that led to the crystallization of the ALD films in the VO$_2$ stoichiometry are summarized. Higher pulse power, temperatures or oxygen pressure lead to the oxydation of the film in the V$_2$O$_5$ stoichiometry, while lower parameters fail to anneal the film. In each experiment the flash duration was 20 ms and the VO$_2$ thickness 50 nm.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Flash power (J/cm$^2$)</th>
<th>Temperature (°C)</th>
<th>Oxygen pressure (mbar)</th>
<th>Grain size (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70</td>
<td>300</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>300</td>
<td>100</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>300</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>90</td>
<td>290</td>
<td>30</td>
<td>smooth film</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>300</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Figure 3.7 | Figure a) and b): AFM characterization of the VO$_2$ films annealed respectively with flash lamp anneal and with the slow anneal technique described in the main text. The flash annealed sample is smoother, with smaller grains. c) The result of an analysis of the average grain diameter conducted with the AFM analysis software Gwyddion is shown. The flash anneal technique allows to tune the average grain size of the film down to 20 nm, while the slow anneal results in grains of 55 nm average size. In the graph, the points represent the normalized number of grains which have a certain diameter. For each film, the normalized grain count was fitted with a Gaussian to obtain the grain size distribution (solid line). d) Typical hysteresis curve of a flash annealed device and slow annealed device. The flash annealed device presents a higher resistivity in the metallic state and a less steep transition. The hysteresis width remains similar between the two films: $\Delta T \approx 12$ K for the flash annealed sample and $\Delta T = 11$ K for the slow annealed sample.

For a quantitative analysis, the films were characterized with atomic force microscopy (AFM) and the images were processed with the software Gwyddion, to calculate the average grain size of the annealed films. Figure 3.7 shows the AFM characterization of a flash annealed film compared to a film that is treated with the slow anneal process described above (400° C). The smaller grain size of the flash annealed film and its increased smoothness result evident. The choice of the set of parameters for the annealing conditions allows to trim the grain size down to 20 nm, as it is shown in figure 3.7 c.

To probe the phase transition characteristics of the film, the resistivity of the film was measured
with a four-probe technique while raising the temperature of the chip in a temperature-controlled cryostat. The typical resistivity vs temperature curves of a flash annealed film and slow annealed film are shown in figure 3.5 d. The slow annealed films usually present a phase transition with an on/off ratio of more than two orders of magnitude. The hysteresis curve of the flash annealed samples varies depending on the processing condition; however, they generally present an on/off ratio lower then two order of magnitudes. The width of the hysteresis, calculated at the middle resistance value in the phase transition, results to be very similar for the two films: $\Delta T = 12^\circ C$ for the flash annealed sample and $\Delta T = 11^\circ C$ for the slow annealed sample. The flash annealed sample completes the phase transition in around 20$^\circ C$, while the slowed annealed sample in around 15$^\circ C$, therefore resulting in a steeper hysteresis curve. The impact of the granular structure of the material and the grain dimension on the device performance are analyzed in the next chapter.

### 3.1.4 Device Processing

From the VO$_2$ films two types of devices were fabricated and investigated: planar devices and crossbar devices. The processing steps for patterning and contacting the scaled devices are shown in figure 3.8.

The processing is conducted on 4-inch Si (100) wafers. A thermal SiO$_2$ of 1 $\mu$m is deposited with plasma-enhanced chemical vapor deposition for providing insulation between VO$_2$ and the substrate. The processing steps then diverge for the planar and the crossbar devices. For the planar devices:

- the VO$_2$ film is deposited and annealed as explained in section 3.1.1 and 3.1.3. The 4-inch wafer can at this point be cut into smaller chips for further processing;
- a negative resist (AR-N from AllResist) is spun and patterned with e-beam lithography;
- the VO$_2$ is etched with a dry inductively coupled plasma (ICP) etching process, subsequently the resist is stripped;
- through the means of a positive resist (PMMA) and e-beam lithography, contacts are patterned;
- a layer of 50 nm gold is evaporated over a layer of 50 nm nickel, which promotes adhesion over the SiO$_2$ layer; lift-off is performed.

For processing the crossbar devices, the 4 inches wafer is usually cut already in smaller chips after the deposition of thermal silicon oxide:

- the SiO$_2$ is patterned with positive resist and e-beam lithography to define the bottom contacts of the device; with a reactive ion etching process, trenches of 20 nm depth are etched in the SiO$_2$. 

49
Chapter 3. Experimental Methods

Figure 3.8 | The processing steps for fabricating planar and crossbar VO$_2$ switches is here schematized. For the planar devices, the VO$_2$ is patterned with electron beam lithography followed by an ICP etch; with a further lithography step the contacts are deposited. For the crossbar switches, the process starts with the patterning of a trench in the SiO$_2$, which is consequently filled with evaporated nickel (for adhesion) and platinum contacts. The VO$_2$ is then deposited and annealed on top of the contacts and patterned with ICP etch. Finally, the top contact is deposited. On the right, an SEM image of a planar and a crossbar device is shown.

- the trenches are filled with a 10 nm nickel / 10 nm platinum evaporated contact, and lift-off is performed;
- the VO$_2$ film is deposited and annealed as explained in section 3.1.1 and 3.1.3;
- a negative resist (AR-N from AllResist) is spun and patterned with e-beam lithography; ICP is used to pattern the VO$_2$ in large areas over the bottom contacts. This patterning is necessary to isolate the devices from one-another;
- through the means of lift-off of a positive resist (PMMA) and e-beam lithography, the top contacts (Ni/Pt) are deposited.

The devices fabricated in this fashion have the shape of simple resistors, whose resistance value is highly dependent on the voltage applied. We have investigated planar devices with variable dimensions between 100 and 1000 nm in length, and 400 to 2000 nm in width, over 50 nm film thickness. It was found that planar devices are more reliable when the width of the device is higher than the length. For the crossbar devices, the area of the device is defined by the crossing of the top and bottom contacts. The devices were therefore fabricated with areas variable from 500 nm $\times$ 500 nm down to 70 nm $\times$ 70 nm. In this configuration, the length of the device is represented by the film thickness; 80 nm thick films were used in this case.
reduced thickness of the device is comparable with the grain size of the VO$_2$ films, so that between the top and the bottom contacts only a net of grains in a parallel fashion is formed.

3.2 Scanning Thermal Microscopy

Since the VO$_2$ phase transition can be triggered by Joule heating occurring inside the device, as it is discussed in section 2.1.2, it is of great interest to understand and measure the temperature distribution inside a VO$_2$ device upon the occurrence of the transition. To this end, a Scanning Thermal Microscope (SThM) was used to characterize the devices. The SThM measures presented in this thesis have been performed in collaboration with F. Könemann, F. Balduini and B. Gotsmann. Large part of the results were obtained during the master thesis work of F. Balduini, under the supervision of B. Gotsmann and myself.

Scanning thermal microscopy is a scanning probe microscopy technique which closely relates to atomic force microscopy, with the addition of a temperature sensor on the scanning tip of the instrument. Located in the IBM noise-free laboratories, the SThM is a state of the art, custom-built instrument that allows to measure the temperature at the nanoscale with a spatial resolution down to 6 nm [259]. The measurements are conducted in ultra-high vacuum. A schematic of the measurement set-up can be seen in figure 3.9. The device under test is activated electrically via a modulated voltage signal, while a cantilever tip scans over the sample and is used to measure the self-heated, spatial temperature distribution of the device.

The scanning cantilever has a nanometer-sharp tip realized in Si [260]; the tip is electrically isolated from the sample by a SiO$_2$ layer that naturally forms upon processing. The cantilever is highly doped with phosphorous atoms ($10^{20}$ cm$^{-3}$), except for the region where the tip is located, of about $2 \times 2$ µm, which is doped at $10^{17}$ cm$^{-3}$. This region serves at the same time as an integrated heater and as an integrated resistor, whose resistance is temperature dependent, and can therefore serve as a heat sensor. The resistance of the temperature sensor is measured via a Wheatstone bridge configuration, and its spectral components demodulated via a lock-in amplifier. The relation between the tip resistance and the tip temperature can be extracted through a calibration procedure, as described in [259]. This calibration procedure requires strong assumptions and it's the major source of uncertainty in the measurements, resulting in 20% inaccuracy in the measured temperature. The sensitivity to the temperature variations along the measurement is significantly higher, in the order of $\Delta T = 10$ mK. For performing a measurement, the cantilever is heated with a constant voltage to an out-of-contact temperature $\Theta$, which is usually chosen around 300°C. When posed into contact with an electrically activated sample, the tip exchanges heat with the device; the heat exchange is translated to a shift in the integrated resistance of the tip, which can be measured with the Wheatstone bridge set-up. In the in-contact configuration, the power $P$ dissipated by the heater can be calculated as:

$$ P = \dot{Q}_c - \dot{Q}_{ts} = \frac{\Theta - T_A}{R_c} + \frac{\Theta - T}{R_{ts}}, $$

(3.1)
Chapter 3. Experimental Methods

Figure 3.9 | Simplified schematic of the SThM setup. A cantilever with an integrated temperature sensor scans over the device under test. The device is electrically activated with a DC voltage and a superimposed AC component. The electrical activation results in a Joule or Peltier heating effect in the device. The temperature at the nanoscale can be measured via the cantilever sensor as a voltage signal over a Wheatstone bridge configuration, which is successively amplified. The temperature rise of the device, as well as the voltage drop across the Wheatstone bridge is modulated with harmonic components proportional to the harmonics of the AC voltage applied to the device. A lock-in amplifier is used to demodulate the signal. Post-processing techniques allow to derive the device temperature from the harmonic components of the Wheatstone bridge voltage. Image courtesy of S. Hönl and F. Könemann.

where \( \dot{Q}_{cl} \) and \( \dot{Q}_{ts} \) denote the heat flux through the cantilever and through the tip, \( T \) is the sample temperature and \( T_A \) is the ambient temperature. \( R_{cl} \) is the thermal resistance between the tip and the cantilever and \( R_{ts} \) the one between the tip to the sensor. \( R_{ts} \) can be very difficult to estimate, as it greatly depends on the point-contact between the tip and the sample. \( R_{ts} \), in fact, varies in dependence of the material the tip gets in contact with. In addition, the value of \( R_{ts} \) depends on how the tip gets in contact with the sample, meaning that, when the tip touches the material, multiple point-of-contacts can be present, given the topography of the sample [261]. However, as explained in detail in [262, 259, 263], the spatial distribution of the temperature of the device can be calculated independently from the value of \( R_{ts} \), when a technique that introduces a modulation in the temperature signal is used. This is generally realized by applying a modulated voltage to the sample, that will translate in modulated self-heating effects in the device, i.e. modulated temperature changes via Joule or Peltier effects. This modulation ultimately allows to link the temperature of the sample with the DC and AC spectral components of the voltage drop across the Wheatstone bridge \( \Delta V_{AC} \) and \( \Delta V_{DC} \):

\[
\Delta T_{DC} = \Delta \Theta_{outofcontact} \times \frac{\Delta V_{AC}}{\Delta V_{DC} - \Delta V_{AC}},
\]

where \( \Delta T = T - T_A \). A complete derivation of this equation can be found in [259, 263]. It has to be mentioned that the proposed derivation of the sample temperature can undergo some variation when the device is non-linear, or the AC modulation of the device voltage signal is
unipolar. For further details about the methods of analysis of the SThM temperature maps of the VO$_2$ samples, we invite the reader to refer to [262]. The SThM technique is here used to explore the phase transition on planar VO$_2$ switches, and to investigate whether the phase change happens uniformly through the device or a filament is formed as reported in previous work on crystalline films [158].

### 3.3 Device Model for Circuit Simulations

For understanding the impact of different networks schemes on the computational capabilities of VO$_2$ coupled-oscillators, and in particular for designing circuits exploring different coupling and biasing schemes of the devices, it is necessary to realize an electrical model of the VO$_2$ which can be used in a circuit simulation framework. Many different simulation schemes have been explored in literature for VO$_2$. For example in [264, 265], a physical model of a 1-D device is implemented to take into account the Joule heating distribution over the electrically-activated phase transition. The model is capable of circuit simulations when the other circuit components are introduced through an analytical description. Other physical models have been explored in TCAD tools [266], however, the integration of the physical model with a circuit simulation environment is not introduced. To implement a device model in commercial circuit simulators, for example in SPICE, multiple works rely on the coding of a compact model, in which the abrupt transition of the device is not described through its physical properties, but rather through the introduction of a non-linear function of the input electrical variables (voltages and currents). An overview of the compact models used to describe the hysteretic behaviour of non-linear devices can be found in [267]. More specifically, compact models of VO$_2$ can be found in literature [268, 164]. In this work, we use a behavioral model proposed by Maffezzoni et al. and described in [164]. This work presents a driving point equivalent model for VO$_2$ two-terminal device which is able to reproduce the hysteretic behavior. The equivalent circuit is depicted in figure 3.10 (a). In its insulating state, the VO$_2$ is modeled as a high impedance resistor which behaves according to the formula $V = R_{\text{ins}} \cdot I$, where $V$ and $I$ are the voltage and the current in the VO$_2$ resistor, and $R_{\text{ins}}$ its insulating resistance value. Equivalently, the device in its metallic state is described by the relation $V = R_{\text{met}} \cdot I$, where $R_{\text{met}}$ is the metallic resistance value. The transition between the metallic and insulating state is modeled with an abrupt (but continuous) transition, and it is decided by a voltage comparator with the following input-output relationship:

$$V_0 = 0.5 \cdot (1 \cdot \tanh(2\alpha V_{in}))$$  (3.3)

where $V_{in} = V^+ - V^-$ is the comparator input, $V_0$ its output voltage, $\alpha$ is the parameter that determines the slope of the transition curve. The voltage $V_0$ varies between 0 and 1 V and it is used to drive the input of the comparator $V^+$ by a voltage-controlled voltage source of gain $\Delta V = V_{\text{TH}} - V_{\text{TL}}$, where $V_{\text{TH}}, V_{\text{TL}}$ are the voltage thresholds for the IMT and the MIT respectively. The VO$_2$ device current is derived by the current that flows into the feedback
Chapter 3. Experimental Methods

Figure 3.10 | **a)** behavioral model of a VO\textsubscript{2} device, as described in [164]. **b)** Current versus voltage simulated curve of the behavioral model. Reproduced from [287], ©IEEE 2018.

resistor \( R_F \), whose conductance is described as \( G_F = \frac{1}{R_{ins}}(1 - V_C) + \frac{1}{R_{met}} V_C \). This model is implemented in Verilog A and used in SPICE simulations over the LTSPICE software, or equivalently in cadence virtuoso. In figure 3.10 (b) the I-V characteristic of a simulated device model with \( V_{TH} = 3V \) and \( V_{TL} = 2V \) is shown. The input voltage is applied between PIN+ and PIN-. At low input voltages the device presents an equivalent resistance \( R_{eq} = R_{ins} \) and a low current \( I_F \) is generated. Upon hitting the voltage threshold \( V_{TH} \) the device equivalent resistance becomes \( R_{eq} = R_{met} \), with an increase of the current \( I_F \). The sharpness of the transition can be tuned through changing the values of \( R \) and \( C \) in the device circuit model. In the following chapters, the VO\textsubscript{2} devices are simulated in coupled oscillator networks utilizing this model and adjusting the device parameters, such as the metallic and the insulating resistances and the voltage threshold, to the experimental values. As shown in the following, a careful tuning of the model parameters can reproduce with good precision the behavior observed in experiments. However, more complex simulation environments are needed to reproduce with more precision the phenomena observed in the devices, such as the occurrence of multiple phase transitions in multi-domain devices. One possible approach in this direction is presented in chapter 4, section 4.1.2. The behavioral model here described is instead employed in circuit simulations with the aim to explore the design and the computation capabilities of a network of VO\textsubscript{2} coupled oscillators.
4 Characterization of the Phase Transition in scaled VO$_2$ Devices

The investigation of the oscillatory neural networks presented in this thesis starts from the study of the relaxation oscillators building blocks: the VO$_2$ devices. ONNs are based on the synchronization phenomena of oscillators. Oscillators with different natural frequencies can experience frequency and phase locking when coupled together [78, 83]. However, high divergence between the natural frequencies caused by variability between the oscillating devices can prevent the synchronization even when a strong coupling is used [269]. To realize large systems of coupled oscillators, the devices should therefore yield high uniformity. Equally important for the success of the coupled-oscillators technology is the CMOS compatibility of the VO$_2$ device fabrication process. Therefore, a core investigation conducted in this thesis regards the fabrication of uniform VO$_2$ devices on a silicon platform. To this aim, we processed and characterized planar and crossbar devices from VO$_2$ films deposited on a SiO$_2$/Si platform and studied the impact of the polycrystallinity of the films on the phase transition and on the variability of scaled devices. Our main findings are discussed in this chapter.

Starting from the planar device, we present evidences of multiple switching events inside the film. We speculate that the origin of the multiple switching behavior is connected to the granular nature of the film. A scanning thermal microscopy study of the planar device is discussed, imaging the formation of current paths inside the device. An investigation conducted across crossbar devices of different dimensions highlights the possibility of reducing the number of phase transition steps with the scaling of the device dimensions. Lastly, we examine the achievement of a single, sharp, switching in a single grain device.

4.1 Characterization of VO$_2$ Planar Devices

Vanadium dioxide planar devices were fabricated from ALD and PLD films as explained in chapter 3, section 3.1. The devices were patterned into stripes, with variable dimensions (length and width) between 300 nm and 1500 nm. Their resistance was measured in vacuum, in a temperature controlled chamber, with a two probe measurement. The two probe measurement introduces a contact resistance, whose value was calculated to be around 500 $\Omega$. The
Chapter 4. Characterization of the Phase Transition in scaled VO\textsubscript{2} Devices

![Hysteresis curves of VO\textsubscript{2} planar devices. a) R-T curve of a VO\textsubscript{2} planar device with 1400 nm (width) × 300 nm (length) dimensions, fabricated from a PLD film of 50 nm thickness. An SEM image of the device is shown in the figure inset. The phase transition cycle is measured in a temperature-controlled chamber. The resistance measurement is conducted supplying a constant current of 100 nA and measuring the voltage across the device. The device presents a phase transition of more than two orders of magnitude, with insulating resistance R\textsub{ins} = 130 k\Omega at 330 K and a metallic resistance R\textsub{met} = 530 \Omega at 355 K. The phase transition is not continuous but proceeds in steps, which are reproducible, with slight variations, over multiple measurement cycles. The step-like behavior is attributed to the switching of single domains, possibly corresponding to the grains observed in the films (of average dimension of 50-80 nm). b) R-T curve of a planar ALD device annealed with the flash lamp anneal technique. The device presents similar dimension to the PLD device, with a thickness of 50 nm, and 1000 nm × 500 nm length and width. However, since the FA device has an average grain dimension of 20 nm, it comprises around 10× the grains respect to the PLD design. Its phase transition results smoother, with only a couple of identifiable steps. This observation suggests a correlation between the grains observable in the VO\textsubscript{2} film and the number of steps in the transition.

hysteresis curves, also referred to as resistance vs. temperature (R-T) curves of the insulator-to-metal (IMT) and the metal-to-insulator (MIT) transitions of the device were obtained applying a constant current and measuring the voltage drop across the device, during a temperature sweep. The applied current was kept to 100 nA, to avoid the self-heating of the device to play a role compared to the chamber temperature in triggering the phase change. A representative result of such measurement for PLD and ALD slow-annealed device is depicted in figure 4.1 a.

Compared to the hysteresis plot of the films shown in chapter 3, section 3.1.3, the hysteresis measurements of the scaled devices is not continuous, but proceeds in clearly defined steps. The steps are reproducible over multiple temperature cycles conducted on the same device, with slight variations, suggesting the IMT and MIT do not occur as a single transition, but rather as multiple, consecutive phase transitions in the VO\textsubscript{2} film. The multi-step transitions of VO\textsubscript{2} were already observed in literature, in crystalline nanowires deposited on lattice-matched substrates [270]. The discrete jumps are attributed to the switching of single domains inside the crystal; it has been shown in fact that the number of steps in the phase change can be reduced by scaling the device dimensions, therefore comprising a smaller number of switching domains in the device [271, 270, 272]. The asymmetry of the transition steps between the insulator-to-metal and the metal-to-insulator phase change has been explained by the
4.1. Characterization of VO₂ Planar Devices

different nucleation mechanism of the two transitions, facilitated respectively by point-defects or thin walls \cite{273, 274}. These findings suggest that, when scaling the device to comprise a single domain, a single, sharp transition can occur; this hypothesis is corroborated by an experimental study from Tsuji et al. on VO₂ nanowires, which show a single phase transition for a 20 nm long wire \cite{275}.

In the case of VO₂ deposited on SiO₂, experimental observations suggest that the switching domains could correspond to the grains that can be identified in the films. As discussed in section 3.1.3, the devices annealed with flash lamp anneal showed an average grain size of around 20 nm, much smaller than what usually obtained with slow annealing techniques (50-80 nm). Therefore, the R-T curve of a flash annealed device with dimensions comparable to the slow annealed device presented in figure 4.1 (a) was investigated. The FA devices, for comparable dimensions to the slow annealed device, given that they present a VO₂ film with lower grain size, comprise a higher number of grains. The typical R-T curve measured for one of such devices is depicted in figure 4.1 (b). The curve appears smoother and only a couple of resistance steps are present. Similarly, the resistivity measurements conducted on the VO₂ blanket films were also continuous and did not present step-wise phase transitions (figure 3.7). The continuous hysteresis observed on the resistivity measurements of the film as well as on the scaled flash annealed devices can be explained by an averaging of step-wise single-grain changes over a higher number of grains.

The impact of the granularity of the film on scaled devices in the electrically-driven phase transition was also investigated. To this aim, the current vs. voltage (I-V) characteristic of the planar devices was measured, sweeping the source current through the device and measuring the voltage. Upon the occurrence of the first phase-transition, an irreversible phase change is reported for all the devices. When investigated with SEM imaging, the devices which undergo the irreversible phase change present a modified morphology, with the clustering of previously separated grains (figure 4.2 (c) and (d)). This irreversible change can lead to failures, lowering the yield of the sample. The mechanism of the failure and of the morphology change in planar VO₂ devices deposited on Al₂O₃ has been studied and documented by Shabalin et al. \cite{276} with current-activated devices. The authors suggest that the device irreversible change and failure is due to current spikes caused by charging of parasitic capacitances upon the abrupt phase transition of the device. The authors indicate the employment of design techniques apt to reduce the parasitics as a possible solution to mitigate the problem. Alternatively, employment of a current source with faster response time can also help to prevent current spikes. Surely, this problem is more pronounced when a voltage source is used to conduct the I-V measurements, as the current spikes are in this case not compensated by the current source. For this reason, for the planar device operation, we found effective to initialize the device with a current source-measurement triggering in a controlled way the irreversible phase change and stabilizing the subsequent phase transitions, before employing the devices in a relaxation oscillator circuit.

In figure 4.2, the initialization cycle of a VO₂ planar device is shown. The phase transition is
Chapter 4. Characterization of the Phase Transition in scaled VO$_2$ Devices

Before the irreversible change
After the irreversible change

Figure 4.2 | Initialization cycle and current vs. voltage measurement of a planar VO$_2$ ALD device with dimension $1300 \text{ nm (width)} \times 200 \text{ nm (length)} \times 50 \text{ nm (thickness)}$. a) Plot of the device resistance upon the sweep of the current source. After an irreversible change, the curve stabilizes. The curve shows multiple phase transition for the electrically-activated device. The insulating resistance is calculated to be $105 \text{ k}\Omega$, while the metallic resistance at $90 \mu\text{A}$ is $5 \text{ k}\Omega$. b) Current vs. voltage characteristic. c) and d) SEM images of a device before and after the irreversible phase change, respectively. Images reproduced from [277], available at: https://doi.org/10.1016/j.sse.2019.107729; licensed under CC-BY 4.0. Full terms at https://creativecommons.org/licenses/by/4.0.

Electrically triggered sourcing a current through the device. The first phase transition results in an irreversible change that lowers the value of insulating resistance; after, the device stabilizes, allowing for a reliable and reproducible current vs. voltage (I-V) characteristic. Noticeably, as registered in the R-T measurements, the phase transition between insulating and metallic state proceeds in steps, associated with switching events of consecutive grains. The number of steps that we record in this measurement is lower compared to what results from the R-T measurements. This can be explained by the formation of a filamentary portion of metallic material connecting the two electrodes, which can expand as the input electric signal is raised, as it is discussed in section 4.1.1.

The irreversible change, combined with the multi-grain switching, contributes to increase the device-to-device variability in the VO$_2$ planar switches. The devices fabricated starting from the PLD films had relatively low yield (ca. 50%), meaning that only in a few devices presented a phase transition after the irreversible change and could therefore be used for experiments on coupled oscillators; they exhibited a resistance variability from the mean value of $\Delta R_{\text{INS, MET}} = 25$–58%, and a threshold variability $\Delta V_{\text{TH}} = 20\%$, measured across 12 devices with equal dimensions. The devices fabricated from the ALD films showed increased reliability; they did not experience failure upon the initialization cycle and the relative variabil-
4.1. Characterization of VO$_2$ Planar Devices

Figure 4.3 | Temperature maps derived with the SThM. a) and c) AFM scans of two VO$_2$ planar devices realized with ALD and slow anneal of the VO$_2$ film at 400° C for 20 minutes. The film presents grains of 80 nm average dimension. b) and d) temperature maps of the devices in their metallic state. The device in b) was measured with a voltage input $V_{in} = 6.5 \pm 1.5$ V and a series resistor $R_s = 100$ kΩ, while the device in d) was measured with $V_{in} = 1.6 \pm 0.2$ V and $R_s = 1$ kΩ. The AC modulation was set at a frequency $f = 1.2$ kHz. The scale bar next to the image refers to the increment of temperature over room temperature produced by the self-heating of the device.

ity of their parameters was reduced to $\Delta R_{INS, MET} = 5$-20% and $\Delta V_{TH} = 10%$. We attribute the better performance of the ALD devices to the more homogeneous characteristics of the films (discussed in section 3.1.1 of chapter 3). The device-to-device variability has an impact on the oscillator performances: the more non uniform the devices are the stronger the coupling required to insure frequency and phase locking between oscillators. Moreover, the multi-step I-V characteristic can induce distortion in the oscillations, as described in the following chapter.

4.1.1 Characterization via Scanning Thermal Microscopy

The scanning thermal microscopy technique (SThM), described in section 3.2, was used to characterize the electrically-activated phase transition in the VO$_2$ planar devices. From literature, it is known that the electrically-activated insulating-to-metallic phase transition in VO$_2$ planar devices happens with the formation of a metallic filament connecting the two electrodes [158, 278, 279, 280]. With the SThM technique, it is possible to measure the temperature distribution of self-heated devices with nanometer resolution. In the scope of this thesis, this characterization has been conducted to explore the phase transition in the polycrystalline VO$_2$ devices realized on Si/SiO$_2$ and to study the effects of the film granularity on the phase transition.

As already explained in section 3.2, with the SThM characterization the AFM image and the temperature map of a device can be derived at the same time. The devices are driven with a voltage source and connected in series with an external resistance. When biased at a voltage
Figure 4.4 | Example of expansion of the VO$_2$ filament with increasing of the sourced voltage. The device under test reports dimensions of 1400 nm (width) × 500 nm (length) × 35 nm (thickness) and average grain size of 20 nm. The device was prepared with flash anneal as per the conditions reported in line 5 of table 3.6. a) Thermal map of the device in its insulating state. b) Map of the device in the metallic state. c) and d) evidence of filament expansion for increased electrical activation. The device was measured with a series resistance $R_s = 10$ kΩ. The applied voltage is indicated on the figure for each experiment, and was further superimposed to an AC voltage of 100 mV @ 1.2 kHz.

lower than the IMT voltage threshold, the temperature signal in the thermal map results to be low and no temperature profile can be detected. When biased at a voltage higher than IMT threshold, the device switches to its metallic state and the thermal map shows the formation of a filament with higher temperature, indicating the portion of the material which undergoes the IMT. Two measurements conducted on VO$_2$ ALD planar devices are depicted in figure 4.3.

Due to the cantilever geometry, the VO$_2$ area right next to the contacts cannot be accessed by the tip and therefore cannot be imaged. The contact separation which results from the AFM map appears shorter than what it is in reality. Nevertheless, the measurements reveal that the phase transition regards only one portion of the material and concerns only a few grains in the device, opening a low impedance current path connecting the two electrodes. Precise analysis of the measurements allow to identify the boundaries between the metallic and the insulating state of VO$_2$ (see appendix, section A.1, figure A.3).

Thermal maps measured on devices activated with progressively higher applied input voltages reveal that by increasing the electrical power, the portion of the material which undergoes
the IMT expands. An example is shown in figure 4.4, where a device fabricated from a flash annealed sample, reporting average grain size of 20 nm is investigated. The I-V curve of the device under investigation is reported in the appendix, section A.1, figure A.1 and shows multiple phase transition steps. The device was tested applying a series of bias voltages below and above the voltage thresholds of the consecutive phase transition steps. For a voltage bias lower than the first step in the insulator to metal phase transition, no significant heating is observed in the device. When biasing the device with an input voltage higher than the first IMT step, a metallic filament is formed between the two electrodes; the current concentrates in the portion of the material which presents a lower resistivity, producing higher heat dissipation in the filament compared to the rest of the device. As the bias voltage is increased, the thermal maps show an evolution of the metallic filament. In particular the analysis reveals that a new portion of the material undergoes the IMT, opening a second current path.

The SThM measurements also possibly reveals how grain boundaries impact the heat distribution in the VO$_2$ device. This, together with more details over the widening of the metallic filament, are discussed in the Appendix A.1.

### 4.1.2 Simulation of Phase Transitions in Planar Devices

In the previous sections, and in particular within the discussion of the thermal characterization conducted on the phase transition in the VO$_2$ devices, it emerged that the steps in the I-V and R-T characteristics can be assigned to the transition of individual grains. To further investigate this behavior, we developed a model of the switching in the devices, taking into account the presence of multiple domains which undergo the IMT. The discussion here reported is part of the master thesis of F. Balduini [281]. This work was conducted with the aim of obtaining a simple model to find the essential parameters which govern the multi-step IMT behavior of the planar switches. The VO$_2$ film is described as a square network of grains, in which each grain is treated as a unique body. The physical characteristics of each grains are defined as a distribution of the properties (temperature of the IMT, temperature of the MIT, insulating and metallic resistance) measured in the experiments or derived from literature (for more details, refer to section A.2 of the appendix). Similar models, which simulate the VO$_2$ as a network of resistors, have previously been proposed in literature [282, 168, 283]. The simulation algorithm can be schematized as following:

- The VO$_2$ device is modeled as a networks of variable, temperature dependent resistors; at room temperature and under no applied voltage, all the resistors are in the insulating phase. The grain boundaries are taken into account as connecting resistances between the grains; their heat dissipation was not introduced in the simulator.

- A voltage is applied to the device and the voltage and currents in each node of the resistor network is calculated. The Joule power dissipated from each grain is derived.
Chapter 4. Characterization of the Phase Transition in scaled VO$_2$ Devices

- The heat equation:
  \[
  C_v \frac{\partial T}{\partial t} = k \nabla^2 T + P_{joule} \frac{V}{V_v} + g \frac{T - T_{sub}}{t}
  \]  
  (4.1)

is solved for each grain, to determine its temperature. $C_v$ is the VO$_2$ thermal capacitance, $k$ its thermal conductivity per unit thickness of the film, $g$ the thermal volumetric conductivity between VO$_2$ and SiO$_2$, $T_{sub}$ the substrate temperature, $V_v$ and $t$ the volume and thickness of the grain.

- Given the temperature developed in each grain, the new resistance value of the grains is calculated and the algorithm is repeated until the steady state is reached.

In the model, the grain boundaries are considered as fixed resistors placed in between the grains, as represented in figure 4.5 (a). The simulated R-T characteristic is depicted in figure 4.5 (b) and, similarly to the experimental curves, it presents numerous IMT steps, caused by the phase transition of single grains inside the device. Having set the grain boundary resistance to be slightly higher than the metallic resistance of the VO$_2$, in the simulation the low-impedance state of the device is dominated not by the resistivity of the single grains, but rather by the resistances of the grain boundaries. In figure 4.5 (c) the comparison between a measured and a simulated I-V curve for a planar device is shown. The simulation model was investigated to reproduce the experimental I-V curve of a device characterized through the SThM. The SThM characterization was used to identify the grains which undergo the IMT. It revealed that the current path in the metallic state was formed by around 10 grains comprehended between the electrodes. The simulation of a device comprising a $2 \times 5$ grain matrix was fitted to the experimental curve, showing the capabilities of this simple model to reproduce the experimental behavior, from the hysteresis width to the presence of distinctive jumps in the phase transition of the device.

**Figure 4.5** | a) The VO$_2$ device is modeled as a network of temperature dependent resistors. The contact impedance, as well as the grain boundaries are taken into account as fixed resistances between each grain unit. b) and c) simulated R-T and I-V curves of planar VO$_2$ devices. The simulations reproduce the multi-step behavior in the electrical characteristics of the device. Similarity to the experimental curve is obtained by fitting the parameters reported in table A.4 of Appendix A.2.
4.2 Crossbar Devices

In this section crossbar devices are presented. Crossbar devices are investigated with the aim of reducing the device-to-device variability. In planar devices, the insulator-to-metal transition appears as a filament between the two contacts, which expands over a series connection of grains. The grain-boundaries are likely to represent high resistances in series with the metallic domains, therefore forming hot spots in the switched device (as discussed in the simulation presented above and in the appendix A.1). As highlighted in [276] and observed in this thesis, the excessive Joule heating can be the cause of the irreversible transition in planar devices. In contrast, the length of the device in the crossbar configuration is represented by the separation between the two contacts at the cross intersection, which corresponds to the thickness of the deposited film. This allows to create short devices, in which a matrix of parallel grains is present between the two electrodes, and no grain boundary is in series to the current path (figure 4.6). Therefore, by reducing the length of the device to comprise only parallel grains, we expect to reduce the filament length to a single grain, possibly eliminating the highly resistive grain boundaries and improving the variability between the devices.

The crossbar devices were fabricated from an 80 nm thick ALD film with the process described in section 3.1.4. The length of the device is represented by the distance between the two contacts, and therefore corresponds to the thickness of the VO$_2$ film, 80 nm. The area of the device was varied between 350 nm $\times$ 350 nm down to the scaled dimensions of 70 nm $\times$ 70 nm. The resistance vs. temperature curve of the crossbar has been investigated. In figure 4.7, the hysteresis cycles of a 250 nm $\times$ 250 nm and a 70 nm $\times$ 70 nm device are shown. Similarly to the planar devices, the R-T curve of crossbar devices shows a multi-step phase transition. In addition, the scaled device clearly shows less steps in the phase transition compared to the wider device. We attribute this characteristic to the reduced number of grains present between the contacts, therefore to the lower number of domains which undergo the IMT.

![Planar Device vs Crossbar Device](image-url)
Figure 4.7 | Hysteresis cycles of two crossbar devices, obtained from a 80 nm thick ALD film annealed at 400°C for 20 minutes. a) Resistance vs. temperature curve of a 250 nm × 250 nm device. As for the planar devices, it is possible to recognize multiple phase transitions in the curve. b) Resistance vs. temperature curve of a 70 nm × 70 nm device. Holding the device smaller dimension, the phase transition presents a lower number of steps, as between the two contacts there are less grains which undergo the IMT.

Figure 4.8 | The resistance vs. current and current vs. voltage characteristic of a crossbar device with 250 nm × 250 nm dimension, obtained sourcing a current and measuring the voltage across the device. The first activation of the device no longer causes an irreversible change. A multi-step phase transition associated with the IMT of different grains is observed also in this case.

Regarding the electrical activation, no irreversible change was observed for the crossbars. An I-V curve of the crossbar devices is shown in figure 4.8. Similarly to the planar device, also the crossbar shows a multi-step phase transition, associated with the consecutive switching of different grains. The variability of the parameters of the crossbar devices was also studied, showing an improvement compared to the planar devices.
4.3. Single Grain Devices

The resistance and threshold variability between devices of the same dimensions is reported to be $\Delta R_{\text{INS, MET}} = 7\text{-}10\%$ and $\Delta V_{\text{TH}} = 12\%$ respectively. The higher uniformity of the crossbar devices allowed to obtain the frequency locking of four oscillating devices, leading to the results described in section 5.2.3.

4.3 Single Grain Devices

The analysis conducted on the crossbar and planar devices highlight the role of multiple phase transitions in the variability between devices of nominally the same dimension. Moreover, resistance vs. temperature graphs reveal that by scaling the device dimensions and therefore incorporating a lower number of grains between the electrodes, the number of steps in the R-T characteristic of VO$_2$ is reduced. This motivated the investigation of the phase transition of a VO$_2$ single-grain device.

To this aim, the nanoparticles presented in figure 3.5 in chapter 3 were contacted and characterized in their electrical properties. In literature, the investigation of the phase transition of such structures has been conducted by measuring the reflectance of an incident laser [139] or the Raman intensity [284] over a temperature sweep of the chip. These measurements cannot address a single particle, but rather reveal the phase transition characteristics of multiple particles at the same time. To the best of our knowledge, no direct measurement of the resistance vs. temperature cycle of a single nanoparticle was conducted before.

We used electron-beam lithography to locate and contact the nanoparticles. A picture of the contacted nanoparticles is shown in figure 4.9 (a). The R-T hysteresis curve of the nanoparticles was studied with the same method used for the crossbar and the planar devices characterization. A temperature sweep was conducted in a temperature-controlled chamber and a source current was used to measure the resistance of the devices. To avoid any influence from the self heating of the device, a probe current of 10 nA was used. Moreover, the temperature ramp had a step of 0.02 K/s, to obtain precise measurements. As shown in 4.9 (b) and (c), the IMT of the nanoparticles is a single, point-sharp transition. The hysteresis width, however, can differ from device to device. Two over seven tested devices present a large hysteresis, of around 55° C (figure 4.9 (c)). The parameters influencing the width of the hysteresis are not well understood in literature. Multiple works suggest that the density of grain boundaries, together with the number of defects within a grain are the key parameters which determine the hysteresis properties [257, 284, 243]. It has been argued that in a continuous film a higher density of grain boundaries (smaller grain size) would increase the availability of nucleating defects, therefore reducing the width of the hysteresis [245]. Similarly, the reduced presence of defects is associated with a widening of the hysteresis. Suh et al. [139] have investigated the hysteresis of VO$_2$ nanoparticles through the reflectance of an incident laser, and, similarly to our findings, they reported hysteresis width in the order of 60 K. They argued that in free-standing nanoparticles the phase transition cannot nucleate at the grain boundaries of the material. The nucleation therefore needs to take place at the location of defects inside...
the particle; when the particle is highly crystalline, the number of defects which can nucleate the phase transition is reduced. Therefore, the higher is the crystal quality, the lower is the probability of nucleating the phase transition, resulting in the wide hysteresis of the film.

We conducted a characterization of the crystal quality of the nanoparticles with Scanning Transmission Electron Microscopy (S-TEM). A lamella was cut with a focused ion beam to examine the cross-section of pristine, non-contacted nanoparticles. The results of the TEM investigation are shown in figure 4.10. From this study, the nanoparticles result to have long-range crystal orientation; some of them appear to be almost defect-free (figure 4.10 (a)) with higher defect concentration on the edges. Other particles, like the one investigated in figure 4.10 (b), are polycrystalline; their different crystal orientation can be easily recognized with the Fast Fourier Transform (FFT) of the TEM image. This study could implicate that the devices which present a wider characteristic have better crystal quality, while polycrystalline devices are associated with a narrower hysteresis. To support this hypothesis, a TEM characterization of the electrically-measured devices is planned as future experiment.

Finally, the electrical activation of the single-grain devices was investigated. The devices could
4.3. Single Grain Devices

**Figure 4.10** | Transmission electron microscopy of a VO$_2$ nanoparticle obtained via solid solid state dewetting. Conducted in collaboration with M. Sousa. a) TEM image of a nanoparticle of around 100 nm dimension, which present long-range crystal order. In a1) and a2), two different areas of the nanoparticle are explored. The Fast Fourtier Transform (FFT) shows that the two areas have the same crystal order and orientation, suggesting a single-crystalline nanoparticle. The crystalline quality of the nanoparticle can be appreciated in the zoomed-in picture in a3). On the edges, however, the material appears to be more defective. b) A second nanoparticle is investigated with TEM. In the area b1) we can notice the coexistence of two different crystal orientations, which underlines the polycrystallinity of the nanoparticle. In the FFT of picture b), peaks corresponding at the two crystal orientations can be recognized. With the FFT analysis conducted on the areas b1) and b2), the peaks corresponding to the two crystal orientations can be distinguished.
be activated electrically utilizing a voltage source, and connecting them in series with a 1 kΩ resistance. The voltage drop across the series resistor was measured and used to calculate, with good approximation, the current passing through the VO₂ device. The result of such a measurement is shown in figure 4.11 and presents a single, sharp phase transition.

The single-grain devices represent a promising solution for reducing the device-to-device variability of the VO₂ phase transition, eliminating the randomness of the grains and grain boundaries. However, the formation of the isolated grains through solid state dewetting does not currently allow to control the position or the dimension of the grains. Further processing is needed to achieve this control and allow a more systematic study of these devices. Some approaches have already been proposed in literature [257]. Moreover, as explained in section 2.2, the hysteresis width is an important parameter which can determine the performances of a single oscillating device. For reducing the power consumption of the oscillators, a narrow hysteresis is desirable. The large hysteresis width shown by the single-grain device is therefore detrimental for a circuit exploitation of these devices. However, if it is indeed possible to reduce the hysteresis with introducing defects in the material, doping could represent a solution for both improving the device performances and pushing the phase transition to higher temperatures.
4.4 Main Achievements

In this chapter, we investigated the fabrication of VO₂ devices on silicon, deepening the understanding of the phase transition mechanism and devising a strategy to integrate VO₂ coupled-oscillators technology with a CMOS compatible process. In summary:

• We fabricated planar VO₂ devices on a silicon platform. We characterized them in their electrical properties, investigating the impact of the film granularity in the device performance. We identified the origins of a multi-step transition in the switching of consecutive grains inside the device.

• We used a state-of-the-art scanning thermal microscopy characterization technique on the planar devices, obtaining images with unprecedented resolutions of the metallic filament formation and evolution in an electrically-activated device.

• We created a simulation environment able to reproduce the behavior of VO₂ planar devices. The tool was used to investigate the impact of grain boundaries in the device thermal distribution.

• We proposed a crossbar configuration on a Si platform, scaling the device down to 70 nm × 70 nm. With this design, we achieved better control on the device phase transition, which resulted in devices with higher uniformity and lower variability.

• Finally, we investigated the realization of a single-grain VO₂ device on silicon. Through a solid state dewetting process, we were able to achieve formation of large VO₂ nanoparticles, which were contacted and studied in their electrical characteristic. The nanoparticles showed a single, sharp, phase transition, therefore eliminating the variability introduced by the presence of multiple transition steps in polycrystalline devices. To the best of our knowledge, this represents the first demonstration of a VO₂ device integrated on Si achieving a single, point-sharp phase transition.
Coupled Oscillator Networks based on VO$_2$ Devices

In the previous chapter we have examined the characteristics of VO$_2$ devices fabricated with different design strategies. In this chapter, we focus our attention on the circuit implementation of coupled oscillators based on the phase transition of the previously described devices. Starting from a proposed circuit implementation of an Oscillatory Neural Network, we present an experimental investigation of the pattern storage and recognition methods. We first address, in experiments and in simulations, the phase synchronization dynamics of two oscillators coupled through a resistive element. We discuss the impact of the non-idealities of the devices presented in chapter 4 on the oscillator trajectories. Via extending the proposed circuit design to three coupled oscillators, we demonstrate the recognition capabilities of the network when the test pattern is encoded in the relative time-delay of the voltage bias of the oscillators. Finally, the exploitation of a 4-coupled oscillator network as an image edge extractor filter is presented.

5.1 VO$_2$ Oscillators: Characteristic and Performances

The volatile, insulator-to-metal phase transition of VO$_2$ devices can be exploited in circuit applications to build compact, energy efficient oscillators. To achieve consecutive, self-sustained IMT and MIT in the device, it is necessary to set the working point in the negative-differential regime of the VO$_2$ I-V curve. As explained in section 2.2, this can be achieved through connection of the VO$_2$ device in series with a resistor or a transistor. A schematic of the two configurations is shown in figure 5.1. Compared to the design with a series resistor, the series transistor allows for tuning of the bias condition, and consequently, the output oscillator frequency, through calibration of the gate voltage [285]. In the following experimental discussion, both circuit configurations will be employed.

The VO$_2$ devices were measured with electrical probes and connected in the circuit configurations via wiring to off-shelf components. The best performing oscillators can operate with an input voltage of 1 V, and present an oscillation amplitude of less than 0.5 V. An example is depicted in figure 5.2 (a). The power consumption was calculated to be $P \approx 20 \ \mu$W. This
value was derived as \( P = V_{DD} \cdot I_{average} \), where \( I_{average} \) is the average current that is sourced from the supply voltage over one oscillating period. Similar figure of merits in the power consumption of VO₂ oscillators have been reported by other groups [78] and are competitive compared to other oscillator technologies (see table 2.11, section 2.4.1). In particular an estimation of the maximum frequency of oscillation was also conducted. As pointed out in section 2.2, the oscillation frequency depends from several parameters: the hysteresis width of the device (threshold voltages), the value of the circuit components and the voltage supply of the oscillators. To conduct the maximum frequency measurements, the oscillators were biased with a 10 kΩ resistor and the highest \( V_{DD} \) voltage allowed to obtain oscillations (see equation 2.11). As the experiments are conducted with externally coupled components, the oscillator circuit suffers from parasitic capacitances introduced through the probes and the wiring connections to the external circuit elements. In order to bring these parasitics to the minimum, the series resistor was integrated directly between the signal and the ground pads of the probe tip, as shown in the inset of figure 5.2 (b). The maximum oscillation frequency was registered to be \( f = 2 \text{ MHz} \) and was obtained with a crossbar device of 70 nm \( \times \) 70 nm dimensions, biased with \( V_{DD} = 3 \text{ V} \). We determined the parasitic capacitance introduced by the probe tip to be \( C_p \approx 80 \text{ pF} \). By reducing the parasitic capacitance, it is expected that the maximum oscillation frequency will further improve.

### 5.2 Coupled Oscillators

In this section the coupling dynamics of the oscillators and their computational capabilities are examined. The oscillatory neural network circuit configuration considered in this manuscript is depicted in figure 5.3. The circuit comprises multiple oscillator units, realized through the series connection of a VO₂ device and a transistor. The oscillators are fully-connected to each
5.2. Coupled Oscillators

Figure 5.2 | a) Waveform of a VO₂ oscillator which works with scaled voltage supply. The oscillator is driven with a 1 V voltage supply and presents oscillation between 0.1 and 0.28 V. Circuit parameters: \( V_{DD} = 1 \) V, \( R_S = 10 \) kΩ, \( C_P = 150 \) nF, \( R_{INS} = 104 \) kΩ, \( R_{MET} = 500 \) Ω, \( V_{TH} = 0.9 \) V, \( V_{TL} = 0.63 \) V. b) and c) Waveform and Fourier spectra of the maximum frequency oscillator, based on a crossbar device of 70 nm × 70 nm dimensions. To reduce electrical parasitics, the series resistor was integrated on the tip of the electrical probe. Circuit parameters: \( V_{DD} = 3 \) V, \( R_S = 10 \) kΩ, \( C_P = 80 \) pF, \( R_{INS} = 92 \) kΩ, \( R_{MET} = 800 \) Ω, \( V_{TH} = 2.5 \) V, \( V_{TL} = 1.5 \) V.

other through electrical coupling elements, specifically resistors or capacitors. As explained in chapter 2, the coupling elements of the ONN represent the memory of the circuit and therefore set the stable phase-relation configurations between the frequency-locked oscillators. In our ONN realization, differently from what is done in other VO₂ implementations \([196, 217]\), the input information is not encoded in the gate voltage of the series transistor, but in the time-delay of the input voltage of the oscillators. In the example of an image recognition process, the gray-scale information of the image is translated in a delay of the activation of the oscillator input voltage. The initial phase difference between the oscillators will therefore correspond to the time-difference set by this input delay. When the relative phase difference does not correspond to one of the memorized patterns, the oscillators relax to the nearest encoded phase configuration, recognizing the pattern. This circuit implementation of ONNs presents two main advantages:

- The circuit exploits the associative memory capabilities of ONNs to perform computation, therefore many patterns can be stored in the network by choosing the coupling elements; to perform recognition between different patterns, the network does not need to be reconfigured, but, provided a non-memorized input pattern, discrimination between all the stored patterns is done in a single calculation. In perspective, the coupling resistances can be implemented with memory elements, such as PCM and RRAMs, which would add reconfigurability to the system.

- The circuit encodes the information in the timing of the signals rather than in their amplitudes. The technology processes all the information, from the input to the output, through time-delay encoding. This makes the network implementation resilient to scaled voltage power supplies.

In the following, the experimental demonstrations of pattern-matching computations realized
Figure 5.3 | Schematic of the oscillatory neural network circuit. The input image gray-scale values are encoded in the delay of the input voltage of the oscillators. The memory of the circuit is provided by the coupling elements. The output is an oscillating waveform at the transistor drain. The information is encoded in the phase difference between the oscillating waveforms. Adapted from [286], ©2020 IEEE.

with this network are presented. The discussion concentrates first on how to store different patterns, i.e. different output phase relations in the network; afterwards, the first experimental demonstration of the input-delay information encoding to output-phase recognition process is presented.

5.2.1 Two Coupled Oscillators based on VO₂ Devices

In this section coupling experiments and simulations between two VO₂ oscillators are presented. The aim of this discussion is to show how different phase relations can be established between the oscillators with the use of different coupling resistance values. For this study, a system of two coupled oscillators is considered. Two coupling schemes are explored in experiment and simulations, the first one using a resistance as a coupling element and the second one introducing in addition a coupling capacitance.

For the experimental demonstration, two VO₂ PLD devices with the same dimensions (1000 nm (width)×700 nm (length)) were contacted and connected to series resistances $R_{S1}$ and $R_{S2}$ to realize single oscillator units. Subsequently, a coupling resistance was used to lock the oscillators in frequency (figure 5.4). The resulting phase and frequency locked waveforms of the oscillators are depicted in figure 5.5. When using as a coupling element a resistance $R_C = 3 \, k\Omega$, the oscillators oscillate in-phase; when increasing the coupling resistance to $R_C = 9 \, k\Omega$, the oscillators oscillate in out-of-phase. The two phase configurations are obtained over multiple experiments changing the coupling resistance between these two values. The phase difference between the signals of the two oscillators was measured registering the crossing of the 1 V threshold for the falling edge of the oscillators. The difference in time between the two signals was then divided by the period of oscillation ($T_{in-phase} = 2.3 \, ms$,
$T_{out-of-phase} = 3.2 \text{ ms}$) and expressed in degrees, resulting in $\Delta \Phi = 1.06 \pm 1.2^\circ$ for the in-phase coupling and $\Delta \Phi = 179 \pm 9^\circ$ for the out-of-phase coupling. Simulations conducted with the circuit model presented in section 3.3 and taking into account non-idealities such as the probe contact resistance were matched to the experiments.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{circuit.png}
\caption{Circuit scheme of two coupled oscillators. Reproduced from [287], ©IEEE 2018.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=1\textwidth]{experiments_simulation.png}
\caption{Experiments and matched simulation of two-coupled oscillators. The phase configuration (in-phase or out-of-phase) can be programmed by changing the value of the coupling resistance. Circuit parameters: $V_{IN1} = V_{IN2} = 3.2 \text{ V}$, $C_{LOAD1} = C_{LOAD2} = 150 \text{ nF}$, $R_{S1} = 26 \text{ k\Omega}$, $R_{S2} = 26 \text{ k\Omega}$; oscillator 1: $V_{TH} = 1.9 \text{ V}$, $V_{TL} = 0.7 \text{ V}$, $R_{NS} = 39. \text{ k\Omega}$, $R_{MET} = 7.6 \text{ k\Omega}$; oscillator 2: $V_{TH} = 2.2 \text{ V}$, $V_{TL} = 0.67 \text{ V}$, $R_{NS} = 23 \text{ k\Omega}$, $R_{MET} = 2.4 \text{ k\Omega}$. Reproduced from [277], licensed under CC-BY4.0. Full terms: https://creativecommons.org/licenses/by/4.0/.
\end{figure}
In the out-of-phase case, both in simulation and experiment, the oscillations present a double peak shape. This is attributed to the low value of the coupling resistance, which ensures a strong coupling between the oscillators: as the VO$_2$ device of oscillator $i$ undergoes the metallic-to-insulating transition, the equivalent impedance seen from the oscillator $j$ changes accordingly. The voltage partition between the series resistance $R_{ij}$ and the equivalent resistance seen at the node $V_{OSC,j}$, which is affected by the MIT, causes the double peak behavior. This high cross-talk between the two oscillators can be reduced by increasing the coupling resistance. For high impedance values of $R_C$, the equivalent resistance seen from the oscillator $V_{OSC,j}$ is less sensitive to the variation of the impedance of oscillator $i$ and the double peak can therefore be suppressed. The PLD devices, presenting a very large variability, could be coupled only with low resistance elements (strong coupling). The ALD planar devices, presenting a higher uniformity, allowed to increase the coupling resistance value of one order of magnitude.

Two ALD devices coupled in the out-of-phase configuration are shown in figure 5.6 (a). With a coupling resistance of 27 kΩ, the double peak behavior is not canceled, however it results smoother compared to the previous case.

**Figure 5.6 | a)** The out-of-phase oscillations of two resistively-coupled ALD planar devices. Circuit parameters: $V_{IN1} = V_{IN2} = 4.7$ V, $C_{LOAD1} = C_{LOAD2} = 150$ nF, $R_c = 27$ kΩ, $R_{S1} = 33$ kΩ, $R_{S2} = 36$ kΩ; oscillator 1: $V_{TH} = 2.1$ V, $V_{TL} = 0.9$ V, $R_{INS} = 37$ kΩ, $R_{MET} = 4.7$ kΩ; oscillator 2: $V_{TH} = 2$ V, $V_{TL} = 0.68$ V, $R_{INS} = 33.5$ kΩ, $R_{MET} = 7$ kΩ. **b, c, d)** In an experiment conducted with the same oscillators we can see an example of the influence of the VO$_2$ multi-step switching behavior on the oscillating waveform. **b)** Waveform of two coupled oscillators. While oscillator 1 presents a conventional waveform, oscillator 2 shows an unexpected plateau in the falling exponential. **c) and d)** Evolution of the resistance of the VO$_2$ devices across an oscillating period. The graphs show that oscillator 1 switches from an insulating to a metallic state, while oscillator 2 presents three difference resistance regimes, with the appearance of a third resistance value between the low impedance and the high impedance states. Circuit parameters: $V_{IN1} = V_{IN2} = 4.7$ V, $C_{LOAD1} = C_{LOAD2} = 150$ nF, $R_c = 29$ kΩ, $R_{S1} = 30$ kΩ, $R_{S2} = 33$ kΩ. Figures reproduced from [277], licensed under CC-BY4.0. Full terms: https://creativecommons.org/licenses/by/4.0/.
In figure 5.6 (b) a case in which the multi-step phase transition of the planar devices affects the oscillation waveform is shown. The waveform of the output voltage of two coupled oscillators is depicted. Oscillator 1 presents a conventional waveform, alternating between the high and low voltage thresholds of the VO\(_2\) phase transition. In contrast, oscillator 2, while presenting a conventional rising exponential, in the falling edge shows an abrupt transition between a falling exponential curve and a slowly-varying, almost steady state voltage. This behavior was investigated deriving the evolution of the VO\(_2\) resistance over time of the two devices, represented in 5.6 (c) and (d). The resistance is calculated from the analysis of the circuit, dividing the measured voltage falling over the VO\(_2\) from the current flowing through it. The first oscillator clearly shows a phase transition between the metallic and the insulating state in correspondence of the rising and falling edge of its output waveform. When investigating the resistance of the second VO\(_2\) device, three regions can be instead identified: a high impedance regime is followed by two distinctive regimes at lower impedance, which are connected to the plateau shown in the voltage oscillating waveform. The two distinctive low impedance regions are linked to the multi-step resistance change presented by the I-V curves of the devices discussed in section 4.1.

Simulations

The experiments presented so far showed the coupling of two VO\(_2\) oscillators in-phase or in out-of-phase configuration. However, when employing ONNs for image recognition, the encoding of gray-scale values in the network is also of interest. The stabilization of the relative phase of two-coupled oscillators to intermediate values was explored through simulations of an hybrid R-C coupling. Figure 5.7 (a) shows the possible output phase configurations that can be memorized when using a purely-resistive coupling. Systematic simulations confirm that, for this circuit scheme, only two output phase configurations are possible: either in-phase or out-of-phase state. The output phase difference of the two oscillators is plotted against the value of the coupling resistance R\(_c\), normalized by the nominal value of the metallic resistance of the VO\(_2\) oscillators. For values of R\(_c\) \leq 3R\(_0\) the oscillator presents an in-phase output configuration. For values of R\(_c\) \geq 4R\(_0\), the oscillators present an out-of-phase configuration. On the bottom, the corresponding output waveforms are plotted. Figure 5.7 (b) shows the output phase states achieved when a hybrid R-C coupling is introduced in the system. The simulations have been performed for a fixed value of the coupling capacitance, varying only the coupling resistance value. The output phase configuration spectrum obtained with this circuit design is much richer: for smaller values of resistance, the resistive-coupling brings the oscillator to lock in an in-phase configuration. The out-of-phase configuration is achieved with a sharp change for R\(_c\) \geq 4R\(_0\) as previously shown for the purely resistive-coupling configuration. However, with the addition of the capacitive element, a further increase in the coupling resistance value brings the phase-difference between the oscillators to stabilize to intermediate values. This investigation reveals the output phase in a system of two coupled oscillators can be varied continuously between 0° and 180° through the design of the R-C coupling, opening the possibility of memorizing not only black-and-white images, as previously demonstrated, but
5.2.2 Pattern Recognition with Three Coupled Oscillators

As briefly discussed in the previous section, the input signal delay between the two oscillators has no influence on the phase difference established by two-coupled oscillators. The oscilla-
5.2. Coupled Oscillators

Figure 5.8 | Simulations (a-b) and experiments (c) of the effect of an input phase delay in a two-coupled oscillator experiment. The relative delay of the inputs doesn't affect the output state of a two-coupled oscillator system; only the coupling weight is determinant for the output phase configuration. The experiments were conducted on the circuit of figure 5.5.

ory neural network theory predicts that multiple stable phase-relations can be encoded in an ONN. To achieve more than one stable phase configuration, it is necessary to scale the system to larger dimensions (more than two coupled oscillators).

In figure 5.9 the results of an experiment on a three coupled oscillator system is shown. In this experiment, the three oscillators are coupled with a set of resistive and capacitive elements (figure 5.9 (a)). The capacitances are fixed and ensure the frequency coupling of the oscillators. The coupling resistances are tuned to memorize two phase configurations, or patterns, in the network, which are depicted in figure 5.9 (b). From the Hebbian learning rule (equation 2.17) the relative, normalized weights for storing the chosen patterns are calculated:

\[ w_{12} = 1, \quad w_{13} = w_{23} = 0. \]  

(5.1)

These weights were matched through empirical search with the coupling resistances value of \( R_{12} = 300 \, \text{k}\Omega \) and \( R_{13,23} = 680 \, \text{k}\Omega \). With the coupling scheme presented, only patterns 1 and 2 from figure 5.9 (b) are stable phase configurations in the oscillating network. To initialize the network to an unstable configuration, the test patterns were encoded in the time delay of the voltage inputs. To demonstrate the effectiveness of this implementation for the phase initialization, the input of oscillators 1 and 3 were kept at a relative fixed time delay: oscillator 1 served as a reference and was initialized at time \( \Delta t = 0 \); oscillator 3 was activated with a delay \( \Delta t = T/2 = 225 \, \mu\text{s} \) respect to oscillator 1, where \( T \) is the oscillation period of the network (\( T = 450 \, \mu\text{s} \)). The delay of the input voltage of oscillator 2 was varied between \( \Delta t_2 = 0 \) and \( \Delta t_2 = T/2 \). An input time-delay of 0 represents a white pixel; an input time delay value of \( T/2 \) represents instead a black pixel; the values in between correspond to a gray-scale variations. Depending on its input, the oscillator 2 is expected to stabilize either in-phase with oscillator 1 (white pixel), therefore recognizing pattern 1, or in phase with oscillator 3 (black pixel), therefore recognizing pattern 2. In figure 5.9 (c) the waveforms for two recognition experiments are shown. Depending whether \( \Delta t_2 \) is closer to 0 or \( T/2 \), the output phase of the oscillator 2 stabilizes at 0 (top graph) or \( T/2 \) (bottom graph). The network is therefore able to recognize one of the memorized patterns when a distorted pattern is given as an input.
Figure 5.9 | a) Schematic representation of the experimental setup for the coupled oscillators image recognition experiment. The VO$_2$ devices were fabricated from the ALD films with slow anneal and in planar configuration. The VO$_2$ devices were tested in vacuum with a probe-card and connected in circuits with off-shelf components. Circuit parameters: $C_C = 1 \text{ pF}$, $R_{12} = 300 \text{k} \Omega$, $R_{13,23} = 680 \text{k} \Omega$. b) Trained patterns and test pattern considered for the experimental demonstration of image recognition. For the test pattern, oscillator 1 and 3 were kept fixed in the out-of-phase configuration, while the input delay of oscillator 2 was varied between the in-phase configuration respect to oscillator 1 (white pixel) and the out-of-phase configuration respect to oscillator 1 (black pixel). c) Experimental results of successful pattern recognition. The input delay of Oscillator 1 and 3 is kept fixed. By varying the input delay of oscillator 2, the two stored images are correctly recognized. d) Pattern recognition for all the entire range of input delays of oscillator 2. Intermediate delays result in recognition of an erroneous image. Reproduced from [286], ©IEEE 2020.
5.2. Coupled Oscillators

The experiment was repeated multiple times for different delays of oscillator 2, as depicted in figure 5.9 (d). The output phase of the oscillators is calculated after a stabilization period of around 10 oscillations, and averaged over multiple experiments; the error bars correspond to the standard deviation of the output phases. When the difference in gray-scale value between the trained pattern and the test pattern is below 20%, successful recognition is achieved. For intermediate time delays, the output phase stabilizes to a spurious pattern.

5.2.3 Feature Edge Extraction with Four Coupled Oscillators

With the coupling of a higher number of oscillators more complex functions can be demonstrated. More specifically, when the aim is to use coupled oscillator networks to perform image analysis, it is convenient to have a 2-D network to match the dimensionality of the image. The smallest 2-D network can be built with 4 oscillator units. In this section, we discuss experiments and simulations of a 4-coupled oscillator network adopted as an image filter. To implement an experimental demonstration of 4-coupled oscillators, it was necessary to optimize the device-to-device variability of the VO₂ switches. This was obtained through the design and fabrication of the crossbar devices. As with the three-coupled oscillators experiment, the oscillators were locked in frequency through resistive and capacitive elements. Their relative output phase was calculated taking the distance between the crossing of the 1 V line in...
Chapter 5. Coupled Oscillator Networks based on VO₂ Devices

the falling edge of the oscillator curves, designing oscillator 1 as the reference (figure 5.10 (a)). The network was programmed to recognize vertical, horizontal and diagonal patterns through the Hebbian Learning Rule. To the best of our knowledge, this is the first demonstration of 4-coupled VO₂ oscillators with memory capabilities realized on a silicon platform. The circuit parameters used to couple the oscillators were: $R_{12}, R_{13}, R_{24}, R_{34} = 82 \, \text{kΩ}$, $R_{23}, R_{14} = 130 \, \text{kΩ}$, $C_c = 5.6 \, \text{nF}$, $V_{GX} = 1.4-1.6 \, \text{V}$, $V_{IN} = 1.8-2.2 \, \text{V}$. The slight variation in the gate voltages of the transistors was used to bring the natural oscillation frequency of the single, uncoupled units to be close. This operation was needed to bridge the differences in the natural frequencies of the oscillators given by a 10\% of device-to-device variability. The horizontal, vertical and diagonal patterns memorized in the network were identified through multiple experiments. The result of the pattern recognition is depicted in figure 5.10 (b). In addition to the three memorized patterns, a fourth pattern where all the oscillators result equally spaced was identified. This spurious pattern is the same that was obtained in the three-coupled oscillators experiments in the area of erroneous recognition.

Systematic experiments on the recognition of noisy input patterns were hindered by random fluctuations of the oscillations and cross-talk noise. For example, referring to the diagonal edge 1 phase diagram in figure 5.10, we can observe that the phase data points of osc. 3 and osc. 4 occasionally spread between the in-phase and out-of-phase configuration respect to osc. 1. In some cases, this noise in the output phase of the oscillators leads to the locking of the system to another stable phase pattern. The input-delay to output-phase inference process was therefore investigated in simulations calibrated on the experimental results, in which the resistance and threshold voltage variability of the VO₂ devices was lowered to 5\%. In simulations the three stored patterns (horizontal, vertical and diagonal) are identified from noisy test patterns encoded in the time-delayed input of the oscillators (figure 5.11 (b-d)). In addition, two spurious patterns are observed: the pattern which sees all the oscillators equally spaced in the phase space plus a pattern in which all the oscillators are in phase with each other (figure 5.11 (a) and (e)). The arising of spurious oscillating patters is predictable, since in this realization the maximum memory capacity of the ONN - discussed in section 2.3.1 - is violated [289, 290]. Moreover, the all-in-phase configuration is predicted to be always a stable pattern for Hopfield networks [67]. The extra phase configurations presented by the network can be harvested as additional information, as discussed in the following. The $2 \times 2$ ONN was used as a filter for an edge-extraction operation performed on an image. To demonstrate this, handwritten digits from the MNIST dataset [291] were considered. The gray-scale pixels of the image were encoded as time-delays in the switching of the input voltage of the oscillators. In particular, the 256 gray-scale levels were mapped in the time-delay interval between 0 and $T/2$, where $T$ is the period of the coupled oscillator system. As shown in figure 5.11(f), when swiping the $2 \times 2$ ONN filter on an image of the MNIST dataset, vertical, horizontal and diagonal edges can be identified. In addition, the background as well as the center of the digits can be identified through the all-in-phase oscillating condition. This example demonstrates how a single network of coupled oscillators can serve as a filter for the feature edge extraction of an image. As discussed in section 2.4.2, the feature edge extraction capabilities of VO₂
5.2. Coupled Oscillators

Figure 5.11 | a-e) Simulation of the input time-delay to output-phase inference on a 4-coupled oscillators network. The circuit parameters are calibrated on the experiments. Vertical, horizontal, diagonal edges as well as the all-in-phase configuration are obtained. f) The ONN is used as a filter on a handwritten number image taken from the MNIST dataset. The filter is able to recognize vertical, horizontal, diagonal edges and the background in parallel, without any need to be reconfigured. In the examples above, the filtering operation was conducted with a stride of 2 for the image above (digit 3) and a stride of 1 for the image below (digit 4).

coupled oscillators were already shown in literature. However, the previous examples did not exploit the associative-memory capabilities of the system, but rather obtained similar results calculating the distance between the input image and the considered edge pattern. The system that is here proposed offers the advantage of being able to calculate the image edge without the need of re-configuring the system for each of the edges (horizontal, vertical, diagonal, background). The operation is conducted in parallel, therefore representing an advantage in terms of computation speed and hardware resources.
5.3 Main Achievements

In this chapter we discussed about the experimental demonstration of Oscillatory Neural Networks based on VO$_2$ devices. In particular, we presented a novel approach to ONN which relied on the encoding of the input information in the relative time-delay of the activation voltage of each oscillator unit and we validated the functionalities of our design in experiments. In summary:

- With a two-coupled oscillators circuit, we demonstrated in experiments and simulations the possibility of encoding different phase relations with tuning the coupling of the system. In particular, we showed a first experimental proof of providing frequency and phase locking with a resistance as the coupling element between two oscillators.

- By expanding the network to a three-coupled oscillators system, we demonstrated the retrieval of saved patterns from a distorted input pattern fed to the oscillatory circuit. The input of the system was encoded in the relative time-delay between the activation voltages of the single oscillators, an important characteristic which was first introduced with the specific circuit design we propose. We therefore realized a system which consistently computes the information in time.

- With four-coupled oscillators, we performed in experiments and simulations an operation of feature edge extraction on black-and-white images.

- Compared to other hardware and software approaches, which need to use different filters to recognize each of the edges, the system we propose is able to discriminate between vertical, horizontal, diagonal edges and uniform parts of the figures in a single computation.
6 **VO₂ Coupled Oscillators as Filters in Convolutional Neural Networks**

In the previous chapter we discussed the experimental demonstrations of oscillatory neural networks based on the phase transition of VO₂ devices, their pattern recognition and edge extraction capabilities. In this chapter, we take the oscillator technology a step forward, extending the exploitation of ONNs to support and accelerate the computation of traditional neural networks, such as Convolutional Neural Networks (CNNs). First, we present a convolutional neural network in which several digital convolutional filters are replaced with a $3 \times 3$ ONN unit. Further, we discuss the implementation of a backpropagation algorithm which can act on the ONN, to improve the recognition performance and to allow the trainability of the ONN filters in the CNN architecture. Finally, we introduce a circuit unit capable of transferring the phase information computed by a first ONN layer to a subsequent ONN layer.

### 6.1 **VO₂ Coupled Oscillators as Analog Filters in Convolutional Neural Networks**

In literature, ONNs are used primarily to recognize distorted or incomplete patterns. As already discussed in section 2.4.2, ONNs can retrieve from a $n \times m$ noisy image, an $n \times m$ output image, but do not provide a classification. Image recognition algorithms, as convolutional neural networks, process instead an input image of an unknown subject and provide as an output a classification vector containing the probabilities that the image in question corresponds to a given class. To understand how ONNs can be applied as hardware accelerators to convolutional neural networks, it is important to have an insight on how a CNN operates to classify an image. It has been proven that the first layers of CNNs usually extract low-level features from the image, as for example horizontal, vertical and diagonal edges [228]. The consecutive layers are able to distinguish more complex patterns, until the classification is achieved. In the last Chapter, a 4-coupled oscillators ONN was employed to extract edge features from an image, similarly to what is done by the first layer convolutional filters of a CNN. CNN utilize mostly filters of $3 \times 3$ dimensions, as they yield a good balance between accuracy and computational burden [8]. Therefore, when we aim to replace convolutional filters with ONNs,
it is not necessary to scale the ONN to the dimension of the image, but it sufficient to use very small networks, which can be more easily implemented in practical realizations. In the following, we perform simulations on a $3 \times 3$ VO$_2$ coupled oscillators network and employ it as a filter in a CNN architecture. The procedure used for this demonstration is summarized as follows:

- A convolutional neural network with a structure borrowed from the VGG-13 has been programmed with TensorFlow and trained for the recognition of the MNIST dataset. The CNN structure is schematized in Table 6.1. 10000 images from the dataset were considered, divided in 6000 training images and 4000 test images. A recognition accuracy of 97% on the test set was achieved.

- A $3 \times 3$ ONN was trained with the Hebbian Learning Rule to recognize horizontal, vertical, diagonal edges and the background of an image. The $3 \times 3$ ONN is used as a filter with stride 2 on the 10000 images of the MNIST dataset, obtaining as an output five $13 \times 13$ images corresponding to the feature maps of each edge.

- The filters that perform the same feature edge extraction operation in the CNN are identified. To this aim, the 64 trained filters of the CNN first layer were convoluted with the MNIST images. The resulting feature maps were activated with a ReLU function and compared with the ONN feature maps via calculation of the mean square error of the images differences. The filters which gave the minimum mean square error were therefore swapped with the ONN.

- Five digital filters corresponding to the extraction of horizontal, vertical, diagonal edges and background were replaced by a single ONN filter. In practice this was done replacing the feature maps of the first layer corresponding to the digital filters with the feature maps obtained with the ONN.

- The subsequent layers of the network were retrained on the new ONN-CNN dataset, reaching recognition accuracy of 95% on the test set.

The simulation environment was built between Python and LTSpiceXVII. In particular, the CNN was coded and trained with a backpropagation algorithm in Python with Tensorflow. The ONN simulations were performed in LTSpiceXVII with the VO$_2$ model described in section 3.3 and device parameters calibrated on the experimental results presented in section 5.2.3. For the ONN-CNN performance evaluation, Python was used to compute the ONN weights and the input delays of the oscillators from the input MNIST image, to call LTSpice and set the circuit parameters to perform the circuit simulations. Finally, the output waveforms were also transposed into a processed image in Python. Some consideration are necessary to fully depict the computational method employed to train the ONN-CNN network. The ONN is composed by 9 coupled oscillators; four patterns, plus the all-in-phase pattern, are stored with the Hebbian Learning Rule in the network. Unquestionably, the amount of stored patterns
6.1. VO₂ Coupled Oscillators as Analog Filters in Convolutional Neural Networks

<table>
<thead>
<tr>
<th>MNIST dataset</th>
<th>27x27x10000</th>
</tr>
</thead>
<tbody>
<tr>
<td>ONN-CNN</td>
<td></td>
</tr>
<tr>
<td>5 ONN filters + 59 CNN filters</td>
<td>3x3x64, stride = 2, padding = same</td>
</tr>
<tr>
<td>CNN 1</td>
<td>3x3x64, stride = 1, padding = same</td>
</tr>
<tr>
<td>Max Pool 1</td>
<td>2x2, stride = 2, padding = same</td>
</tr>
<tr>
<td>CNN 2 (x2)</td>
<td>3x3x128, stride = 1, padding = same</td>
</tr>
<tr>
<td>Max Pool 2</td>
<td>2x2, stride = 2, padding = same</td>
</tr>
<tr>
<td>CNN 3 (x2)</td>
<td>3x3x256, stride = 1, padding = same</td>
</tr>
<tr>
<td>Max Pool 3</td>
<td>2x2, stride = 2, padding = same</td>
</tr>
<tr>
<td>Fully connected 1</td>
<td>4096</td>
</tr>
<tr>
<td>Fully connected 2</td>
<td>1000</td>
</tr>
<tr>
<td>Fully connected 3</td>
<td>10</td>
</tr>
</tbody>
</table>

**Table 6.1** | Structure of the CNN under consideration. The convolutional neural network extracts simple features such as horizontal, vertical and diagonal edges in the first layer of convolutional filters. Subsequent layers extract more complex feature maps, until recognition is achieved.

**Figure 6.2** | **a)** Filters of the CNN first layer which select horizontal, vertical and diagonal edges on the feature maps. **b)** Original MNIST images, before undergoing the filtering operated by the ONN. **c)** A single ONN filter replaces five CNN filters. Different configurations of the output phase of the oscillators are linked to the recognition of a determined edge. In this way, the spurious patterns present in the ONN are harvested as additional information. **d)** The output of the ONN filter highlights the edges of the image. **e)** The image is expanded to the original size and post-processed to show the effectiveness of the feature edge extraction operation performed by the ONN. Figure reproduced from [288], licensed under CC-BY 4.0. Full terms: https://creativecommons.org/licenses/by/4.0/.
defies the maximum memory capacity for the network to perform exact recognition. Therefore, as expected, multiple spurious patterns arise. The edge information is however not corrupted by the spurious patterns, but can be retrieved by considering the phase relations between key oscillators, as depicted in figure 6.2. Moreover, the output information from the $3 \times 3$ filter is a $3 \times 3$ image encoded in the relative phase of the oscillators. The relative phase is computed with post-processing and the result of this operation is stored in the five filters corresponding to vertical, horizontal and diagonal edges. Of course, a complete realization of the network would need to include an hardware conversion of this information. This point will be addressed in the following sections.

Finally, the CNN and the ONN-CNN have a gap in recognition performance of around 2%. The reason for the worsening of the neural network performances is attributed to the differences between the ONN and the digital filter outputs and in particular to the occasional failure of the ONN to recognize the correct edge. To rectify this, we have developed a backpropagation algorithm which can act on the coupling weights of the ONN and which is discussed in details in the next session. Despite the reduction in recognition performances, the proposed CNN-ONN implementation could yield high technological advantages. In fact, as a single ONN is used to replace 5 digital filters, it allows for a reduction of the number of parameters that need to be trained by the network: 45 parameters undergo training for 5 CNN filters of $3 \times 3$ pixels size, however only 36 parameters need to be trained for a single ONN that performs all filtering actions. Assuming that the ONN is used to implement all the convolutional filters in the network, the number of parameters to be trained is reduced of 20%. This can represent an important advantage in terms of speed and power consumption when training larger networks. In addition, as in an ONN the processing of the five filters happens in parallel, whilst in the standard CNN these five convolution actions are performed sequentially, the number of accesses to the memory is proportionally reduced, therefore allowing in perspective a reduction of the training time and operation of the network.

### 6.2 Backpropagation Algorithm applied to the ONN

The training of a neural network is usually achieved in software through backpropagation, an algorithm for supervised learning which uses a gradient descend. Given the output classification vector $y$ of a neural network, and given vector $\hat{y}$ associated to the right labels of the dataset, the similarity of the two vectors is computed through the means of a cost function $C(y, \hat{y})$. As the classification vector $y$ of the network depends on the weights $w$ of all the layers of the network, the function $C(y, \hat{y})$ is also dependent on the values of the weights in each layer. The aim of the backpropagation algorithm is to minimize $C(y, \hat{y})$ respect to the weights $w$. This optimization is done calculating the derivative of $C(y, \hat{y})$ respect to the weights of each layer. The result of the calculation is then used to optimize the weights in this fashion:

$$w_{i+1} = w_i - \eta \cdot \frac{\partial C(y, \hat{y})}{\partial w_i} \quad (6.1)$$

88
6.2. Backpropagation Algorithm applied to the ONN

\[ \phi \propto w \cdot t_D \]  

where \( \phi \) is the ensemble of oscillator phases, \( t_D \) is the vector of input-time delays and \( w \) represent the weight matrix of the network. The backpropagation algorithm employed is schematized in figure 6.3 (a). The weight-updates \( \Delta w \) for the ONN layer are calculated from the derivative of the cost function according to this expression:

\[ \Delta w = \eta \sigma (\hat{y} - y) t_D^T \]  

where \( T \) denotes the vector transpose, \( \sigma(\hat{y} - y) \) is the derivative of the cost function and \( \eta \) is a learning parameter (typically \( 0 < \eta < 1 \)). The algorithm is tested on a edge recognition task where the \( 3 \times 3 \) ONN has previously failed. In figure 6.3 (b) the evolution of the phase error through eight learning epochs of the backpropagation algorithm is represented. Initially, the ONN converges to a wrong recognition of the image as a background feature. However, through
the backpropagation optimization of the weights, after training the correct feature retrieved. The capacity of the backpropagation algorithm to train the ONN into correctly recognizing the features has been tested on different patterns with progressively high gray-scale deviation from the stored pattern. Only relying on the training through Hebbian Learning Rule, when the search pattern differs of around 25% from the memorized pattern, the recognition gets inaccurate. However, after the training, the ONN is able to correctly recognize patterns with a 35% deviation from the stored patterns.

The extension of the backpropagation algorithm to the entire ONN-CNN is yet to be implemented, but is expected to boost the recognition performance, as well as to allow direct training of the ONN in the CNN system. The results of this work resulted in a patent application.

6.3 Phase-Detector Circuit for ONN Second Layer

Previously, the replacement of several filters in a convolutional neural network with an oscillatory neural network hardware platform has been discussed. In this section, we take a look to the architecture of an ONN system and we examine how to interface such a circuit with standard digital computational units. Moreover, neural networks are generally multi-layered structures: it is therefore important to address how the information contained in the phase of an ONN unit can be transferred to the next ONN layer. The discussion here presented regards some design ideas which were still not brought to practical realization and which should serve as an outlook for the complete design of an ONN accelerator platform.

Computing with oscillators presents the advantage that the information is encoded and processed in the timing of the signals, rather than in their amplitude, therefore the technology does not suffer from scaled supply voltages. However, the translation of a digital input in a time-delay and vice versa, a phase-output in a digital information, is not trivial. Lately, since time-mode signal processing is being researched for diverse applications, precise circuit implementations of digital-to-time converters (DTCs) as well as time-to-digital converters (TDCs) have been proposed [292].

The digital-to-time conversion of the information can be obtained by simple circuits, exploiting the time-delay of logic gates. The most straightforward implementation of a delay unit consists on a chain of inverters [293]. Similarly, the realization of a delayed signal controlled by the clock of a processor can be easily obtained through the design of shift registers and counters [294]. More complex implementations of DTCs have also been proposed in order to obtain very precise results [295, 296, 297]. The resolution of the DTC can pose a constrain on the maximum frequency of operation of CNNs. For processing an image with 256 gray-scale values, the input information needs to be encoded in a time delay with a least significant bit precision of \( \Delta t = \frac{1}{256} \frac{1}{f} \), where \( f \) is the frequency of the locked oscillators. If we assume that the minimum time-delay generated by the system is comparable with the clock-frequency of a processor, with the example of a Pentium 4 processor, which has a clock frequency of
around 4 GHz, the maximum oscillation frequency of the ONN would be limited to around 8 MHz. In reality, if the delay is generated by the propagation delay of a logic circuit, such as an inverter chain, the minimum delay can be significantly smaller. For example, in [297], a DTC with a least significant bit resolution of 22 ps is demonstrated, allowing for a theoretical realization of an ONN which operates at around 80 MHz.

Time-to-digital converter (TDC) circuits have been extensively studied and developed for high-precision phase detection in high-frequency phase locked loops (PLLs) applications. Many implementations, spanning from analog to all-digital TDCs have been proposed in literature and are commonly employed in frequency synthesizers, high frequency transmitters and receivers. Latest implementations of TDCs can reach a resolution of delay-detection in a signal of 6 ps [298, 299]. However, it is expected that this fine resolution will not be needed in application to ONN-CNN platforms, as the output information of a neural network, i.e. the classification vector, might not need a very fine level of accuracy, but, given the reduction and extrapolation of information, just the distinction of in-phase and out-of-phase signals.

As TDCs and DCTs circuits can be rather complex, area and power consuming circuits, it is not practical to envision a digital-to-time, followed by a time-to-digital conversion for each ONN layer in a multi-layer network. In the perspective of expanding the ONN filters to all the CNN layers, it is important to design a circuit which can transfer the phase information from a layer $i$ to the layer $i + 1$. In figure 6.4, a proposed circuit which is able to perform this task is presented. Figure 6.4 (a) shows the circuit schematic of the layer $i$ of a multi-layer ONN, while figure 6.4 (b) the connection circuit between layer $i$ and layer $i + 1$. The connection circuit is a phase detector based on a VO$_2$ device. It comprises a summing amplifier which detects the oscillation signals coming from the oscillatory units of the layer $i$. A second amplifier which sees a diode connected in negative feedback provides the non-linear activation of the layer. At the output, an oscillator unit is connected. The circuit acts as a majority gate: the output oscillator locks in phase with the majority of the input components that retain equal phase. This can be well understood when looking at the behavior of the connection circuit when these input combinations are applied: oscillator [1,5,9] (rising edge), oscillator [4,5,6] (horizontal edge), oscillator [2,5,8] (vertical edge) and oscillator [3,5,7] (falling edge). In the example related to the waveforms (c-f) in figure 6.4, the first ONN layer is synchronized on a falling edge, therefore oscillators [1,5,9] present a 180$^\circ$ phase compared to a reference oscillator, while the others present a 0$^\circ$ phase. The connection circuit for the falling edge detection sees all three oscillators at 180$^\circ$, and its output oscillator is also oscillating at the same phase. The connection circuits for rising, horizontal and vertical edge see two oscillator phases at 0$^\circ$ and one oscillator phase at 180$^\circ$; the output oscillators for these circuits stabilize at a phase of 0$^\circ$. The connection circuit is ultimately able to control the phase of the next stage oscillators depending on the phase of the previous stage. The discussed circuit implementation resulted in a patent application.
Figure 6.4 | a) Schematic of one ONN layer. b) Schematic of the connection circuit between two layers c-f) Waveforms linking the phase of the connection circuit output oscillator with the feature extracted by the previous ONN layer.
6.4 Benchmark

A benchmark of the ONN accelerator compared to conventional CPUs and GPUs in performing the convolution operation was proposed in the publication [288], licenced under CC-BY and it is here reproduced. The analysis has been conducted not considering the peripheral circuitry (TDCs, DTCs and connection layers) that the ONN system will require, therefore could not be benchmarked against other accelerator for neural networks, where a complete system comparison should be performed. This discussion should be taken as indication of the potentiality of the ONN technology.

For the comparison, we assume the extension of the ONN filters to the entire first layer of the CNN presented in 6.1. The first layer of the CNN consists of 64 filters of $3 \times 3$ dimension passing through a $27 \times 27$ pixel image with a stride of 2, accounting to total of $13 \times 13$ operations per filter. Assuming multiple ONN filters working in parallel, and assuming that each ONN can perform 5 filtering actions inherently, a total amount of $13 \times 13 \times 64/5 = 2200$ ONNs is required. This corresponds roughly to 20'000 oscillator units, coupled through 80'000 memristors. For a $\text{VO}_2$ and memristor device dimension of 100 nm $\times$ 100 nm, the total estimated area for the ONN layer would be around 0.002 mm$^2$. The power consumption of the convolutional layer was calculated referring to the demonstration offered in [300] and in section 5.1, where the oscillators were shown to operate at a power $P = 20 \mu W$ with a scaled supply voltage $V_{DD} < 1$ V and $f = 3$ MHz frequency operation. The total energy for the ONN to process one image with 64 filters at 3 MHz, including a waiting time of 5 oscillating periods for the output stabilization, is calculated as:

$$P \cdot f \cdot 5 = 0.6 \mu J/frame$$

The total energy consumption of the coupling memristor in estimated to be $P = 3.4 \mu J/frame$, when the average memristor value is $R = 100 \text{k}\Omega$ with a voltage drop of $V = 0.7$ V. As a perspective, it is expected that a $\text{VO}_2$ oscillator can be driven with $1 \mu W @ 0.3$ V and 20 MHz, upon scaling of the device dimensions. Moreover, through improvement of the device uniformity the coupling strength could be weakened, allowing 1 M$\Omega$ coupling resistance [196]. This would allow to reduce the energy consumption of the ONN system to 3 nJ/frame.

The same convolution of the CNN first layer filters, when operated on a GPU, requires to perform $(13 \times 13)$ convolutions $\times$ 64 filters $\times$ $(3 \times 3)$ pixels/filter = 97344 multiply-accumulation operations, that correspond to around 200'000 flops. In Intel’s CPU Core i9, which runs 1 TFLOP/s at 95 W, the total energy accounts for 20 $\mu J/frame$; in the NVIDIA Tesla V100 GPU, that operates 120 TFLOP/s @ 300W, the total energy is 500 nJ/frame. The ONN system with the current device specification can already operate at a lower power consumption compared to a conventional CPU. Moreover, through scaling and device optimization, it is expected to outperform the top GPU available on the market. Experimental demonstration of a complete ONN system is however required to validate this prediction.
6.5 Main Achievements

In this chapter, we developed a method to employ coupled-oscillators systems as hardware accelerators for convolutional neural networks. In summary:

- With the example of classification of handwritten digits, we successfully demonstrated that oscillatory neural network hardware can be employed to accelerate the convolution operation in convolutional neural networks.

- We demonstrated that a single ONN filter can replace up to five digital convolutional filters in a CNN.

- A backpropagation algorithm which can act on the ONN hardware was developed.

- A connection circuit which allows for multi-layer ONN structures was designed.

- This novel approach allows in perspective to reduce up to 20% the number of trainable parameters in CNNs, without reducing the number of filters employed by the recognition algorithm, therefore improving speed and energy efficiency without lowering the recognition performances.
The research towards the development of a neuromorphic computing hardware as accelerator for neural network applications is motivated by the need of overcoming von Neumann's bottleneck. The strategy followed by this research, similarly to what is done with non-volatile memories for vector-matrix multiplication acceleration, is to find suitable solutions to perform in-memory computing. Oscillatory neural networks (ONNs) are known for their associative memory capabilities, which can be employed to retrieve from distorted or incomplete data the information stored in the memory of the system. Traditionally, these networks have been used for pattern recognition applications. Hardware implementations of oscillatory neural networks have recently gained momentum since the research of novel, non-linear electrical devices allowed for the design of compact, energy efficient oscillators. However, until now the practical realization of an ONN hardware was hindered by the need of scaling the network to large dimensions, in order to achieve the landscape of functionalities required for the technology to be competitive. Being ONNs fully-connected networks, the experimental demonstration of a large ONN system faces the challenge of implementing a high number of interconnections between the oscillatory units. Moreover, mismatch between the electrical components have proved to be detrimental for the frequency locking of the oscillators and for the computational precision of the system.

The main innovation presented in this work regards the development of a hardware platform that exploits small networks of coupled oscillators, limited to a $3 \times 3$ matrix size, which are designed to accelerate the filtering operations computed in convolutional neural networks (CNNs) algorithms. With the approach we propose, we eliminate the problems connected to the realization of large ONN circuits, promoting instead a modular approach to computation, which envisions several smaller networks working in parallel to extract specific data features. With the case example of image classification operated by a convolutional neural networks, we propose ONNs as an hardware implementation of the convolutional filters present in the algorithms. Through the exploitation of the associative memory capabilities of ONNs, it is possible to use a single ONN unit to replace multiple digital filters of the CNN, therefore developing an hardware platform capable of reducing the training parameters and consequently
Chapter 7. Conclusion and Outlook

boost the speed and power performance of the algorithm. In the work presented in this thesis, the foundation for the development of this technology is laid. A comprehensive research was conducted, to address the possibilities as well as the challenges faced by the technology from the device realization, to the implementation of the circuits and the integration of the platform with industry-standard algorithms. A summary of the principal achievements of this research is given in figure 7.1.

For the industrial development of ONNs, it is necessary to identify a technique which allows to fabricate compact, scalable oscillators. Therefore, an important part of the research presented in this thesis focused on the development of fabrication methods to integrate the metal-insulator transition material VO$_2$ on silicon with a CMOS compatible process. In Chapter 3 we discussed an exploration of deposition and annealing methods, including the first demonstration of grain size tuning in VO$_2$ on SiO$_2$/Si through the employment of a flash lamp anneal technique. The results of the material investigation have been presented in Chapter 4, where we analysed the phase transition characteristics of nano-devices realized in three different geometries: planar devices, crossbar devices and single grain devices. As the material is conventionally deposited on lattice-matched substrates, the first step in our research was to identify the impact of the granularity of the material, which arises when the VO$_2$ is deposited on SiO$_2$/Si, on its transition characteristics. In particular, we observed a multi-step phase transition characteristic in the planar and the crossbar devices, connected to the consecutive switching of different domains inside the VO$_2$ layer. We discussed experimental evidence suggesting that the multi-step switching behavior originates from the subsequent phase transition of single grains in the device. To further characterize this behavior, we employed a state-of-the-art scanning thermal microscopy technique to derive thermal maps of electrically-activated planar devices. This characterization method allowed to identify the metallic path that is formed in a planar device after the IMT transition and to study the evolution of the filament upon increasing the electrical power supply. The formation of a current path, as well as the subsequent switching of multiple grains inside the device, contribute to the high device-to-device variability recorded for planar devices. With the introduction of the crossbar design, which comprises a single layer of parallel grains between the two metal contacts and eliminates grain boundaries in the current path, we achieved a reduction of the device-to-device variability and higher yield. Finally, we explored the possibility of obtaining a single, sharp phase transition of VO$_2$ on silicon with the processing of single-grain nanoparticles. The nanoparticles were fabricated through solid state dewetting of VO$_2$ films on the SiO$_2$/Si substrate. With this design we offered a first demonstration of single-grain VO$_2$ devices on a silicon platform, which retain a single, point-sharp phase transition. The single-grain switches represent an important advance in controlling the insulator-to-metal phase change of VO$_2$ on silicon.

In Chapter 5 we investigated the realization of compact, power efficient oscillators based on the phase transition of VO$_2$ devices. The best-performing oscillators operated at a scaled voltage supply $V_{DD} < 1 \, \text{V}$, with low power consumption of $P = 20 \, \mu \text{W}$. The maximum frequency of operation of the device was measured to be $f = 2 \, \text{MHz}$, limited by electrical time constant.
**Figure 7.1** | Summary of the principal results discussed in this thesis. Top: planar and crossbar VO$_2$ devices were fabricated and characterized; through contacting of a single-grain device a sharp phase transition is achieved. Middle: the connection of single VO$_2$ oscillators is explored; different phase patterns are memorized and recognized in experimental demonstrations of ONNs. Bottom: one ONN filter is used to replace up to five digital filters in a CNN; the training of the ONN with a backpropagation algorithm and a circuit implementation of subsequent ONN layers are discussed. As an outlook for this technology, the ONN filters should be expanded to all the CNN layers and trained with the backpropagation algorithm to obtain image classification.
Chapter 7. Conclusion and Outlook

associated with the parasitic capacitance introduced by the measurement setup.

The realization of oscillatory neural network circuits was investigated through the frequency-locking of VO\textsubscript{2} oscillators with resistive and capacitive elements. Compared to other approaches, which encode the input information in the amplitude of the driving voltages of the circuit, the ONN design we propose relies on the encoding of the information in the relative timing of the signals. This approach avoids the noise and uncertainty on the least significant bits compared to the most significant bits which is present in amplitude-encoded technologies and it is therefore resilient to scaled voltage power supplies. Through experiments and simulations, we investigated the storage of the information in the output phase difference of the oscillators. In our design, the coupling elements between the oscillators represent the memory of the system. Our study reveals that while a purely-resistive coupling allows the encoding of either the out-of-phase or the in-phase configuration in the system, a hybrid capacitive-resistive coupling enables the storage of intermediate phase-relations, therefore increasing the information capacity of the system. We conducted three- and four-coupled oscillators experiments, demonstrating the pattern-matching capabilities of the ONN design. In particular, the reported experiments show a first demonstration of the effectiveness of the information encoding in the timing of the input voltage signals; with this processing scheme, a physical implementation of pattern recognition with ONNs was implemented. Moreover, with four, fully-coupled VO\textsubscript{2} oscillators we demonstrated in experiments and simulations the storage of up to five patterns in the phase relations of the ONN outputs. The ONN unit is able to discriminate in a single comparison the five patterns from distorted inputs, without the need of being reconfigured. This was applied to MNIST dataset images for a feature edge extraction operation, in which horizontal, vertical, diagonal features and the background can be discriminated.

The exploitation of the ONN pattern recognition capabilities for accelerating the computation of industrial-standard convolutional neural networks have been presented in Chapter 6. Through simulation, a 3 × 3 oscillatory neural network was integrated in the computational flow of a CNN algorithm. We demonstrated that a single ONN unit can replace up to five convolutional neural network filters, recognizing in with one filtering action horizontal, vertical, diagonal edges and the background of an image. The resulting CNN-ONN was tested on the MNIST dataset, were it achieved 95% of recognition accuracy. In table 7.2 all the image processing architecture proposed so far with ONNs have been summarized, together with the contribution that this work brings to the field. In particular, the ONN design for feature edge extraction operation was here employed for the first time as a filter in convolutional neural networks. From a system perspective, a backpropagation algorithm tailored on the ONN was designed, with the aim of implementing the training of the ONN alongside the other layers in the CNN, as well as to improve its recognition performances. Moreover, the design of a circuit to cascade ONN filters to form multi-layered networks is presented.

Ultimately, the research presented in this thesis aimed to produce an evaluation of the potentiality of VO\textsubscript{2}-based ONN technology starting from the realization of compact devices up to
an initial exploration the integration of the ONN platform to support neural networks computations. The device and circuit research helped to gain a clear vision of the computational capabilities offered by this technology, as well as the challenges which remain open for future research to address.

**Table 7.2** | The summary of the techniques used to filter and process images with coupled oscillators technology presented in table 2.13 is here re-proposed with the addition of the contribution of this thesis to the field. The information from [196, 216, 217].

<table>
<thead>
<tr>
<th>Technique</th>
<th>Input</th>
<th>Output</th>
<th>Function</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Shukla et al.</td>
<td>Gate voltage</td>
<td>Amplitude of the voltage signal</td>
<td>Edge and saliency detection</td>
<td>2-pixel comparison, 8 comparison per pixel</td>
</tr>
<tr>
<td>b) Cotter et al.</td>
<td>Frequency deviation</td>
<td>Synchronization time</td>
<td>Edge and saliency detection</td>
<td>9-pixel comparison with reference edge; one comparison per each edge needed</td>
</tr>
<tr>
<td>c) Tsai et al.</td>
<td>Gate voltage / modulated resistance</td>
<td>Amplitude of the voltage signal</td>
<td>Edge and saliency detection, color detection, 9-pixel comparison with reference pattern</td>
<td>one comparison per each function needed; programmable function; edge detection through the calculation of the deviation of 3 neighboring pixels in each direction.</td>
</tr>
<tr>
<td>This work</td>
<td>Time delay of input voltage</td>
<td>Relative phase of oscillation</td>
<td>4-pixel edge detection, CNN filter</td>
<td>all edges recognized with one comparison, programmable function via coupling tuning</td>
</tr>
</tbody>
</table>

### 7.1 Future Directions

In this work, we presented the development of the ONN technology from the device to the system architecture. Although we strongly believe that employing oscillatory technology for neural networks computation can bring decisive advantages in terms of reducing hardware resources, decrease the power consumption and increase the speed of neural networks training and inference, future work is needed to further assess the potential of this neuro-inspired technology respect to other approaches.

From the device point of view, further progress is required to reduce the VO$_2$ device-to-device variability that so far hinders the scaling of the oscillatory neural network to more than four-coupled units. As we identified that the device-to-device uniformity suffers from multi-steps switching in the devices, we believe that the realization of single-grain devices can bring the technology to the performances needed to expand the networks. To this aim, further research
is needed to localize and control the dimensions of the VO$_2$ single-grain nanoparticles. Other important figure of merits should be addressed with tailored experiments. In particular, the maximum oscillation frequency could be investigated by integrating all the oscillator components on a chip, therefore eliminating the parasitic elements that are yet responsible for the limitation in the speed of the oscillators. The scaling of the VO$_2$ devices to lower dimensions should also be considered. As a single grain device was contacted in this thesis, resulting in promising electrical characteristics, we believe that the device dimensions can be scaled down to 10 nm size or less without apparent limitation to the switching functionalities. From an energy perspective, narrow hysteresis curves in the VO$_2$ R-T characteristic are desirable, to bring the IMT and MIT voltage thresholds close together, resulting in a lower swing of the oscillations. As discussed in this thesis, VO$_2$ single grain devices presented a larger hysteresis compared to the multi-grain switches. This phenomenon is explained in literature by the difficulties in nucleating the transition when the crystal quality of the nanoparticle is high and a lower number of defects is present. The widening of the hysteresis could potentially hinder the scalability of the device to a single grain of nanometer dimensions; however, artificial creation of defects, such as with the introduction of dopants, might represent a solution to control the hysteresis width of the device. The introduction of dopants in the VO$_2$ film (for example Germanium) should anyway be considered to increase the phase transition temperature of the device up to 100° C [145, 147], allowing the system to function in the temperature range required for commercial applications.

As the device uniformity improves, experimental realizations of 3×3 ONN image filters such as the ones discussed in simulations should be realized. Similarly, the demonstration of larger ONN systems could be tackled. As the ONN presented in this thesis is a fully-connected network, the scaling of the network to comprise a high number of oscillators poses a challenge in the realization of a large coupling matrix, which would represent the bottleneck in the area consumption of the ASIC. Moreover, given the high sensibility of the ONN performance on device-to-device variations, the implementation of very large ONN systems might result challenging to realize. We therefore believe that the approach here proposed to use multiple, parallel ONN filters with a constrained number of nodes yields a most favorable exploitation of the associative memory of this system. Ultimately, the physical realization of a complete ONN system is needed to assess the overall performances of the architecture and to compare it with state-of-the-art specialized neural network accelerators. The next steps for this research envision the expansion of the ONN as convolutional filters across the entire convolutional neural network architecture. The training of the network should be supported by introducing reconfigurable devices as coupling elements of the network, for example employing programmable memory units, such as phase change memories or resistive RAMs.

The Phase-Change Switch project financed by the Horizon 2020 helped us in the past years to research and ultimately achieve the realization of scaled VO$_2$ devices integrated on silicon. In the framework of the H2020 project NeurONN, we are working towards demonstrating a re-configurable ONN architecture integrated with 2D memristor devices, which will allow the learning and acceleration of neural network algorithms.
A.1 SThM Characterization

In this Appendix the results of the SThM measurements on VO$_2$ devices are discussed in more details.

As reported in Section 4.1.1, thermal maps conducted on a ALD Flash Annealed device reveal the nature of the multiple-step I-V characteristics of the VO$_2$ planar devices. The device under test presented multiple steps in the I-V characteristic, which is depicted in figure A.1 a. At the first activation, the device presented multiple phase transitions which were investigated with the SThM technique. However, after stressing the device with very high applied voltage, the I-V characteristic and the resistance change smoothed, as shown in figure A.1 b. The voltage value of biased used for the thermal maps presented in figure 4.4 are referenced to the device I-V after the measurements.

As already discussed in section 4.1.1, the multiple phase transitions of the device are connected to the expansion of the metallic filament that is formed following the first phase transition. This is highlighted by figure A.2 a, in which the difference in the thermal signal between the device activated at 0.8 V and the device activated at 1.8 V is shown. The brighter parts of the graph highlight a region in which the temperature increased by applying a higher voltage; darker region represents equal or slightly decreased temperature in the measurement conducted at 1.8 V. From this graph the expansion of the filament results evident. In particular, the seven bright zones, highlighted by the numbers, are identified as grains which undergo the phase transition upon rising the voltage, therefore bringing to an expansion of the metallic domain and to the evolution of the current path.

The analysis of the temperature profile also reveals the boundaries of the metallic phase in the switched device. Local temperature maxima are indicative of local heat sources. Assuming a reasonably homogeneous film and spreading geometry the temperature profile of the region around the maximum where Joule heating is present is concave. For the areas where no heat sources are present, but only thermal conduction along the film and conduction into the
Appendix A. Appendix

Figure A.1 | IV characteristic of the VO$_2$ for which the filament expansion was investigated. a) Originally the device presented multiple, defined phase transitions, which can here be identified in the steps present in the I-V curve. The device was biased at the voltages corresponding to the states before and after the steps in the I-V curves and for each bias point a thermal scan was performed. b) After multiple measurements, the device I-V degraded to a smoother curve, possibly because of high voltages used to activate the device, which led to a new irreversible change. The thermal maps discussed in the main text refer to this current-voltage characteristic.

Figure A.2 | Plot of the difference in thermal signal between the map taken at 0.8 V and the one at 1.8 V. An increase in temperature is highlighted by a brighter color; darker colors indicate areas where the temperature remained constant or slightly decreased. The points (1,2) refer to a shift of the thermal peak of the device. The areas (3,4) identify the grains which switched from the insulating state at 0.8 V to the metallic state at 1.8 V of bias. With (5,6,7) we highlight the formation of a second current filament in the device.
A.2 Simulation of Multi-Grain Switching in VO$_2$ Devices

In the following, further details on the simulation presented in section 4.1.2 are discussed. In particular, table A.4 presents the values of the parameters used for the simulations presented in 4.5. Referring to equation 4.1, the temperature $T$ is calculated for each grain. The volume and volumetric heat capacitance are constant and referred to each grain; the differential step of the Laplace operator is equal to the grain diameter. The parameter $k$ refers to the VO$_2$ thermal conductivity, whose value can be found in literature ($k = 8 \frac{J}{mK}$) [301]. However, as in our model the VO$_2$ is not treated as an homogeneous film, but the heat equation is treated separately for each grain, the thermal conductivity per unit thickness $k$ is substituted by the thermal conductivity of the grain boundary interface $k_{gb} = 10^7 - 10^8 \frac{W}{m^2K}$, multiplied by the grain diameter. The value of the grain boundary interface, as well as the value of the grain boundary resistance ($\approx 10xR_{MET}$) was calibrated in the simulations to reproduce the R-T and I-V characteristics of the experimental devices. The thermal conductivity of the SiO$_2$ substrate was derived from the SThM measurements as $g = 10^8 \frac{W}{m^2K}$, value which is in agreement with what reported in literature [301]. The substrate temperature was fixed at 21°C. The value of the volumetric heat capacitance was also taken from literature [301].
Appendix A. Appendix

Figure A.3 | Thermal map obtained from a device annealed with FA at 300°C, 90 J/cm², 20 mbar oxygen pressure. a) I-V curve of the device. c) Temperature map of the device obtained with a bias of $V_{IN} = 3$ V ± 0.4 V @ 1.2 kHz. b) and f) 1D temperature profile along the black lines shown in panel c. e) The inset shows the temperature profile across the grain boundaries of the device, which result as heated spots in the measurements. This effect could be indicating that the grain boundaries represent high resistances across the current path; however, as AFM is an edge-sensitive technique, a further investigation is needed to rule out the possibility of this effect being an artifact of the measurements.
### A.2. Simulation of Multi-Grain Switching in VO$_2$ Devices

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grain diameter</td>
<td>50 nm</td>
</tr>
<tr>
<td>Insulating resistivity</td>
<td>0.5 $\Omega$ cm</td>
</tr>
<tr>
<td>Metallic resistivity</td>
<td>0.05 $\Omega$ cm</td>
</tr>
<tr>
<td>Standard deviation for the resistivity</td>
<td>5%</td>
</tr>
<tr>
<td>Series resistance</td>
<td>15 k$\Omega$</td>
</tr>
<tr>
<td>Volumetric heat capacity</td>
<td>$1.5 \times 10^6 \frac{J}{m^3 K}$</td>
</tr>
<tr>
<td>Substrate thermal conductivity</td>
<td>$10^8 \frac{W}{m^2 K}$</td>
</tr>
<tr>
<td>Grain boundaries thermal conductivity</td>
<td>$10^7 \frac{W}{m^2 K}$</td>
</tr>
<tr>
<td>Integration time</td>
<td>7 ns</td>
</tr>
<tr>
<td>Temperature for the IMT</td>
<td>67°F</td>
</tr>
<tr>
<td>Temperature for the MIT</td>
<td>57°F</td>
</tr>
<tr>
<td>Standard deviation for the transition temperatures</td>
<td>5°F</td>
</tr>
<tr>
<td>From $dR_{\text{INS}}/dT$</td>
<td>1 k$\Omega$/K</td>
</tr>
</tbody>
</table>

**Table A.4** | Values used for the simulation presented in section 4.1.2.
References


References


References


References


References


References


References


References


References


References


References


References


References


References


References


Curriculum Vitae

Personal Data

Name               Elisabetta Corti
Date of Birth      21.09.1993
Place of Birth     Como, Italia
Citizen of         Italia

Education

2017 – 2021  École polytechnique fédérale de Lausanne, Switzerland
             PhD candidate, Electrical Engineering
2015 – 2017  Politecnico di Milano, Italia
             Master of Science, Electronics Engineering, 110 "cum Laude"
2012 – 2015  Politecnico di Milano, Italia
             Bachelor of Science, Electronics Engineering, 110 "cum Laude"

Professional Experience

2017 – 2020  IBM Research – Zurich, Switzerland
             Predoctoral Scientist, Science & Technology Department
03 – 09/2017 imec, Belgium
             Research Intern, Master thesis project

Activities and Certificates

2018 – 2020  Peer-Review, upon invitation for various scientific journals
Curriculum Vitae

2015 – 2017  **Idea League Challenge Programme**  
Exchange Program on business, economics and engineering, with modules in RWTH Aachen, Politecnico di Milano, Chalmers University, TU Delft and ETH Zurich

07/2017  **SENG**, Summer Camp for Elite Students, Hong Kong University of technology

03/2015  **ATHENS Exchange Program**, Warsaw University of Technology, Poland  
Electromagnetic Waves Engineering and Applications

Awards and Recognitions

09/2018  **First Patent Achievement Award**, IBM Research, Zurich, Switzerland

2015 and 2017  **Premio di Studio Giovanni Zampese** (2x), Scholarship, Cantu, Italia

2012 – 2017  **Merit-Based Scholarships**, Politecnico di Milano (3x), Italia

Conference Presentations

(Journal Publications are listed in List of Publications)

12/2020  **Material Research Society Autumn Meeting**, Boston (online), USA  
Scaled VO$_2$ Oscillators for Neural Network Applications, **talk**

05/2019  **Material and Process Graduate Symposium**, Zurich (online), Switzerland  
Grain-size Tuning of VO$_2$ Films on Si using Millisecond Flash Annealing, **poster**

11/2019  **Semicon Europa 2019**, Munich, Germany  
**Invented talk**, Advances in Energy Efficient Neuromorphic Computing: Ready for Artificial Intelligence at the Edge?

06/2018  **Swiss Physical Society Meeting**, Zürich, Switzerland  
Neuromorphic Computing with Coupled VO$_2$ Oscillators, **poster**

03/2019  **IEEE EUROSIP-ULIS International Conference**, Grenoble, France  
VO$_2$ Oscillators Coupling for Neuromorphic Computation, **talk**

11/2018  **IEEE International Conference on Rebooting Computing**, Washington, USA  
Resistive Coupled VO$_2$ Oscillators for Image Recognition, **talk**