UCLouvain
Institute for Information
and Communication Technologies,
Electronics and Applied Mathematics

Potential of Vanadium Dioxide for Artificial Sensory Neurons

Noémie BIDOUL

Thesis submitted in partial fulfillment
of the requirements for the degree of
Docteur en sciences de l'ingénieur et technologie

Dissertation committee:

Prof. Denis FLANDRE	*Supervisor*	UCLouvain, Belgium
Prof. Laurent FRANCIS	*President*	UCLouvain, Belgium
Prof. Jean-Pierre RASKIN	*Secretary*	UCLouvain, Belgium
Dr. Léopold VAN BRANDT		UCLouvain, Belgium
Prof. Elisabetta CHICCA		RUG, Netherlands
Prof. Adrian M. IONESCU		EPFL, Switzerland

August 2024

Contents

Contents . . . . . **i**

Abstract . . . . . **v**

Author's publication list . . . . . **vii**

List of acronyms . . . . . **ix**

Acknowledgments . . . . . **xi**

Introduction **1**

1 The landscape of neuromorphic hardware **7**
1.1 Neuromorphic processing systems 8
1.1.1 Artificial Neural Networks 8
1.1.2 Spiking Neural Networks (SNNs) 10
1.2 Neuromorphic sensors 16
1.2.1 Sensitive synapses 16
1.2.2 Sensitive neurons 17
1.3 Vanadium dioxide for neuromorphic applications 24
1.3.1 Transition properties 25
1.3.2 Resistive-switching VO_2 devices 31
1.3.3 VO_2 neuromorphic sensors 34
1.4 Key messages 38

2 Understanding the stochasticity of resistive switching in VO_2 . **41**
2.1 VO_2 thin film synthesis 42
2.1.1 Choice of substrate 42

2.1.2 Deposition technique 43
2.1.3 Impact of oxygen concentration 45
2.1.4 Impact of annealing temperature 51
2.2 Two-terminal device fabrication 53
2.2.1 Process flow 53
2.2.2 Material characterization of small- versus large-grains films . . 56
2.2.3 Two-terminal devices electrical characteristics 58
2.3 Tuning the stochasticity of resistive switching 59
2.3.1 On the origin of stochasticity 60
2.3.2 The impact of grain size 61
2.3.3 The impact of channel length 64
2.3.4 Discussion 65
2.4 Key messages 71

3 Temperature-sensitive stochastic spiking neurons: different approaches to encoding 73
3.1 Stochastic spiking neurons 75
3.1.1 The negative differential resistance condition 75
3.1.2 Circuit interface 77
3.1.3 Analytical dynamic model 79
3.1.4 Dynamic stochastic analysis 81
3.2 Rate-based temperature encoding 85
3.2.1 Sensitivity and Limit of Detection 85
3.2.2 Extended temperature and frequency operation range 90
3.2.3 Benchmarking 93
3.3 Stochastic bursting temperature encoding 98
3.3.1 Conditions for bursting 98
3.3.2 Temperature-dependent bursting 100
3.3.3 Burst metrics definition 101
3.3.4 Burst metrics statistics 103
3.3.5 Burst metrics temperature dependence 105
3.4 Key messages 107

4 Assessing the environmental impact of VO_2 neuromorphic hardware 109
4.1 Basics of LCA 111
4.1.1 The four stages of LCA 111
4.1.2 LCA of AI and neuromorphic technologies 114
4.2 LCA methodology 115

4.2.1 Goal and scope . . . 115
4.2.2 Inventory . . . 116
4.2.3 Impact assessment . . . 119
4.3 LCA results . . . 120
4.4 Interpretation . . . 122
4.4.1 Uncertainty & scenario analysis . . . 122
4.4.2 Eco-design perspectives . . . 125
4.4.3 Comparison with CMOS-based neurons . . . 127
4.5 Key messages . . . 130

Conclusions and Perspectives . . . **133**

Bibliography . . . **145**

Appendix . . . **I**
A Appendix A . . . I
B Appendix B . . . IV

Abstract

ENGINEERS have often drawn inspiration from biology in their quest for more efficient systems. In the realm of micro-electronics and computer science, such design practices have resulted in the emergence of the neuromorphic engineering research field, which aims to design electronic hardware and software that replicates the computing paradigms of the brain. Indeed, the human brain can perform complex processing, inference, and generation tasks while relying on an inherently noisy and imperfect hardware, with a power consumption equivalent to that of a few light bulbs. The quest towards replicating the brain computing software can be tackled through several angles. The first would be a top-down approach, where bio-inspired machine learning algorithms are run on conventional CMOS-based Von Neumann processor architectures. An alternative is a bottom-up approach in which novel hardware primitives are designed to emulate the behavior of individual neurons. In such approach, rather than being designed independently from the computing platform, the algorithm is defined by the neuronal architecture itself. Like in biological neurons, information is encoded in the timing of electrical spikes produced by these emerging devices. A myriad of implementations of computational neurons and synapses have popped up over the last decade: some analog or digital, based on conventional CMOS processes or using novel materials. What is still missing is a bio-inspired interface with the real world: that is, devices that would be able to emulate the sensory neurons that convert information such as temperature, light, or sound into electrical spikes, in other to feed neuromorphic processing systems.

In this thesis, we propose a specific implementation of such sensory spiking neurons based on vanadium dioxide (VO_2), a phase-change material. VO_2 possesses the peculiar property of transitioning from a semiconductor to a metallic phase around a temperature 68 °C. Such phenomenon can be leveraged to design devices which demonstrate a reversible transition from high to low resistive states, allowing to emulate the electrical spikes of neurons. Akin to phenomena taking place in biological neurons, such resistive switching is intrinsically stochastic, while the physical mechanisms behind these variations are not yet fully identified. Depending on the neuron use case, one might desire to minimize or harness such stochasticity. However, due to the poor understanding of the phenomenon, there is a lack of guidelines to design devices tailored for both cases. In addition, most of the research has been focused on designing computational rather than sensory VO_2 neurons. The phase transition in VO_2 can be triggered by a wide

variety of stimuli, including temperature, light, and deformation. This opens the possibility of performing stimulu transduction directly within the neuron, and to simultaneously encode it into the timing of spikes. Finally, although neuromorphic technologies are often put forward as a lever to perform energy-efficient information processing and possibly to curb the environmental footprint related to the use phase of large AI models, there is little research on the environmental impact of the fabrication of such emerging devices.

This work attempts to bridge these gaps, starting from the vanadium dioxide material properties and building up towards application level. Chapter 1 introduces the rich landscape of neuromorphic computing and maps VO_2 neurons among other existing implementations, highlighting their peculiar properties and discussing their benefits compared to other solutions. Chapter 2 presents the synthesis of polycrystalline VO_2 thin films and the design of micro-resistors exhibiting resistive switching. It then studies how specific material properties and device design parameters can affect the resulting neuron stochastic behavior. Our findings suggest that the number of grains that compose a VO_2 neuron has a significant impact on its stochasticity, and that the latter can thus be impacted by the film grain size and device dimensions. Our observations lead to significant improvements of the present understanding of stochasticity in resistive-switching devices. Chapter 3 demonstrates how the VO_2 micro-resistors can be used within a simple circuit composed of one transistor and one capacitor to produce electrical spikes. We discuss how the intrinsic cycle-to-cycle variations of the device I-V characteristics, studied in detail in a static approach in Chapter 2, result in a dynamic cycle-to-cycle variations in the spike duration. Static and dynamic stochasticities are bridged through an analytical dynamic model, which is verified by both static and dynamic empirical data and allows us to determine the dominant causes of spike rate stochasticity. We then take the use case of temperature as specific stimulus, to explore different ways of translating information into the timing of spikes. Through a more classical approach, we demonstrate encoding of the temperature into the spike rate of VO_2 neuron. We investigate how the dynamic stochasticity impedes the Limit of Detection (LoD) in such scenario. We then propose a novel, more bio-plausible and energy-efficient encoding technique by biasing the VO_2 neuron at the edge of its stable spiking regime, yielding sparse stochastic spikes bursts. We study the specific burst metrics in which the stimulus is encoded, as well as the statistical distribution of these bursts and underlying physical phenomena. Finally, Chapter 4 takes a step back to discuss the possible environmental impacts of such technological development. We quantify the direct environmental impact of the device production through a cradle-to-gate Life Cycle Assessment, identifying critical process steps that dominate indicators such as primary energy demand and greenhouse gasses emissions. We propose eco-design routes to minimize such footprint and compare this footprint to that of alternative CMOS-based implementations.

Author's publication list

Related journal papers

JP1. N. Bidoul, N. Roisin and D. Flandre, "Tuning the Intrinsic Stochasticity of Resistive Switching in VO2 Microresistors", *Nano Letters*, vol. 24, no. 21, pp 6201-6209, May. 2024. DOI: 10.1021/acs.nanolett.4c00184

JP2. N. Bidoul, L. Van Brandt and D. Flandre, "Bio-inspired Encoding of Heat Using VO2 Neuron Operated in Stochastic Bursting Regime", *IEEE Electron Device Letters*, submitted.

Related conference papers

CP1. N. Bidoul, T. Rosca, A. M. Ionescu and D. Flandre, "Static and Dynamic Stochastic Analysis of a Temperature-Sensitive VO2 Spiking Neuron", in *IEEE 53rd European Solid-State Device Research Conference (ESSDERC)*, Sept. 2023, pp 81-84. DOI: 10.1109/ESSDERC59256.2023.10268509

CP2. N. Bidoul, P. Raux, T. Van Aerschot, A. Pip, C. Renaux, S. Faniel, D. Flandre and J.-P. Raskin, "Process-based Life Cycle Assessment of a Vanadium Dioxide Spiking Neuron", in *Electronics Goes Green (EGG)*, Jun. 2024. *Best Paper Award*. DOI: 10.23919/EGG62010.2024.10631188

CP3. F. Roth, N. Bidoul, T. Rosca, M. Dorpinghaus, D. Flandre, A. M. Ionescu and G. Fettweis, "Spike-Based Sensing and Communication for Highly Energy-Efficient Sensor Edge Nodes", in *IEEE International Online Symposium on Joint Communications & Sensing (JC& S)*, Mar. 2022. DOI: 10.1109/JCS54387.2022.9743501

Related conference oral presentations

CO1. N. Bidoul, and D. Flandre, "Bio-inspired Encoding of Heat Using VO2 Neuron Operated in Stochastic Bursting Regime", in *Neuronics*, Feb. 2024. DOI: 10.29363/nanoge.neuronics.2024.014

CO2. N. Bidoul, B. Huet, F. Urena Begara, J.-P. Raskin, and D. Flandre, "Tuning the stochasticity of VO2 neurons firing-threshold through grain size engineering", in *Neuromorphic Materials, Devices, Circuits and Systems (NeuMatDeCas)*, Jan. 2023. DOI: 10.29363/nanoge.neumatdecas.2023.050.2024.014

Related conference posters

CPO1. N. Bidoul, B. Huet, F. Urena Begara, J.-P. Raskin, and D. Flandre, "Control of sputtered VO2 thin films crystallinity towards optimization of neuromorphic 2T devices", in *Micro and Nano Engineering Eurosensors (MNE)*, Sept. 2022.

List of acronyms

AI	Artificial Intelligence
ANN	Artifical Neural Network
CED	Cumulated Energy Demand
CMOS	Complementary Metal-Oxide-Semiconductor
DNN	Deep Neural Network
GHG	Green House Gases
GWP	Global Warming Potential
H&H	Hodgkin & Huxley
HRS	High resistive state
IC	Integrated Circuits
IF	Integrate-and-Fire
IMT	Insulator-to-metal transition
IoT	Internet of Things
LCA	Life Cycle Asessment
LIF	Leaky Integrate-and-Fire
LoD	Limit of Detection
LRS	Low resistive state
LTD	Long-Term Depression
LTP	Long-Term Potentiation
MIT	Metal-to-insulator transition
NDR	Negative Differential Resistance
NIR	Near-infrared
PLD	Pulsed Laser Deposition
POU	Point Of Use
PPD	Paired-Pulse Depression
PPF	Paired-Pulse Facilitation
PSC	Post Synaptic Current
RF	Radio-frequency
SEM	Scanning Electron Microscope
SFDP	Spike Frequency Dependent Plasticity
SNN	Spiking Neural Network
SNR	Signal-to-Noise Ratio
TLM	Transfer Length Measurements
TRL	Technology Readiness Level
UPW	Ultra Pure Water
UV	Ultraviolet
VIS	Visible
XRD	X-Ray Diffraction

Acknowledgments

Comme vous pouvez l'imaginer, une thèse de doctorat ne peut pas s'accomplir seule, et je n'aurais pas réussi sans le support des nombreuses personnes qui m'ont entourée ces quatre dernières années.

First, I wanted to express my gratitude to all the members of my jury: Prof. Adrian Ionescu, Prof. Elisabetta Chicca, Prof. Jean-Pierre Raskin, Dr. Léopold Van Brandt, Prof. Laurent Francis and Prof. Denis Flandre. Thank you for your interest, for taking the time to carefully review my work, and for the insightful discussions we had during the private defense. It was a very enriching experience and allowed me to significantly improve this manuscript. I especially wanted to thank Elisabetta and Adrian for taking some time off their busy schedule to travel to Belgium and attend my private defense ! Elisabetta, thank you for your captivating talk on neuromorphic obstacle avoidance that you gave before my defense.

J'aimerais aussi évidemment remercier plus spécifiquement mon promoteur, Prof. Denis Flandre, pour son soutien durant ces quatre années. Merci pour ton intérêt sincère dans mes travaux, pour ta disponibilité lors des réunions de suivi qui se prolongeaient souvent en longues discussions techniques, et d'où je ressortait inévitablement avec un regain de motivation et d'inspiration. Ton expertise, ta créativité et ta capacité à apporter un nouvel éclairage aux problèmes techniques m'ont énormément apporté. Merci de m'avoir poussée à publier mes travaux même si je n'en étais pas 100 % satisfaite (l'est-on un jour ?), d'avoir relu et corrigé mes articles avec attention et minutie, parfois (souvent) sous contrainte de temps serrée (et d'avoir donc toléré mon côté "dernière minute"). Merci aussi pour les longs débats sur des sujets de société variés, qui m'ont permis d'affuter mon argumentaire...

Je tenais également à remercier mon comité d'accompagnement: Dr. Nicolas André, Prof. Dimitri Lederer et Prof. Laurent Francis (à nouveau), pour leur guidance lors de mon épreuve de confirmation. Merci encore aussi au Prof. Jean-Pierre Raskin: même si tu ne faisais pas partie de mon comité, j'ai beaucoup apprécié nos discussions scientifiques et la supervision des mémoires que j'ai pu encadrer avec toi sur les sujets liés à ma thèse.

Un énorme merci à mes collègues, des plus récent.es à celles et ceux qui ont déjà quitté l'UCLouvain, et dont beaucoup sont devenu.es des ami.es: Nicolas Roisin, Loïc, Clémentine, Anne-Sophie, Éléonore, Ottilie, Tom, Quentin, Romain Scaffidi, Thibault Delhaye, Margo, Ramy, Justine, Nicolas Brusselmans, Alex, Martin Vanbrabant, Martin Lefebvre, Augustin, Louis, Thibault Pirson, Nicolas Marchal, Roselien, Grégoire, Clara, Karim, Édouard, Pieter, Yiyi, Pol, ... Merci d'avoir égayé mes journées de travail, et pour les discussions diverses et variées à la cafétaria. Merci à celles et ceux avec qui j'ai eu la chance de partir en conférence, pour les très chouettes moments (et ceux dont je ne me souviens pas). Aux membres des bureaux A189 et A188: merci de m'avoir écoutée parler des divers problèmes liés à ma thèse (même quand ça ne vous intéressait pas), de m'avoir tirée en pause café ou en terrasse (même quand je n'en avais "pas le temps", et heureusement !). Votre bonne humeur et votre humour éclaire ce couloir, entendre vos éclats de rire va me manquer ! Une mention spéciale à Nicolas qui a relu ma thèse sur le chemin de ses vacances (ainsi que pratiquement tout mes articles) avec une minutie incroyable. Merci aussi pour tes figures *Blender* qui ont embelli cette thèse ! Merci à Loïc qui a assuré la défense de mon dernier mémorant et m'a permis de souffler après la remise de mon manuscrit; merci pour le très chouette voyage en Suisse, et d'avoir été un partenaire de galère sur ce sujet complexe qu'est le VO_2. Merci à Clémentine qui a souvent été la première oreille pour écouter mes déboires ou éponger mon stress (dès son premier jour). Vous êtes toutes et tous des chercheur.euses et des personnes en or !

Côté expérimental, je n'aurais pas pu réussir cette thèse sans le soutien précieux de toutes les personnes qui gravitent autour des plateformes technologiques *Welcome* et *Winfab*. Un tout grand merci à Christian, Sébastien, Miloud, Nicolas André, Mickey, Valeriya, Pascal, Souley, Benoit et Pierre pour votre grande disponibilité, votre flexibilité et votre pédagogie. J'ai une pensée particulière pour Nicolas André, qui m'a fait découvrir le monde magique de la microfabrication pendant mon mémoire, et sans qui je n'aurais probablement pas pensé à faire une thèse. Un merci spécifique également à Christian et Sébastien pour leur aide dans la réalisation de l'analyse de cycle de vie de mes dispositifs. Merci à Ferran de m'avoir formée au Raman, merci à Benjamin de m'avoir permis d'utiliser son four, sans lequel je n'aurais jamais su fabriquer mon VO_2. Merci à Pierre pour l'inspirante collaboration sur le thème des SNNs. Merci aussi à Brigitte et François pour leur très efficace support informatique.

Un tout grand merci également à toutes les membres du personnel administratif: Anne, Christel, Isabelle, Marie-Hélène, Nathalie et Émilie, qui m'ont soutenue dans un grand nombre de démarches indispensables au bon déroulement de ma thèse (toujours avec rapidité et bonne humeur), et qui contribuent à faire d'ELEN et ICTEAM un lieu où il fait bon vivre.

Merci aux divers mémorants que j'ai eu le plaisir d'encadrer, et en particulier à Pauline et à Tom, qui ont grandement contribué aux résultats présentés dans le Chapitre 4.

A côté de la thèse en tant que telle, j'ai eu la chance de m'investir dans toute une série de "projets parallèles" qui m'ont permis de respirer et m'épanouir dans des actions alignées avec mes valeurs, que je ne retrouvais pas toujours dans mes travaux techniques. Merci donc à toute l'équipe organisatrice de l'école doctorale *SICT*: Thibault Pirson, Antoine, Grégoire, Thibault Delhaye, Alex, Nicolas Moreau, Louis, Lou, Jan, David, Jean-Pierre, ... Merci pour ces deux merveilleuses années d'organisation, de motivation, d'idéalisme, de pragmatisme et de discussions inspirantes. Ce projet m'a apporté de fameux questionnements par rapport à ma thèse et a grandement contribué à construire ma vision de la place de la technologie dans notre société, et c'est grâce à vous. Merci aussi à toute l'équipe d'*EPL Transition*, pour votre motivation sans bornes, et pour continuer à tenter d'inclure les réflexions liées à la transition dans notre cursus: Marie-Charlotte, Isabelle, Delphine, Myriam, Marie-Amélie, Robin, et l'ensemble des membres du groupe de travail. J'ai énormément apprécié contribuer à ce travail de démocratie participative avec vous, même si ce n'était pas toujours facile. Merci à toute l'équipe d'*IngénieuxSud* pour la très épanouissante expérience d'encadrement de cours, et tous les échanges enrichissants sur le thème du partenariat Nord-Sud. Enfin, merci à toutes les personnes que j'ai pu cotoyer dans *Ingés en Transition*: Sébastien, Anne, Bastien, Martin, Paolo, Antoine, Zakarie, Nicolas Ghuys, Aurélia, Jeanne, Rémy, Ysaline, Mélanie, Thibaut Heremans, Laurine, Marie-Charlotte, Thibault Pirson, Alex, Nicolas Roisin, Louis, Robin, Ramy, Justine, Nicolas Brusselmans, Margo, Oriane, Simon, Loïc, Clémentine, Quentin, Mathieu, Clara et tous les nouveaux et nouvelles que j'oublie... Merci à celles et ceux du début d'avoir cru à ce projet fou lorsque l'on vous l'a présenté à ce temps de midi de février 2022. Merci à toutes et tous de vous en être emparé.e, de l'avoir fait vivre et grandir au delà de ce que j'aurais pu imaginer, d'y avoir chacun.e apporté votre coeur, votre temps et votre talent... J'ai très hâte de voir la suite. Longue vie à la Horde, j'ai adoré construire tout ça ensemble, vous allez me manquer !

Je voulais aussi remercier du fond du coeur mes amis et amies pour m'avoir encouragée pendant ces quatre ans, même si ce que je faisais devait vous sembler un peu abstrait. Merci pour le soutien dans les moments de doute, pour m'avoir écoutée et changé les idées quand j'en avais besoin, pour votre présence, chacune unique ! Merci à Alice, Milèna, Ella Mae, Ornella, Gilles, JB, Jeanne, et aussi à tout mon groupe de musique Freaky Mornings pour les géniales pauses musicales: Oriane, Greg, Gab, Pierre, Oli, Simon et Ugo.

Merci à toute ma famille et en particulier à mes parents d'avoir tout mis en place pour que je puisse accomplir mes objectifs. Merci Papa, pour m'avoir transmis ta passion pour la science. Merci Maman, de m'avoir légué ta volonté d'entreprendre et ta créativité. Et surtout, merci à tous les deux d'avoir cru en moi.

Enfin, merci à Antoine: sans toi, ça fait bien longtemps que j'aurais arrêté cette thèse. Merci pour la confiance que tu m'insuffles, et pour le soutien infaillible dans les hauts comme dans les bas. Tu m'inspires plus que tu ne le penses, scientifiquement et (surtout) dans tout le reste.

Introduction

Do the best you can until you know better.
Then when you know better, do better.
– Maya Angelou

Context

FOR decades, the improvement of computing performances has been driven by the famous Moore's law. Subsequent miniaturization steps allowed to follow this golden rule, which predicts that the number of transistors within a given area of integrated circuit (IC) would double every two years. This evolution has allowed engineers to design microprocessors with increased computing capabilities, transforming our world and, more recently, enabling the processing of ever-growing flows of data through AI models. However, with transistor dimensions now down to a few nanometers, the technology is approaching its physical limits and Moore's law is slowing down [1]. On the other hand, the amount of data generated *at the edge*[1] and requiring processing is expected to grow rapidly with the spread of Internet-of-Things (IoT) sensors [2, 3]. A first approach to tackle the problem would be to rationalize the amount of such data by reflecting on the scenarios in which they are truly relevant. In a complementary way, we can find radically less energy-hungry means to generate and process such data[2].

A community of researchers is actively studying the second option, and they are turning to biology for inspiration. Indeed, in terms energy-efficient decision making based on real-time sensory information, biological brains such as that of humans ($\sim$ 30 W) [5] or honey-bees (less than 10 µW) are yet unmatched by conventional computers. To reach such performance, the field of *neuromorphic engineering* underlines fundamental *paradigm shifts* with respect to conventional, Von Neumann computers. One of them relates to the encoding of information. Instead of encoding it digitally in bits and acquiring/updating it at a predefined

[1] In information and communication technologies, the edge designates data-emitting devices that are located at the periphery of the network, but are not permanently connected to this network. Examples are smartphones, connected watches, air quality monitors, ...

[2] For this strategy to be successful, awareness must be raised so that rebound effects due to new use-case scenarios do not cancel the benefits of technology optimizations on a global scale [4].

and fixed rate, biological neurons encode information into the *timing of electrical spikes* or action potentials. This approach allows to leverage the *sparsity* of events that is inherent to numerous sensing scenarios, in which relevant variations in stimulus information only take place at specific moments in time. This evolution has started to take place in the field of Machine Learning (ML): *Spiking Neural Networks* (SNNs), a bio-inspired kind of Artificial Neural Network (ANN), are the first kind of hardware to implement this paradigm.

However, there is still a missing link in the chain. To showcase their full potential, SNNs must be fed with a sparse, spike-encoded version of the stimulus information. A conventional approach would consist in sampling the output of analog sensors at a fixed rate using Analog-to-Digital Converters (ADCs), and convert it into spikes using a dedicated Digital Signal Processing (DSP) unit. But what if, akin to biological sensory neurons, we could bypass these blocks and their associated overhead, by designing compact sensors with a stimulus-dependent spiking behavior? This question opened up a new research field, that of *neuromorphic sensors*. Present CMOS-based implementations of such spiking sensors, referred-to as event-based cameras, suffer from two main limitations. First, they have a large pixel area [6], hindering scalability perspectives. Second, they are mostly limited to transducing optical stimuli in the visible domain [7], since they rely on the transduction capabilities of silicon.

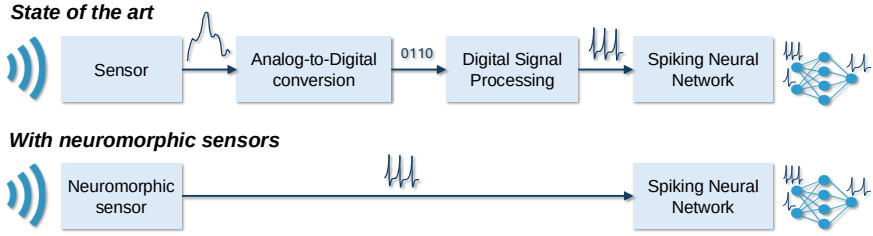


Fig. 0.1 Architectural paradigm shift brought by neuromorphic sensors.

In the recent years, researchers have turned towards *memristors* based on all kind of emerging materials to implement neuromorphic primitives [8–10]. As their conductivity can be modulated through electrical stimulation, they can emulate spiking behavior in a very compact way. More recently, they have been interfaced with sensors to design *sensitive neurons*, a form of neuromorphic sensors. In certain rare cases, stimulus transduction can even be performed directly within the memristive device. Phase-change material vanadium dioxide (VO_2) is one of the materials used to fabricate such memristors. VO_2 memristive devices benefit from three main assets to implement sensitive neurons. First, they can exhibit *volatile* switching from high- to low-resistive states under electrical stimulation, thanks to

their insulator-to-metal transition properties (whereas some other types of memristors exhibit non-volatile switching properties). Secondly, they are able to enter self-sustained oscillating regimes, allowing them to emulate complex neuronal behavior. Thirdly, their insulator-to-metal transition can be modulated by a diversity of stimuli types, enabling transduction functionalities without relying on an external sensor. However, as VO_2 spiking sensors are a fairly new research topic, their large-scale development is still hindered by a number of *limitations*, which will be tackled in this thesis.

Thesis objectives

The overall objective of this work is to explore some of the main present limitations of VO_2 spiking sensors. Rather than to try to improve the state-of-the-art (SoA) on specific key performance indicators (KPI), we want to gain understanding on the causes of these limitations, to derive guidelines on how they should be overcome.

Limitation #1 A first non-ideality relates to the inherent stochasticity of the resistive switching phenomenon, which is not unique to VO_2 memristors. This stochasticity induces cycle-to-cycle variations in the device electrical properties, which can be significant in certain cases. Depending on the given neuromorphic applications the device is intended for, the designer will wish to harness or mitigate these variations. However, given the present poor understanding of the causes behind such stochasticity, designers are rather clueless on how to proceed to maximize or minimize it. Therefore, we aim at answering the following question:

How can specific material properties and device design parameters impact the resulting VO_2 two-terminal devices cycle-to-cycle stochasticity?

Limitation #2 When VO_2 memristors are used to design sensitive spiking neurons which perform rate-coding of an input stimulus, one observes cycle-to-cycle variations in the durations of spikes. These variations can deteriorate the sensor Limit of Detection (LoD). However, how this *spike rate jitter* relates to the cycle-to-cycle variations in VO_2 electrical characteristics is not well understood. In addition, the LoD of VO_2 sensitive neurons is seldom quantified, although it can greatly affect the sensor energy efficiency: in a device with large spike rate jitter, one will have to measure the spike rate over a longer period of time to achieve a given resolution. We thus ask the following question:

What are the dominant causes of dynamic stochasticity in rate-coded spiking VO_2 sensors? How does this stochasticity impact the sensor Limit of Detection?

Limitation #3 Presently, all VO_2 spiking sensors perform rate-coding of the input stimulus. However, due to the specific way in which spikes are generated by VO_2 devices, there are no *refractory period* in between spikes, which are produced consecutively. As soon as a given spike is over, the next one will be produced: the information is actually encoded in the *duration of a spike* (and as a consequence, in the spike rate). This behavior differs from that of biological neurons, which emit action potentials of roughly constant duration independently of the input stimulus, but exhibit more or less long *refractory periods* or silent times in between spiking events, also implementing rate-coding. It's this specific feature that fosters the sparsity that is sought for in SNNs, and is presently lacking in VO_2 neurons. We thus ask ourselves:

Can VO_2 resistive-switching devices be used to design sensitive spiking neurons with sparser encoding techniques than plain rate-coding?

Limitation #4 Just as we did ourselves in the above context, VO_2 spiking sensors - and neuromorphic hardware in general - are often introduced by showcasing their energy-efficiency potential, which could allow to reduce the energy consumption associated to data generation and processing. Narratives suggesting to use neuromorphic technologies to strive for sustainable AI data centers [11] or green AI [12] are common. However, these works usually don't take into account the environmental impacts associated to the *fabrication* of these technologies. Although work attempting to quantifiy these impacts for mainstream CMOS technologies exist, this approach has never been undertaken for any kind of neuromorphic hardware. Indeed, these assessment studies are expected to be even more challenging when applied to emerging devices such as memristors. We thus ask the question:

What are the direct environmental impacts associated to the fabrication of VO_2 memristive devices? Can they be reduced through eco-design, and how do they compare to those of CMOS spiking neurons?

Thesis outline

To tackle these four main research questions, our work is structured as followed. Chapter 1 provides a first introductive dive from the top abstraction level of SNNs towards the specific field of VO_2 spiking sensors and their peculiar properties. Throughout the rest of the thesis, the specific results of our work will be presented as we climb our way back from the material physics and device aspects (Chapter 2) towards the application level (Chapter 3), ending with environmental impact assessment considerations (Chapter 4). Fig 0.2 sums up this structure.

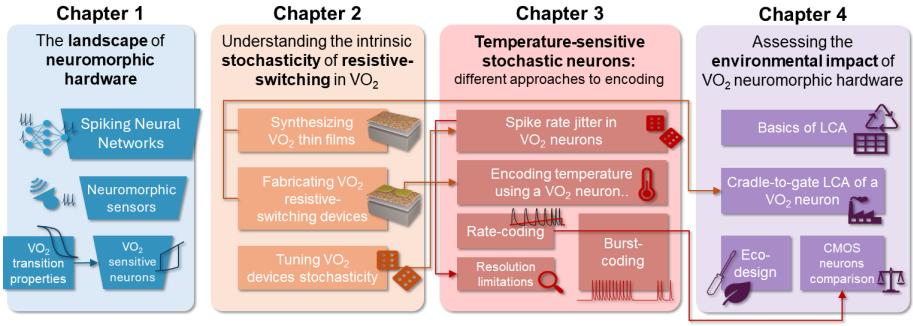


Fig. 0.2 Illustration of the thesis outline.

Chapter 1 provides an introduction to the concept of neuromorphic hardware and explains its main paradigm shift. Spiking Neural Networks are presented, together with their need for neuromorphic sensors. We then detail present CMOS implementations of such sensors, before showing how four main categories of memristive devices can provide a compact alternative. We then focus on the specific case of VO_2 memristive sensors. To understand their working principles and assets, a preface on VO_2 transition properties and multi-stimuli responsiveness is provided. Finally, we list the main types of existing VO_2 spiking sensors, so that the reader will be able to grasp their present limitations discussed in the following chapters.

Chapter 2 tackles Limitation #1 by improving the understanding of stochastic resistive switching in VO_2 two-terminal devices. Its end goal is to derive design guidelines to control this stochasticity. To that end, we synthesize several polycrystalline VO_2 thin films with various grain size, and use them to fabricate two-terminal memristive devices of different dimensions. We perform high-resolution electrical measurements of the devices I-V characteristics over hundreds of cycles. Through these measurements, we discover that the magnitude of the cycle-to-cycle stochasticity in the device electrical parameters is correlated to the number of grains along the device channel length. We show that minimizing grain size and maximizing device length allows to mitigate this stochastic behavior. To the best of our knowledge, this is the first study that takes an empirical, comparative approach to identify certain parameters that impact stochasticity. In addition, we provide for the first time hypotheses explaining the bi-modal distributions in VO_2 devices electrical parameters. This phenomenon has been observed previously in literature, but left unexplained. These conclusions could be especially useful to the memristive device research community, as they could apply to memristors based on other materials than VO_2.

Chapter 3 starts by investigating Limitation #2. By embedding the two-terminal devices fabricated in Chapter 2 within a simple, compact circuit, we design an artificial spiking neuron. We use it to study the relationship between the static cycle-to-cycle stochasticity studied in Chapter 2, and the spike rate jitter observed during the neuron dynamic operation. Using an analytical stochastic dynamic model, we find that the dominant sources of spike rate jitter are the cycle-to-cycle variations of the device electrical parameters in its insulating state. We then show that we can use this neuron to perform rate-coding of temperature, and quantify how the spike rate jitter impacts the sensor Limit of Detection. Interestingly, we prove that the energy-resolution trade-off induced by the spike rate jitter can be relaxed when increasing the spike rate.

The second part of this chapter investigates Limitation #3 by demonstrating a novel operating regime of VO_2 neurons. When biased at the limit of the region where they demonstrate stable spiking behavior, we show that they start emitting *bursts of spikes* separated by quiescent periods of time. We prove that stimulus information (temperature, in our case) can be encoded into the temporal characteristics of these bursts. This bursting behavior resembles the response of certain insect biological receptors to noxious heat, and exhibits a form of self-induced sparsity. To the best of our knowledge, it is the first time that this kind of behavior is emulated by an artificial sensory neuron.

Chapter 4 addresses Limitation #4 and reports a cradle-to-gate Life Cycle Assessment (LCA) to quantify the environmental impacts of the fabrication of our VO_2 device detailed in Chapter 2. By constructing a high-precision inventory of the process elementary flows (input materials and equipment electricity consumption), we estimate their impact on three metrics: global warming potential, cumulated energy demand and ultra-pure water consumption. We find that the use of gold for the electrical contacts dominates the footprint on the first two metrics, and show that it can be significantly reduced by using aluminium as an alternative. First characterizations show that performance seem to be preserved when using aluminium contacts. This conclusion could have positive impacts on the future development of VO_2 memristors, as almost all of them presently use high-impact precious metals for their contacts. This study is one of its kind, as it is the first time that LCA is applied to a neuromorphic technology. We use the results to provide a preliminary comparison to the environmental impacts of CMOS neurons.

Finally, we conclude by benchmarking the specific VO_2 neurons developed within this work compared to other VO_2-, memristor- and CMOS-based neuromorphic technologies, in terms of energy efficiency and neuron area. We discuss the limitations and perspectives of our work, and conclude on some final thoughts regarding the variety of possible applications of VO_2 neurons.

1

The landscape of neuromorphic hardware

What if, every time I started to invent something,
I asked: "How would nature solve this"?
– Janine Benyus

THE goal of this chapter is to **position the VO_2 devices** developed in this work **within the rich landscape of neuromorphic hardware**, and to highlight the ways in which they could complement and interface with existing neuromorphic platforms. This chapter starts from the higher abstraction level of neuromorphic architectures, zooming progressively towards the component level. First, we define neuromorphic processing systems, along with their main applications and essential characteristics. We explain why neuromorphic sensors are required to leverage the full potential of such platforms. We then detail the different types of neuromorphic sensors. We briefly present the different technological implementations allowing to create such hardware primitives, before digging into the specific case of vanadium dioxide-based devices. We present the phase transition properties of this peculiar material, and show how they can be leveraged to design resistive-switching devices able to emulate neurons. Finally, we present existing realizations of VO_2 neuromorphic sensors, and highlight the key challenges associated with their development which will be addressed in this thesis. The objective of this chapter is not to present an exhaustive state of the art of neuromorphic technologies at the higher abstraction level, but rather to provide the reader with a synthetic view of the research field. A detailed literature review is however presented for VO_2-based neuromorphic sensors.

1.1 Neuromorphic processing systems

The brain has always been a source of inspiration for computer scientists and electrical engineers, partly thanks to its impressive computing performance: with about 10^{11} trillions of neurons connected by around 10^{21} of synapses, it is able to perform complex recognition, inference and creative tasks at a power consumption of merely 30 W [5]. In comparison, the power consumption of supercomputers designed with a similar number of transistors lies in the megawatt range [13]. In addition, such supercomputers are designed to perform highly specific tasks (i.e. large scale, systematic computations), and do not offer the versatility and flexibility of the brain. It is generally considered that the development of bio-inspired artificial intelligence algorithms and hardware unfolded through three consecutive generations of architectures [14]: Artificial Neural Networks (ANNs), Deep Neural Networks (DNNs) (which are essentially larger ANNs), and Spiking Neural Networks (SNNs).

1.1.1 Artificial Neural Networks

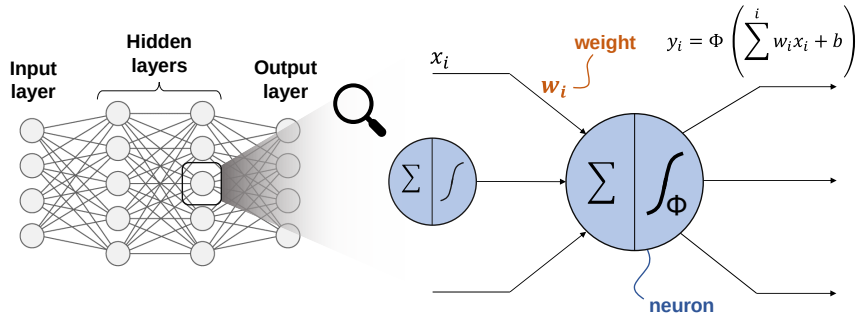


Fig. 1.1 Architecture of an Artificial Neural Network. Neurons perform a weighted summation of inputs, add a bias to this sum, and apply a non-linear function ϕ to the result.

ANNs are densely connected networks of individual neurons and synapses, as illustrated in Fig. 1.1. Neurons perform a linear combination of the data received from their input neurons, add a bias term, and apply a non-linear function (i.e. activation function) to the result. Synapses hold the network weights that are used as multiplying constants in the linear combination performed by the neurons. Updating these weights over time allows to implement learning within the network. The most common learning method is the backpropagation algorithm [15], in which synapses weights are updated from the output towards the input layers in a way that minimizes the error on the result. As Moore's law unlocked ever growing computing power capacities, the number of neuron layers in ANNs increased, yielding

the Deep Neural Networks (DNNs) that are the backbone of today's AI. These architectures contain large numbers of so-called hidden layers in between their input and output neuron layers, enabling the extraction of complex, abstract features. When run on conventional Von Neumann architecture, the storage/update of the network weights, and the computing of network states are performed on distinct locations (namely, in the CPU and in the memory), as illustrated in Fig. 1.2. The constant data shuttling between these two components creates overhead referred to as the *Von Neumann bottleneck*, limiting operations throughput since CPU speed has scaled faster than memory access speed [16].

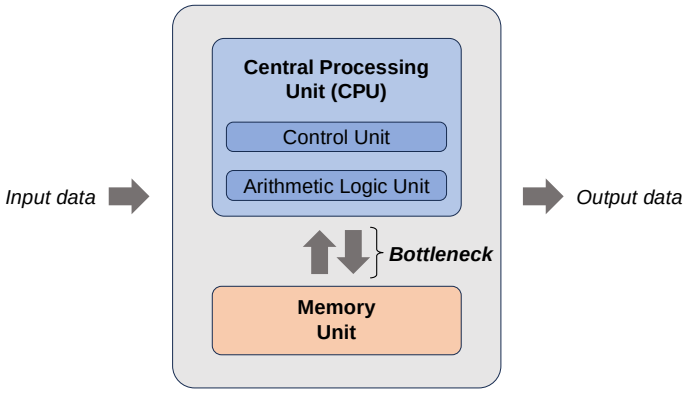


Fig. 1.2 Illustration of the Von Neumann bottleneck.

Solving the Von Neumann Bottleneck

To address this issue, a **first key paradigm shift brought by neuromorphic systems is the co-location of the memory and computing units**, a concept which gave rise to the field of in-memory computing (IMC). In-memory computing was first implemented in a digital fashion in the 1960s through logical in-memory arrays [17]. IMC leverages the fact that machine learning algorithms perform matrix-vector multiplication to compute the output of a neuron layer, and rely on a physical hardware implementation of such matrixes to perform these multiplications [18]. State of the art implementations can be digital [19] or mixed-signal [20, 21]. In their most literal interpretation, bio-inspired neuromorphic platforms implementing IMC design principles would thus be composed of densely connected hardware primitives emulating each individual neurons and synapses. In such implementation, the ANN is directly embedded within the hardware architecture, rather than run as a set of instructions on a CPU. For practical density and flexibility purposes, neuromorphic hardware platforms often implement multiplexing, meaning they emulate large network architectures on a smaller subset of neurons and synapses.

Today, a form of memory co-location that is gaining popularity within neuromorphic hardwares are memristive crossbar arrays [22], in which two fully connected layers of neurons are located at the column and row entries of a matrix, while synapses are located at the matrix crosspoints (Fig. 1.3). In these architectures, synapses are emulated by single devices called *memristors* which we detail further below, leveraging the peculiar physical properties of emerging materials.

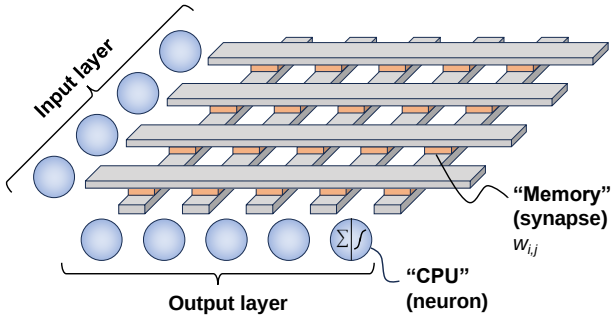


Fig. 1.3 Architecture of a memristive crossbar array.

1.1.2 Spiking Neural Networks (SNNs)

The **second key component of neuromorphic systems lies in the way they encode information**. Conventional processors encode information digitally in bits of zeros and ones, with a resolution defined by the number of bits. In contrast, a specific category of neuromorphic systems called Spiking Neural Networks (SNNs) takes further inspiration from the brain by encoding information in *all-or-nothing* events under the form of electrical spikes. In this case, information content is embedded in the timing of these spikes, following different possible codes. The architecture of SNNs is illustrated in Fig. 1.4. In SNNs, synapses and neurons definitions are extended with respect to ANNs.

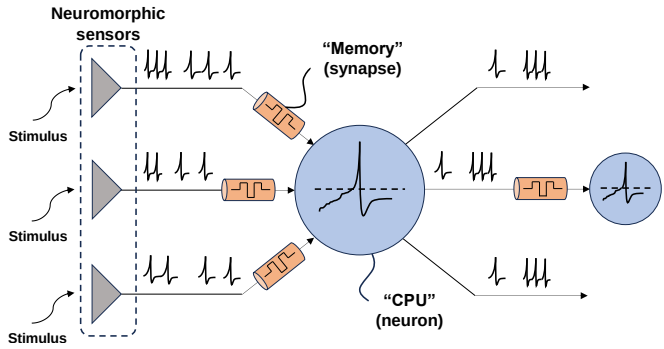


Fig. 1.4 Architecture of a Spiking Neural Network (SNN). Inspired by [23].

Synapses

As in ANNs, synapses are responsible for implementing one of the different existing learning rules, hence emulating the *memory* and *plasticity* (i.e. the ability to adapt to new information) of the brain. The key function of synapses in Spiking Neural Networks are to change their intrinsic state depending on certain characteristics of the electrical spikes that they receive from the neurons they are connected to. This variation impacts the way subsequent spikes are propagated to the neurons connected downwards of the synapse. The specific spikes characteristics that impact the synapse internal state depend on the learning rule. The most common is Spike Timing Dependent Plasticity (STDP) [24], in which the synapse internal state varies as a function of the relative timing of the pre- and post-synaptic spikes. If the neuron connected upstream of the synapse fires before (resp. after) the neuron connected downstream of it, expressing a form of causality (resp. non-causality), the synapse will reinforce (resp. decrease) the connexion between the two neurons. This phenomenon was first theorized by Donald Hebb in 1949 [25] under the maxim *"neurons that fire together wire together"*, or Hebbian learning rule. In neurons, it typically translates to long-term potentiation or LTP (resp. long-term depression or LTD). Synapses can be implemented analogically using CMOS technology [26, 27]. However, a category of analog implementation that has gained more interest in the recent years are *memristors*.

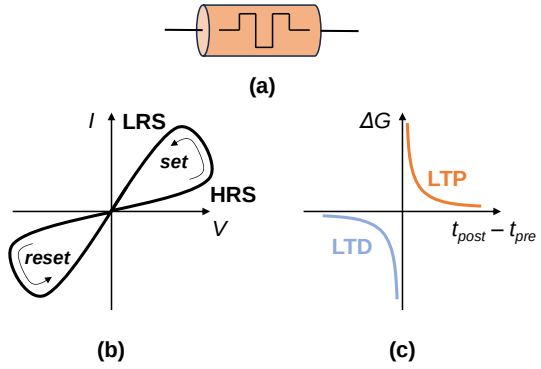


Fig. 1.5 Key characteristics of a memristor : **(a)** symbol, **(b)** theoretical I-V curve, and **(c)** implementation of Spike Timing Dependent Plasticity.

Memristors are passive electronic elements whose conductance varies in response to electrical stimulation. They are able to undergo a transition from a high-resistive state (HRS) to a low-resistive state (HRS). When these memristors are used to implement synapses, such variation must be *non-volatile*, meaning it must persist after the removal of the applied electrical stimulus for a more or less long timescale, or unless some form of reset is applied to them. Depending on the

specific timescale of this decay, they can implement long- or short-term plasticity. First theorized in 1971 by Chua [28], the memristor has been linked to physical implementations only 37 years later [29]. Today, a large variety of emerging materials are exploited for their peculiar physical properties allowing to design non-volatile memristors, as highlighted in multiple reviews [8–10]. They are often referred to as *resistive-switching devices* for their ability to switch from their HRS to LRS. The key characteristics of memristors are illustrated in Fig. 1.5.

Neurons

The most simple model of a biological spiking neuron was first introduced in the early 1900s [30] by Louis Lapicque. It is referred to as the Integrate-and-Fire (IF) model: like in ANNs, the neuron still performs a non-linear function of the sum of the inputs it receives. It does so by integrating input spikes over time and firing a spike or action potential once its internal potential (an analogy to the membrane potential in biological neurons) reaches a certain threshold. An extension of this model is the leaky Integrate-and-Fire (LIF) neuron [31], which includes the leaking of the neuron internal potential over time and can be modeled as an RC circuit (Fig. 1.6a). More complex models exist, such as the exponential Integrate-and-Fire (EIF) [32] where the membrane potential follows an exponential function upon threshold crossing; the adaptive exponential Integrate-and-Fire (AdEx) [33] which includes an adaptation current; and the Hodgkin and Huxley (H&H) model [34].

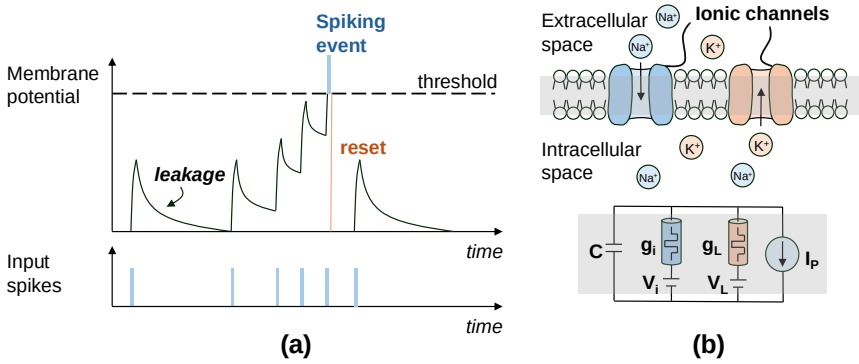


Fig. 1.6 Illustration of two neuron models: **(a)** the leaky Integrate-and-Fire (LIF) model and **(b)** the Hodgkin & Huxley (H&H) model. Adapted from [35].

This model captures the potassium and sodium ion channels of neurons which control how polarizing currents charge and discharge the neuron membrane, through a set of differential equations with four state variables. Each element of the biological neuron is modeled by an electric component (Fig. 1.6b): its membrane is modeled by a capacitor, its ionic and leaking channels are modeled by conductances, ionic pumps are modeled by a current source, and difference in ionic

concentrations between the membrane exterior and interiors are modeled by voltage sources. Generally speaking, there is a trade-off between the neuron models complexity and biological fidelity. Different implementations of artificial neurons have been proposed, both digital in CMOS technology [36] or (2) analog. Analog neurons, more specifically, can be realized both using CMOS transistors operated in subthreshold mode (first demonstrated by the pioneers of the neuromorphic field, Carver Mead and Misha Mahowald, in the early 1990s [37, 38]). Just like for synapses, a category of analog implementations that gained attention in the recent years are those based on memristors using emerging materials [39], which we further discuss in Section 1.2.2.

Hardware implementations

Despite the rising popularity of emerging analog memristive devices that demonstrate synapse or neuron functionalities, large-scale hardware implementations of SNNs still mostly rely on CMOS technology and are all at least partly digital. A distinction can be made between *mixed-signal* and *digital* implementations. Mixed-signal implementations typically perform their computations with analog neurons (the latter often relying on subthreshold CMOS dynamics), while the routing of spikes through the network is handled digitally. Examples of such chips are DYNAPs [40], ROLLs [41], BrainScaleS [42] and Neurogrid [43]. On the other hand, fully digital SNNs are used as accelerators to emulate large cognitive models and include examples such as TrueNorth (*IBM*) [44], Loihi (*Intel*) [45], ODIN [46] and MorphIC [47] (*UCLouvain*) neuromorphic chips. While analog/mixed-signal implementations are more biologically plausible, they require longer design times and provide less portability from one technological node to the next [48]. For smaller problems such as pattern recognition involving very small numbers of pixels, SNNs composed exclusively of memristive neurons and synapses have been demonstrated theoretically [49] and experimentally [50], the latter being illustrated in Fig. 1.7.

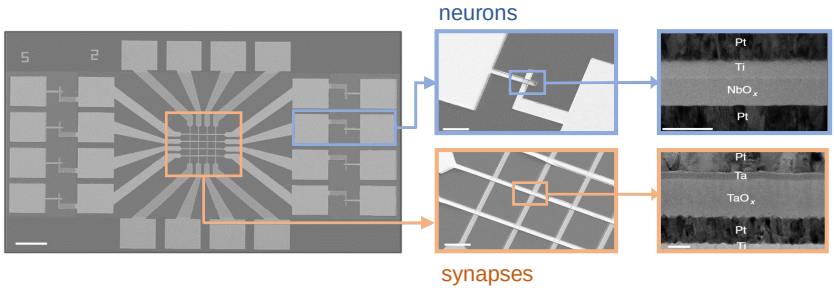


Fig. 1.7 Spiking Neural Network fully implemented using memristive devices for neurons and synapses. Adapted from [50].

These architectures offer high scalability perspectives as neurons and synapses can be implemented with single devices. However, devices non-idealities limit the further deployment of fully memristive architectures. These include endurance issues, non-linearity of the memristive synapses conductance under electrical stimulation, and device-to-device variability. Mixed-signal, hybrid memristive-CMOS implementations can mitigate these non-idealities [51]. The drift of the memristive synapses internal state over time also limits the lifetime of the system memory. Finally, in fully memristive SNNs, information of each image pixel must still be encoded into spikes in a software fashion before being fed to the hardware implementing the network. Demonstrations of fully analog spiking systems from sensing to processing are still missing.

Information encoding

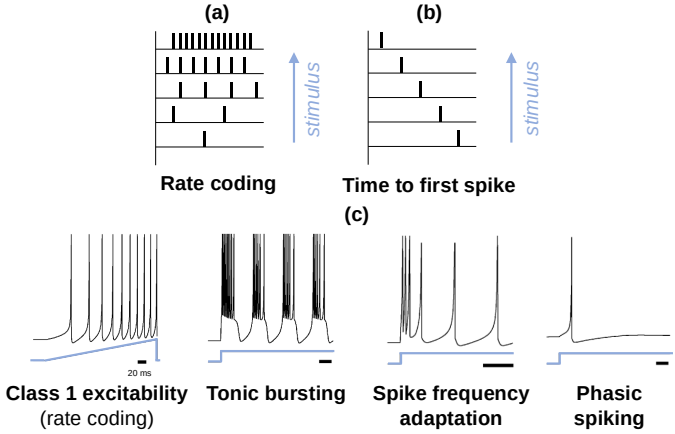


Fig. 1.8 Different types of neuronal code: **(a)** rate coding, **(b)** time to first spike, **(c)** Four out of the 20 Izhikevich behaviors. Adapted from [52].

Several neural code have been observed in biological neurons. The most common is rate coding (Fig. 1.8a), in which information is encoded into the frequency of spiking events. Izhikevich reported the most prominent spiking behaviors of biological neurons [52], a few of which are illustrated in Fig. 1.8c. Rate coding can be assimilated to the *Class 1 excitability* among the Izhikevich behaviors. With rate coding, one needs to witness a certain amount of consecutive spikes to be able to deduce their information content. Other codes exist, such as *time to first spike* (Fig. 1.8b), which allows to encode information in single spikes through the delay between the stimulus onset and the spiking event. Spike-based information encoding leverages event sparsity: with rate coding, the lower the stimulus power or information content, the lower the spike rate, meaning that the network is essentially asleep in the absence of meaningful information (it is however noteworthy that

time-to-first-spike encoding is more energy efficient than rate coding). In "at-the-edge" scenarios in which real-time processing of sensory information is required, and in which the data has a high temporal dimension (i.e. the data only shows relevant variations during short periods of time), such sparsity could be leveraged to perform energy-efficient inference tasks such as pattern recognition. This however requires a hardware with low static power consumption in the absence of spikes.

Energy efficiency and need for spiking sensors

Following such logic, several works have attempted to prove the potential of SNNs compared to conventional ANNs in terms of energy efficiency. So far, conclusions are mitigated. Dampfhoffer et al. [53] estimated theoretically that SNNs could outperform ANNs in energy efficiency under conditions of very high spike sparsity, less than 1.5 spikes per inference. Lemaire et al. [54] modelled an up to 8-fold superiority in energy efficiency of SNNs compared to ANNs. However, Bhattacharjee et al. [55] show that theoretical estimations can underestimate actual measured energy consumption, due to several hardware bottlenecks including device-to-device variations. Davidson and Furber [56] highlight that when using rate coding, event sparsity is not sufficient for SNNs to outperform ANNs from an energy viewpoint. In their work, Frenkel et al. [48] benchmark their digital SNN implementation compared to other SNNs and ANNs using N-MNIST, a spiking version of the MNIST dataset [57], and show that in most case SNNs do not outperform ANNs in terms of energy efficiency. However, they underline the fact that the N-MNIST dataset is constructed from static images, thus not conveying the temporal dynamics that might benefit SNNs. Benchmarking using SNNs-dedicated datasets[1] with a prominent temporal dimension such as the Heidelberg dataset [60] would be desirable. Spiking datasets are needed as the use of SNNs for at-the-edge recognition tasks requires sensory information to be encoded under the form of spikes. This can be performed by converting the analog output of conventional sensors into the spiking domain, sampling it at a predefined rate and converting it into spikes using Analog-to-Digital converters (ADCs) and Digital Signal Processing (DSPs) units [61]. However, such clock-based hardware blocks induce overhead and would impede the system from truly benefiting from event sparsity in a real-life scenario of embedded application. In consequence, engineers have been trying to design *neuromorphic sensors* able to directly encode information into the spiking domain, in a compact and energy-efficient manner.

[1] By opposition to ANNs-tailored ones such as MNIST [58] or CIFAR10 [59], both based on static inputs.

1.2 Neuromorphic sensors

Just like the hardware primitives of neuromorphic processing systems, neuromorphic sensors can either take the form of *sensitive synapses* or *sensitive neurons*, each of which can be implemented in different fashion.

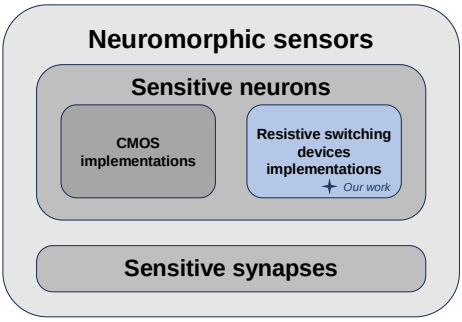


Fig. 1.9 Cartography of neuromorphic sensors implementations. Our work relates to sensitive neurons implemented using resistive-switching devices.

1.2.1 Sensitive synapses

Sensitive synapses show a change in their internal state in response to stimuli, but are not able to generate spikes by themselves and thus require to be interfaced with an upstream artificial neuron, or artificially stimulated with spikes. Sensitive synapses encode stimulus information into the network weights, with the effect of strengthening or weakening connections between specific neurons. In recent years, an increasing number of sensitive synapses have been implemented using non-volatile memristors [62–68] (introduced in Section 1.1.2), in which transduction is directly performed by the memristor. The specific physical mechanism depends on the used material. The most basic form of memory implemented by these sensitive synapses are Long Term Potentiation (LTP) or Long Term Depression (LTD), a persistent change in device conductance under stimulus power, impacting the post-synaptic current (PSC). Two main forms or LTP/LTD are reported. In the first case, the potentiation is performed through input voltage spikes, and the stimulus power modulates the magnitude of the conductance variation [62] (Fig. 1.10a). The magnitude of conductance change will also depend on the input spikes frequency (spike frequency dependent plasticity or SFDP), duration and power. Alternatively, the potentiation can be directly triggered by applying stimulus pulses (e.g. light pulses for photosensitive neurons [63–65]; see Fig. 1.10b). Here again, dependence on stimulus pulse frequency, duration and power is demonstrated. Typically, long

term depression or resetting of the synapse initial state after LTP will be performed by applying voltage pulses of opposite polarity. Other types of synaptic behavior implented by memristive sensitive synapses are paired-pulse facilitation (PPF) or depression (PPD), where postsynaptic currents are temporarily increased or reduced by previous pulses (Fig. 1.10c-d). Once again, this behavior can be triggered by voltage input pulses modulated by the stimulus of interest, or directly triggered by stimulus pulses (mainly in light sensitive synapses). Currently, memristor-based sensitive synapses have been demonstrated for strain [62, 63, 65, 67] and optical stimuli (in the ultraviolet or UV [65, 68], visible or VIS [63] and near-infrared or NIR [64] spectra). For light sensors, most implementations rely on phototransistor architectures based on emerging materials. In all these implementations, transduction is performed directly by the memristive device.

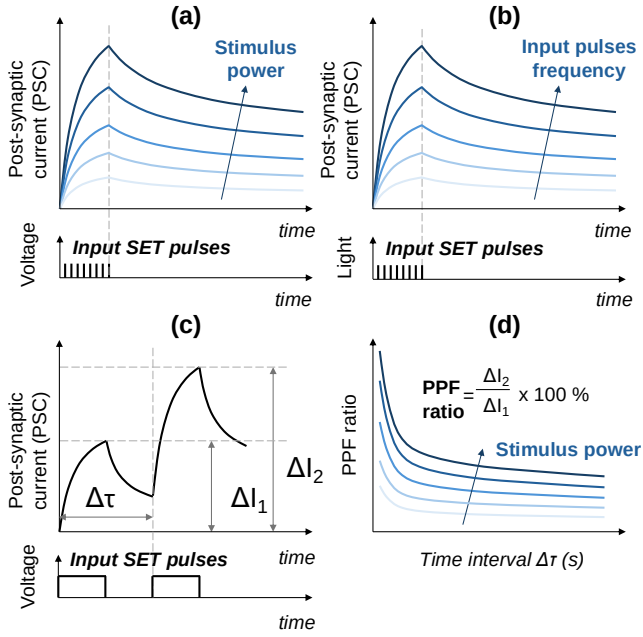


Fig. 1.10 Different behaviors of sensitive synapses. **(a)** Long-term potentiation (LTP) triggered by input voltage pulses, and modulated by stimulus power. **(b)** LTP triggered by stimulus pulses (here, light), and modulated by pulses frequency. **(c)** Definition of paired-pulse facilitation (PPF) ratio. **(d)** Modulation of PPF by an external stimulus.

1.2.2 Sensitive neurons

Sensitive neurons, on the other side, are able to generate electrical spikes under certain bias conditions, and to encode stimulus information into the timing of these spikes. They thus simultaneously fulfill the functions of stimulus transduction and spike generation.

CMOS implementations

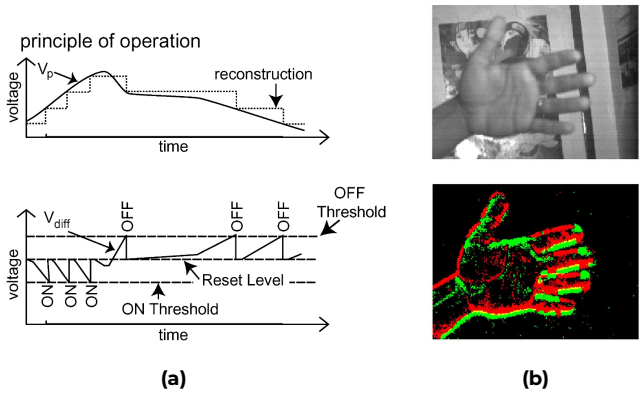


Fig. 1.11 Working principle of an event camera. **(a)** Spikes generation from the signal of an analog light sensor. **(b)** Example of generated image, red (resp. green) spikes correspond to negative (resp. positive) brightness change. Adapted from [7, 69].

The first generation of such sensors to receive attention in the neuromorphic research field are so-called event-cameras. In such cameras, each pixel is independent and generates asynchronous "all-or-nothing" events (spikes) in response to brightness change instead of absolute brightness. In addition to carying temporal and spatial information, the spikes have a polarity depending on whether the brightness change is positive or negative. The concept is illustrated in Fig. 1.11. The idea of event cameras, or dynamic vision sensors (DVS), was first introduced in the early 90s by Mahowald and Mead as an inspiration from the way the biological retina operates [70]. It was then further developed by of Todi Brucker's lab in ETHZ [7] in the 2000s. Despite their bio-plausibility, they suffer from three main limitations. First, a large pixel size, as each pixel typically relies on a large number of transistors[2]. Second, the fact that most implementations limit themselves to sensing light in the visible spectrum, as they use CMOS-based photodiodes and are thus restricted to the transduction mechanisms taking place in silicon. A few demonstrations of CMOS-based spiking sensors could however be adapted to other stimuli than optical ones by interfacing conventional sensors with CMOS-based analog or digital circuits [71–73]. The third limitation relates to a relatively high pixel power consumption, as they require a biasing current leading to significant static consumption (while conventional CMOS image sensors are purely dynamic, with no static power consumption in the pixel array). All in all, the main drawback of all

[2]There has however been progress made on that matter, with a collaboration between Sony and Prophesee demonstrating an event camera with a 24 µm² pixel size [6], thanks to the 3D-stacking of a backside illuminated image sensor and processing chip.

CMOS implementations of spiking sensors is that they require a significant number of active elements for the spike generation function, posing scalability concerns, mostly as Moore's law is slowly reaching its limits. While techniques such as 3D stacking and chiplet interconnects [74] could mitigate the Moore's law slowdown, they cannot totally prevent it.

Resistive-switching device implementations

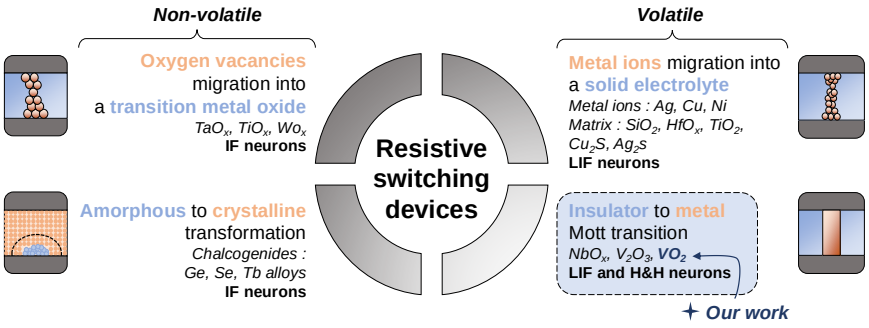


Fig. 1.12 The four main families of resistive-switching devices, together with their respective physical mechanisms, typical materials, and type of neuron models that they can implement. Our work belongs to the family of Mott memristors. Schematics show the state of the device in high-resistive state (HRS) in each four cases. Inspired from [39].

An alternative is to follow a similar approach than for sensitive synapses, and to turn to memristors to implement *computational neurons* that perform spike generation function, in a way that requires a smaller number of circuit components. The modulation of the timing of these spikes by an external stimulus can then be implemented by interfacing the computational neuron with an external sensor, to build *sensory neurons*[3]. This design paradigm exploits the resistive-switching properties of memristors based on emerging materials (introduced in Section 1.1.2). The nonlinearity of their I-V curve makes it possible to use them to implement IF and LIF neurons, and even more complex models such as the H&H neuron. Typically, the two-terminal resistive devices will integrate input spikes through a variety of possible physical mechanisms, leading to the formation of a conductive filament. The formation of this filament is concomitant to the generation of a current or voltage spike due to the sudden decrease of conductance, as the device switches from its high- to low-resistive state. To mimic the neuron behavior, this filament must then resorb in the absence of electrical spikes so that the device returns to its HRS. Depending on the underlying physical mechanism, this relaxation will either be spon-

[3]This distinction between *computational neurons*, which solely perform spiking functions, and *sensory neurons*, which implement spiking *and* sensing functions, will be used systematically throughout this thesis.

taneous (corresponding to the case of the *volatile* memristor, in which the resistive switching is unstable), or artificial through the application of a RESET pulse (for *non-volatile* memristor, in which the resistive switching persists over time). Fig. 1.12 quickly summarizes the four main physical phenomena behind resistive switching: oxygen vacancies formation, amorphous to crystalline transition, metal ions migration, and Mott-type insulator-to-metal transition. We discuss each of these four categories in more details in the following paragraphs (see Fig. 1.13 to 1.16).

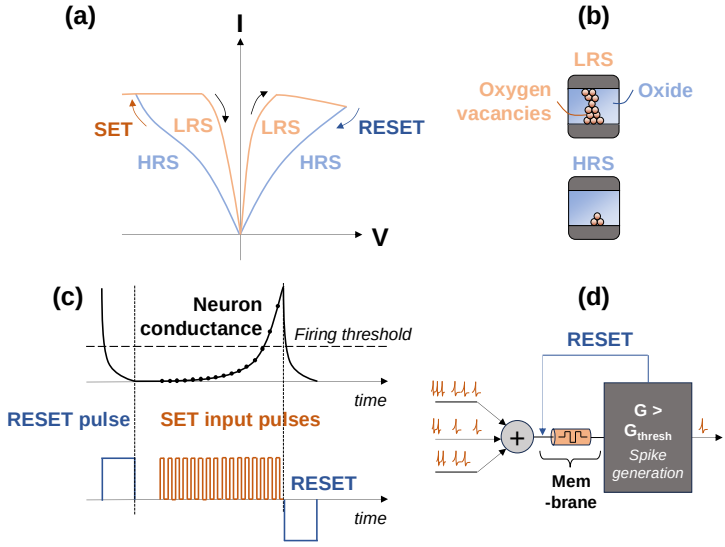


Fig. 1.13 Operation principle of a resistive-switching neuron based on **oxygen vacancies migration**. **(a)** Typical I-V curve. The HRS is shown in orange, the LRS in blue. **(b)** High- and low-resistive states, showing the oxygen vacancies (orange) within the oxide matrix (blue). **(c)** Integration of input pulses and spike firing, followed by relaxation through a RESET pulse. **(d)** The memristor integrates inputs and stores the neuron internal potential, while external circuitry is required for threshold detection and resetting of the neuron state. Inspired from [75,76].

A first type of resistive-switching device illustrated in Fig. 1.13 is based on the **migration of oxygen vacancies into an oxide** such as HfO_x, TaO_x, $Pr_xCa_yMnO_3$. In such devices, the transition from high- to low-resistive state takes place through a soft breakdown of the oxide under a sufficiently high applied electric field, resulting in the formation of a conductive pathway of oxygen vacancies (Fig. 1.13b). For this filament to resorb, an electric field of opposite polarity must be applied to trigger recombination of the ions (Fig. 1.13a). Such resistive switching is thus non-volatile, but can be leveraged to implement an IF behavior [75,76]: the device integrates SET input spikes until its internal state reaches a critical level of ionic transport and a conductive filament of oxygen vacancies connects the device two

terminals, after which the conductance increases suddenly (Fig. 1.13c). These types of memristors essentially store the neuron membrane potential, while external circuitry is required to detect threshold crossing (Fig. 1.13d). Once the spike has been fired, the neuron must then be reverted back to its HRS by applying a RESET pulse of reversed polarity. This feedback mechanism also requires additional circuitry which limits the compactness of the neuron.

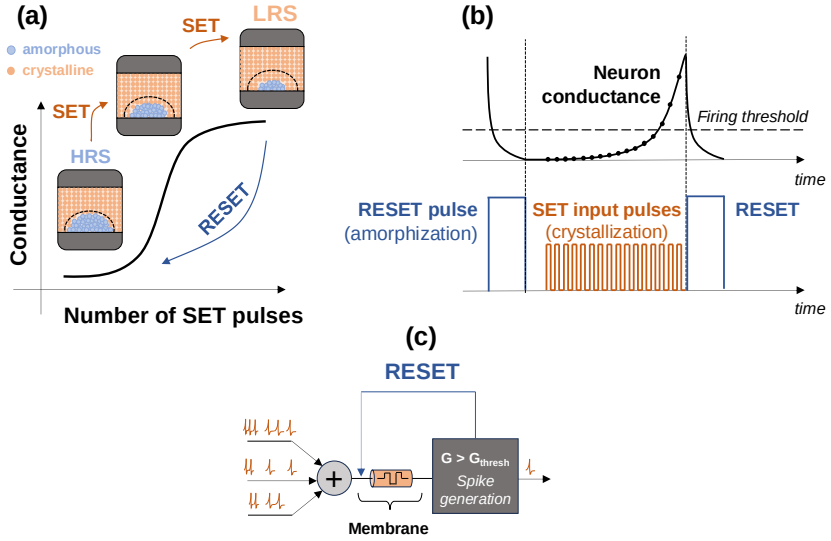


Fig. 1.14 Operation principle of a resistive-switching neuron based on **amorphous to crystalline transition**. **(a)** Conductance variation with number of input SET pulses. The initially large amorphous region (blue) progressively shrinks as the input voltage pulses induce crystallization (orange), increasing the device conductance. The conductance can then be reset to its low value by recreating the amorphous region through a high-power voltage pulse. **(b)** Integration of input pulses and spike firing, followed by relaxation through a RESET pulse. **(c)** Here again, the memristor integrates input and stores the neuron internal potential, while external circuitry is required for threshold detection, spike generation and resetting of neuron state. Inspired from [77].

A second type of resistive-switching illustrated in Fig. 1.14 exploits the progressive **re-crystallization of an amorphous region** within the memristor channel, triggered through Joule heating by applying short voltage pulses of relatively low power (Fig. 1.14a). Typical materials implementing such reversible crystallization are chalcogenides such as Ge, Se and Tb alloys [77–79]. Similarly to oxygen vacancy migration, this phenomenon is not volatile and must be reverted back by applying a longer, higher-power voltage pulse (Fig. 1.14b) to heat up a significant part of the device above the material melting point, re-creating a large amorphous region upon quenching through a glass transition. Once again, the threshold crossing detection, spike firing and RESET functions involve additional

circuit elements (Fig. 1.14c). In this case, there is no conductive filament that fully connects the two electrodes, but rather an amorphous region that shrinks progressively through crystallization, and is then reset back through melting and quenching.

A third type of resistive-switching mechanism illustrated in Fig. 1.15 is based on the **migration of metal ions from a metallic electrode** (Ag, Cu, Ni) **into a solid electrolyte material** (Ag_2S, Cu_2S, SiO_2, SiO_x, ...). Similarly to oxygen vacancy migration, metal cations will migrate from the electrode under the action of an electric field (Fig. 1.15b). A conductive filament progressively forms under the action of repeated electrical pulses (integration); the abrupt increase of conductance once a conductive filament of cations forms across the device mimics the firing of a spike. In the absence of applied electric field, the conductive path will resorb naturally through diffusion as the metallic nanoparticles tend to regroup in isolated clusters to minimize their interfacial energy, so that no RESET pulse is needed (Fig. 1.15a). This volatility allows to implement the leakage behavior of LIF neurons [80–82] (Fig. 1.15c). The relaxation time can range from the tens of micro- to the tens of milliseconds range[4]. In case of too fast relaxation compared to the input dynamics, a parallel capacitance will be added to the memristor to perform integration of input spikes (contrary to non-volatile device in which this integration is performed within the memristor itself).

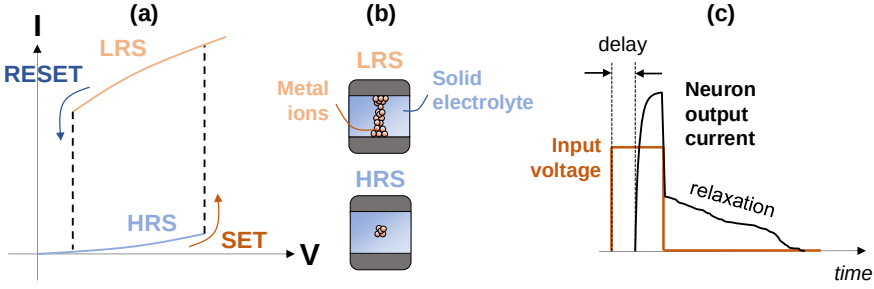


Fig. 1.15 Operation principle of a resistive-switching neuron based on **metal ions migration**. **(a)** Typical I-V curve. **(b)** High- and low-resistive states. **(c)** Integration, firing and spontaneous relaxtion upon input voltage pulse. Inspired from [82].

Finally, a fourth type of resistive switching illustrated in Fig. 1.16 exploits the reversible **insulator-to-metal transition** (IMT) in Mott insulators such as NbO_x, V_2O_3 and VO_2. This transition corresponds to a change of energy band and crystalline structure in the material, and is accompanied by a drastic reduction in resistivity. Interestingly, this transition can be triggered by several types of stimuli, depending on the specific material. The most generic stimulus which is observed

[4]Typically, biological neural codes operate in the 1-100 millisecond timescales [52], with each spike lasting for about 1 millisecond [83].

to trigger such transition is temperature: the transition temperature T_{IMT} is around 1080 K for NbO_2, 340 K for VO_2, 415 K for V_3O_5. In small dimensions (micrometer-scale) two-terminal devices, the transition can be triggered by an applied voltage through a debated combination of electrical and Joule heating effects[5].

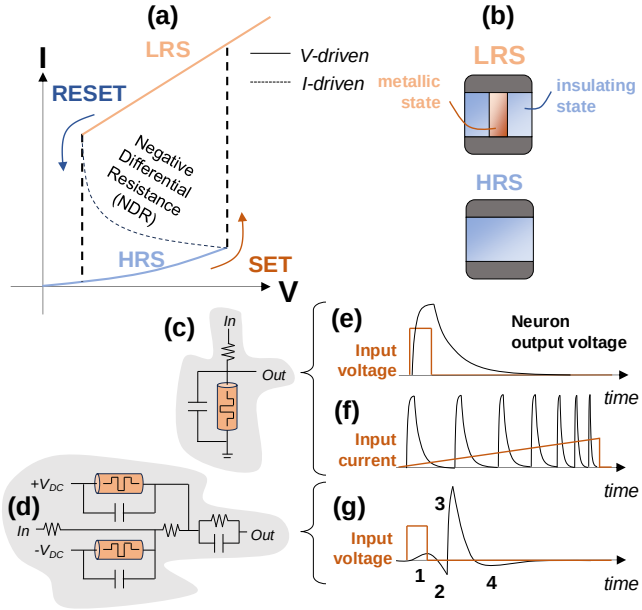


Fig. 1.16 Operation principle of a resistive-switching neuron based on **Mott insulator-to-metal transition**. **(a)** Typical I-V curve. **(b)** High- and low-resistive states. Circuits allowing to design LIF and rate-coded neurons **(c)** and H&H neurons **(d)**. **(e)** Firing of a spike under input voltage pulse (LIF neuron). **(f)** Self-sustained oscillations in case of device presenting a NDR, with rate coding of input current. **(g)** Bio-plausible spiking behavior as implemented in a H&H Mott neuron with four specific stages: resting, hyperpolarization, depolarization and refractory undershoot. Inspired from [84, 85].

The result is the formation of a metallic filament (LRS, Fig. 1.16b), which resorbs as soon as the applied voltage is removed, allowing the device to return to its HRS. Mott insulators are characterized by their ultrafast transition dynamics: both the transition from HRS to LRS (integration) and LRS to HRS (relaxation) take place in the nanosecond timescale, allowing to design neuromorphic devices with accelerated timescales with respect to biological ones. Here again, a parallel capacitance is thus commonly used in computational neurons to integrate input spikes and design LIF neuron. Under certain conditions, resistive-switching devices based on Mott insulators can present a range of *negative differential resistance* (NDR) in their I-V curve (Fig. 1.16a). By setting a current corresponding to this specific

[5]The literature debate around the triggering mechanism will be detailed in Section 1.3.2.

range, the device can be set in an astable regime in which the filament will form and resorb in a cyclic fashion, resulting in a self-sustained spiking behavior. In this specific regime, the spike rate can be modulated by the level of input current, implementing a rate-coding behavior (Fig. 1.16c,e,f). When combining two Mott memristors together, one can model the potassium and sodium ion channels of the H&H neuron model and capture more complex behaviors (Fig. 1.16d,g) [84–86].

Since the early 2010s, these four types of resistive-switching devices allow to implement computational neurons [78, 84]. Theoretically, they also offer the possibility of being interfaced with sensors to design sensitive neurons. However, demonstrations of the latter have only emerged in recent years. Most implementations rely on the two types of *volatile* memristors, namely resistive-switching devices based on Mott insulators [87–97] and metal-ions migration mechanisms [64, 98–102], thanks to their ability to mimic more complex neuron behavior than plain IF (i.e. LIF, H&H). Most Mott memristors implementing sensitive neurons rely on the use of NbO_x [87–90] and VO_2 [91–97] as active material. Regarding sensitive neurons based on metal-ions migration, the metal used preferentially as active electrode is Ag[6], while the most common solid electrolyte materials are SiO_2 [64, 101] and TaO_x [100, 102]. In most cases, the sensing function is fulfilled by an external sensor whose analog output is fed to the neuron. However, in certain niobium oxide [88], vanadium oxide neurons [91–94] and metal-ion migration neurons [98–100], stimulus transduction is performed directly by the memristor itself by exploiting the dependence of the device I-V characteristics on stimulus power (which can be due to various physical mechanisms depending on the specific material). Such implementations where the sensing is directly embedded within the volatile memristor have been reported for thermal [92, 94, 100], optical [91, 98, 99] and RF [91, 93] stimuli. On the other hand, strain sensing is typically implemented through external sensors [88–90, 95, 101, 103]. The compactness of integrated-sensing implementations is an interesting feature, especially in scenarios in which large number of sensors must be integrated in a reduced area.

1.3 Vanadium dioxide for neuromorphic applications

We now dig more specifically into implementations of neuromorphic sensors based on Mott insulator vanadium dioxide (VO_2). We first detail how the insulator-to-metal transition takes place in VO_2 thin films, and what are the dominant physical mechanisms put forward in the literature to explain such phenomenon. An interesting property of VO_2 - compared to other Mott insulators such as NbO_x - is

[6]Thanks to the fast diffusion dynamics of Ag ions in a variety of electrolyte materials. Cu can be used as an alternative, but the electrode will be subject to oxidation.

that its phase transition can be triggered through a wide range of stimuli including thermal, optical, strain, magnetic, electro-chemical and electrical [104]. This multi-stimuli responsiveness makes it an ideal candidate to implement neuromorphic sensors with transduction directly embedded within the memristive device. We quickly review how these different stimuli can affect the insulator-to-metal transition, or even directly trigger it in certain cases. We then focus on the electrically triggered transition, and show how it can be harnessed in micro-scale VO_2 two-terminal devices. We present the existing architectures of VO_2 spiking sensors based on such two-terminal devices. Finally, we evoke the remaining challenges associated to the deployment of VO_2 spiking sensors, which will be tackled throughout the different chapters of this thesis.

1.3.1 Transition properties

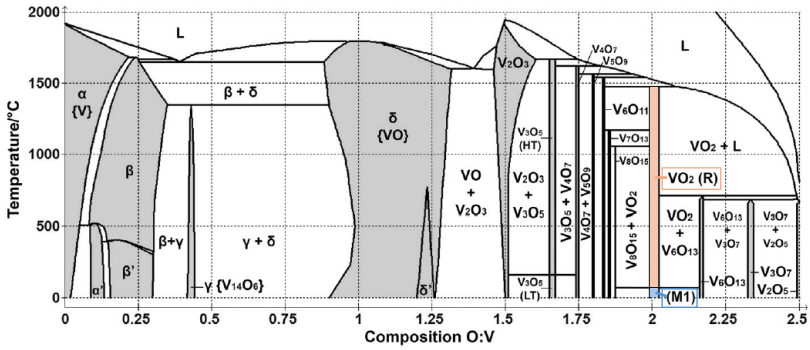


Fig. 1.17 Phase diagram of vanadium oxides as function of their oxygen/vanadium content and temperature. Greyed areas correspond to single phases, other correspond to mixture of phases. The insulating and metallic phases of VO_2 are highlighted in blue and orange, respectively. Adapted from [105].

Vanadium dioxide (VO_2) is one of the numerous existing stoichiometries of vanadium oxides, as shown in the phase diagram of Fig. 1.17. Among these different phases, five are thermodynamically stable, meaning they are stable as pure compounds: V_2O_3, V_3O_5, VO_2, V_3O_7, and V_2O_5. Other phases are referred to as unstable or metastable, and are likely to decompose into one of the stable phases. Vanadium oxide stoichiometries can also be decomposed into (1) those in which vanadium atoms have single valence (there are four of them: VO, V_2O_3, VO_2 and V_2O_5), and (2) those with a mixed-valence of vanadium. The latter either belong to the family of Magnéli series (V_nO_{2n-1}, $3 \leq n \leq 9$) or Wasley series (V_nO_{2n+1}, $n \geq 3$). Thermodynamically stable phases have applications as batteries, catalysts and supercapacitors. In addition, they all present some form of phase transition, taking place at various temperatures [106]. However, most of them are either too

low (around -113 °C for V_2O_3 [107], -268 °C for V_3O_7 [108]) or too high (around 157 °C for V_3O_5 [109], 257 °C for V_2O_5 [110]) for practical applications in electronic, as CMOS circuits typically operate from room temperature to 125 °C [111].

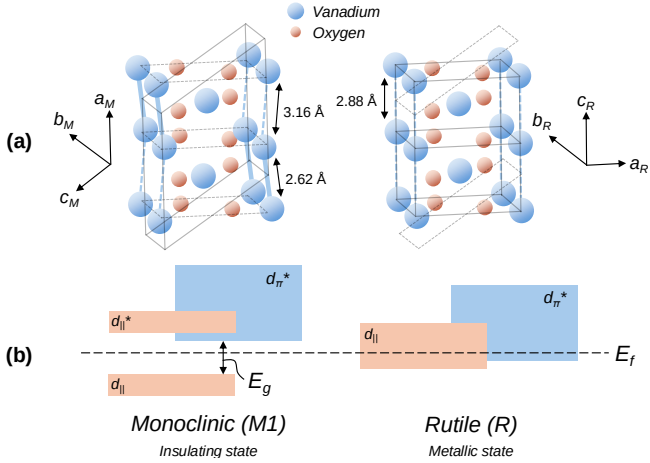


Fig. 1.18 Difference in **(a)** crystalline structure and **(b)** band structures between the insulating and metallic vanadium dioxide phases. Adapted from [112]. In **(a)**, the solid blue line show the V-V dimers.

However, VO_2 stands out as an interesting candidate thanks to its insulator-to-metal transition taking place close to room temperature. This transition was first observed by Morin et al. in 1959 [107], triggered thermally at a temperature T_{IMT} of around 340 K (68 °C). The two phases have different crystalline structures [104] (shown on Fig. 1.18a). In its metallic phase, VO_2 has a rutile (R) tetragonal crystalline structure, in which vanadium ions are arranged in a linear, symmetric way. The average distance between the nearest V-V pairs along the c_R axis is identical for all pairs (about 2.88 Å). In its insulating phase, VO_2 has a monoclinic (M1) phase of low symmetry, and one witnesses a dimerization[7] of the V atoms, with dimers rotating in the c_M direction. As the Vanadium dimers rotate, two different V-V bond lengths appear (2.65 Åand 3.12 Å), and the bonds form zig-zag patterns. The two phases also have distinct band structures [113] (shown on Fig. 1.18b). In the metallic phase, there is a superposition of the d_{π}^* and $d_{||}$ bands, and the Fermi level falls within this superposition zone. As a result, there is an absence of bandgap and VO_2 shows the conductivity of a metal. In the insulating phase, the $d_{||}$ band is split into two bonding and antibonding states, while the d_{π}^* band is moved upward the Fermi level, yielding to the formation of a bandgap E_g of $\sim$ 0.6 - 0.7 eV.

[7] Dimerization qualifies the process of joining two entities by a chemical bond (in this case, two V atoms forming a V-V dimer).

While VO_2 is often referred to as a Mott insulator, Mott's theory is only one of the two major theories regarding the physical mechanism causing the transition between both states [114]. The first theory, or *Peierls* model, states that it is the structural phase transition which drives the IMT transition and band structure change, through lattice distortion and associated electron-phonon interactions. This theory was first introduced by Goodenough [115] in 1971. In his work, the transition is induced by lattice distortion. The insulator behavior is attributed to the dimerization of V-V bonds which show two distinct lengths (3.16 and 2.62 Å). Since the d-electrons of the vanadium atoms are bound to the V-V bonds, this results in the in the splitting or localization of the *d*-orbitals, hence the insulating property. On the other hand, in the metallic state, the fact that the V-V distance (2.88 Å) is smaller than the critical interaction distance (2.94 Å) causes the *d*-orbital electrons to be shared by all the vanadium atoms, resulting in the metallic behavior [104]. The second theory, brought by Mott in 1975 [116], elaborates that lattice distortion is not sufficient to trigger the IMT. Mott attributes the phase transition to electron-electron interactions (correlation). In this theory, the insulator-to-insulator transition IMT (referred to as the *Mott-Hubbard* IMT) takes place when electrical carriers reach a critical density n_c, upon which Coulomb repulsion will screen the electrons away from the ion cores, and the system will become metallic. Band structure change takes place simultaneously. In such theory, the structural phase transition (SPT), i.e. the change in crystalline structure, is a *consequence* of the IMT. Whether the IMT is driven by lattice distortion (Peierls model) or electron-electron interactions (Mott-Hubard model) has been a long standing debate. As the transition takes place on ultrafast timescales, it is hard to decorrelate structural and electronic mechanisms. Several works have however identified the existence of an intermediate monoclinic, metal-like phase occuring during the IMT [117–119], hinting a cooperation between the Mott and Peierls mechanisms. In general, most recent works tend to reconciliate both theories by describing "Mott-assisted Peierls" and "Peierls-Mott" transitions.

The most commonly observed IMT is triggered by heating a VO_2 thin film, typically a few hundreds of nanometers-thick. Such thin films can be deposited through various techniques, the most common including sputtering, Pulsed-Laser Deposition (PLD), and Atomic Layer Deposition (ALD). They can be polycrystalline of randomly oriented or epitaxial nature, depending on the chosen substrate. The transition is accompanied by a drastic reduction in the thin film resistivity (shown in Fig. 1.19), characterized by the resistivity ratio ρ_{OFF}/ρ_{ON} which can reach several orders of magnitude, as high as 10^5 [120]. The hysteretic nature of the transition can be understood intuitively from the energy landscape model, as IMT and MIT transitions are associated to different heights of energy barrier jumps [121]. Such transition properties can also be observed in other structures, such as VO_2 nanowires, nanoparticles, and various device topologies.

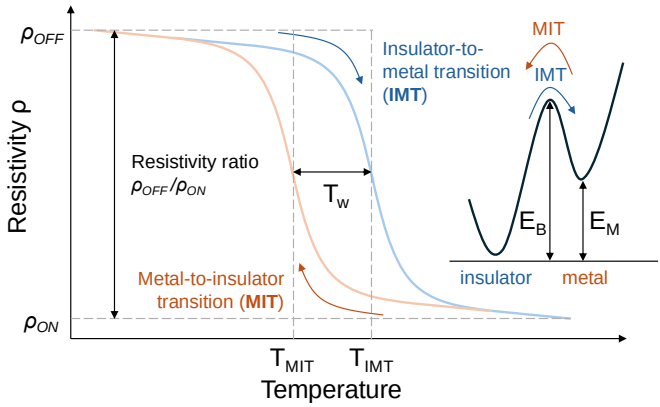


Fig. 1.19 Thermally-triggered hysteretic transition typically observed in a VO_2 thin film resistivity (*y axis is logarithmic*). Inset shows an energy landscape model giving an intuition for the cause of the hysteretic behavior.

Modulating the thermally-triggered VO_2 transition

The transition properties (transition temperature, hysteresis width and resistivity ratio) can be modulated through different parameters. On one hand, these parameters can be *intrinsic* to the VO_2 thin film and set during its microfabrication (Fig. 1.20a-d), such as choice of substrate [122, 123], film thickness and grain size [124, 125] (in the case of polycrystalline films), deposition technique, oxygen content or metallicity [126,127], presence of defects like oxygen vacancies [128–131], doping [132], and residual stress [133, 134]. Among these different parameters, some are highly correlated and can difficultly be studied independently (there will thus be interactions between certain of the modulation mechanisms shown in Fig. 1.20). For example, the type of substrate will impact residual stress [133], grain structure and the presence of interfacial defects. Grain size will increase with film thickness [125], and defects and oxygen vacancies will be present in higher concentration in small-grains films [135, 136]. Interestingly, transition properties can also be modulated *in situ* by external stimuli (Fig. 1.20e-i), including strain [137–139], optical [68, 140, 141], electro-chemical [122, 142, 143], magnetic [144] or electrical [134, 145, 146]. The physical phenomena behind such modulations depend on the specific stimulus. In the case of strain, it can be attributed to a reduction (resp. increase) of bandgap in the case of compressive (resp. tensile) stress [137]. Regarding UV light, light induced oxygen-vacancies have been identified as explaining hypothesis [68, 140, 147]. Electro-chemical modulations involve field-induced oxygen vacancies [122]. Magnetic modulation has been obtained through hydrogen treatment, inducing ferromagnetically coupled dimers [144]; magnetic properties can also be attributed to oxygen vacancies [104]. Electrical modulation is achieved by controling carrier density through an external electric field [145].

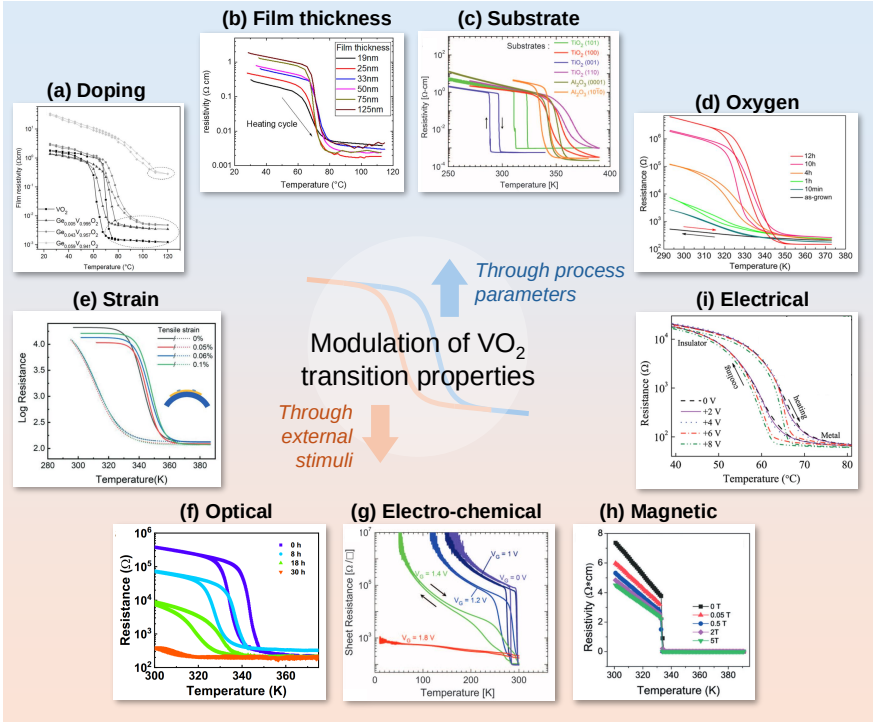


Fig. 1.20 Modulation of VO_2 thermally-triggered phase transition through process parameters (a)-(d) and external stimuli (e)-(i). **(a)** Modulation through germanium doping of VO_2 thin films [132]. **(b)** Modulation through variation of VO_2 thin films thickness [125]. **(c)** Modulation by growing identical VO_2 thin films on various orientations of TiO_2 and Al_2O_3 substrates [122]. **(d)** Modulation through introduction of oxygen vacancies by low-temperature annealing process of VO_2 nanonets [130]. **(e)** Modulation through tensile strain applied on a VO_2 thin film [137]. **(f)** Modulation by irradiating VO_2 thin films under UV light for increasing periods of time [68]. **(g)** Modulation through oxygen-vacancies creation induced by liquid ionic gating in a three-terminal VO_2 device [122]. **(h)** Modulation through application of magnetic field accross VO_2 nanowires [144]. **(i)** Modulation through application of an external electric field across a VO_2 three-terminal device [145].

Triggering the VO_2 transition through various stimuli

Beside thermal triggering, the insulator-to-metal transition can also be *directly induced* from the various stimuli mentioned above (see Fig. 1.21). Strain-induced transition has been reported in VO_2 nanowires at strain levels of ∼2 % (Fig. 1.21a) [148]. Photoinduced transition has also been reported under femtosecond laser pulses in the visible domain (700 - 800 nm) (Fig. 1.21b) [112, 149–151]. In this case, the transition takes place through an ultrafast transition of similar femtosecond timescale, to an intermediate metallic-like monoclinic phase. It is induced by carriers generation and change in Coulomb interactions. This intermediate state is unstable, and the modification in lattice potential then drives the VO_2 into its metallic state [112].

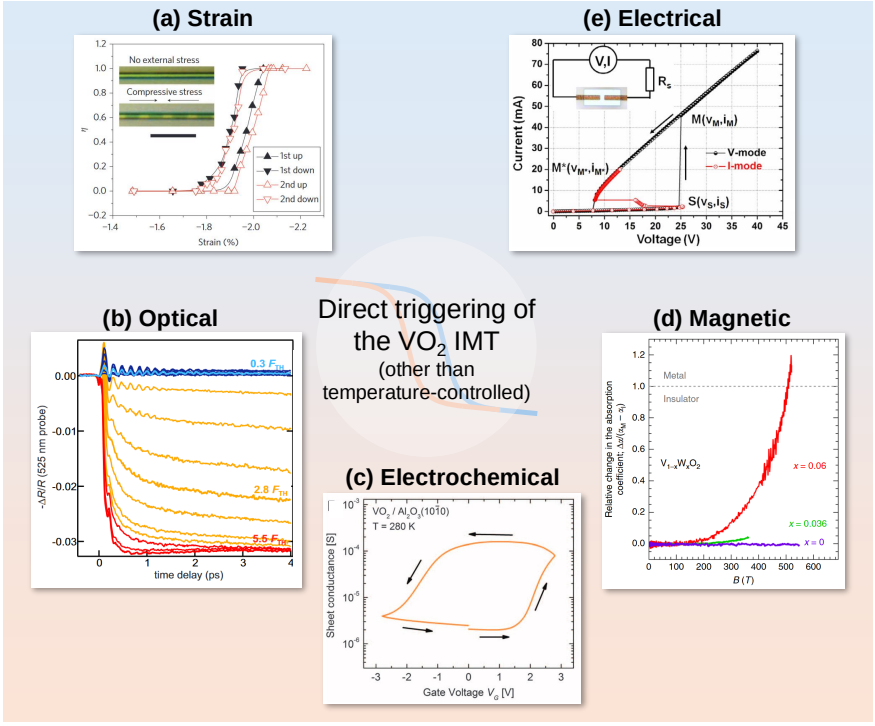


Fig. 1.21 Insulator-to-metal transition triggered through various stimuli, other than temperature-controlled. **(a)** Strain-induced IMT in a VO_2 nanowire. Y axis shows the fraction of metallic state in the wire, determined optically (yellow zones in the inset, whereas green zones correspond to the insulating state) [148]. **(b)** Photoinduced reflectivity change in a VO_2 thin film, induced at different fluences of laser excitation (λ = 800 nm). Reflectivity is used as a proxy to measure the phase transition. [112] **(c)** Dependence of a VO_2 three-terminal device sheet resistance on gate voltage, showing IMT induced through field-induced oxygen vacancies effect. The gate oxide is an organic ionic liquid [122]. **(d)** IMT induced through magnetic field in a W-doped VO_2 thin film. Change in absorption coefficient is used as a proxy to track the IMT [152]. **(e)** IMT induced by an applied voltage (black) and applied current (red) in a two-terminal VO_2 device [153].

In the UV domain, insulator-to-metal transition can be triggered through the generation of oxygen vacancies, which is a slower process taking place in hours timescale [68, 140]. The insulator-to-metal transition can also be directly triggered by field-induced oxygen vacancies, using ionic liquid as gate insulator [122] in a VO_2 three-terminal device (Fig. 1.21c). Regarding magnetic stimuli, Matsuda et al. have shown that a sufficiently high magnetic field could trigger IMT in a W-doped VO_2 film, and attributed it to the a spin effect that induced delocalization of electrons in V ions (Fig. 1.21d) [152]. Finally, the insulator-to-metal transition can also be triggered through an applied voltage in three- or two-terminal devices. Three-terminal devices have a structure akin to field-effect transistors (FET),

in which VO_2 is used as the channel material. Contrary to electrochemical triggering, IMT doesn't involve chemical reactions and is triggered by electrostatical screening through Coulomb interactions, so that the bandgap collapses and the channel becomes metallic [114]. The two-terminal device topology corresponds to the resistive-switching devices introduced in Section 1.2.2. In such devices, IMT can be triggered through a debated combination of Joule heating and electrical effects. We take a closer look at them in the next section.

1.3.2 Resistive-switching VO_2 devices

The first observation of the insulator-to-metal transition in a VO_2 two-terminal device through an applied voltage is traced back to the work of Stefanovitch [154] in 2000. Since then, a large body of works have repeated similar observations ([153, 155–158] to name a few). The typical I-V curve of a two-terminal VO_2 device is illustrated in Fig. 1.22a, along with its key parameters.

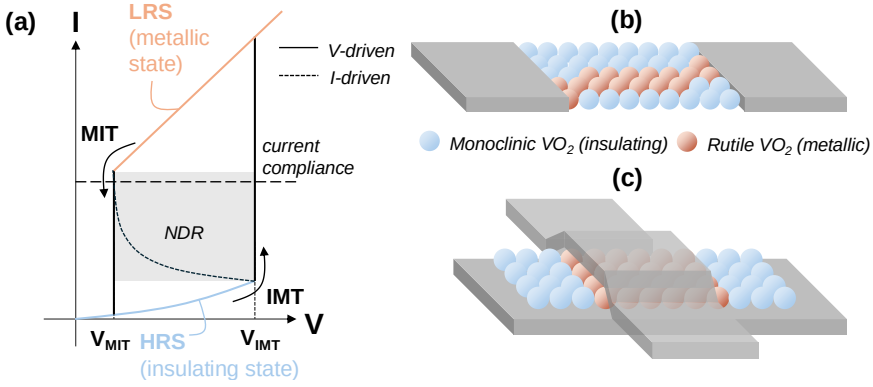


Fig. 1.22 **(a)** Typical voltage driven (solid line) and current-driven (dotted line) I-V curve of a VO_2 two-terminal device. Two possible structures of VO_2 two-terminal devices: **(b)** planar and **(c)** vertical or so-called "crossbar". Device structures illustrations are inspired from [159].

Similarly to the thermally-driven transition, the I-V curve shows an hysteresis, with two distinct threshold voltages V_{IMT} and V_{MIT} associated to the two transitions. The transition can be triggered in voltage- or current-driven mode. In voltage-driven mode, a compliance is typically applied to protect the device from overheating, as the transition is quite abrupt. The current-driven transition shows a negative differential resistance (NDR) region, which is related to the superlinear dependence of VO_2 resistivity in temperature [160]. The exact phenomenon behind the electrically triggered IMT is still under debate. A first theory attributes it to the field induced direct injections of holes or electrons, as initially proposed by Stefanovitch [154]. Other works suggest that the dominant mechanism is Joule

heating, through which portions of the VO_2 device would locally be brought to the transition temperature [161, 162]. While both purely thermal and purely field-induced behaviors have been observed in different VO_2 devices topologies (thin films and three-terminal devices, respectively), in two-terminal devices, the two phenomena are expected to be strongly coupled. Which mechanism is the dominant one depends highly on the device topology and on the properties of the synthesized VO_2.

Two terminal devices can be fabricated through two main topologies, illustrated in Fig. 1.22 vertical and planar. The vertical topology has advantages in terms of scalable integration in crossbar arrays (introduced in Section 1.1.1). Vertical devices have also been claimed to induce less device-to-device variability [163]. Indeed, in planar devices, transition has been shown to take place through a filamentary process [121, 164, 165], as illustrated in Fig. 1.22b. Once the applied voltage is sufficiently high, certain isolated domains inside the VO_2 channel will undergo IMT. Voltage drops will redistribute across these domains neighbours, precipitating the formation of a metallic filament. In certain planar devices, excessive heating can take place during the first transition, as all current is redirected towards the filament, locally resulting in very high current densities [120]. This can permanently alter the grain structure and I-V curve of the device [166]. After this first cycle - sometimes referred to as electroforming -, transition properties can still persist, but device-to-device variability will be induced due to change in the defice ON and OFF state resistances. In vertical devices, transition is expected to take place in a more uniform fashion as shown in Fig. 1.22c. The device is thus expected to experience lower temperatures during transition, removing the electroforming behavior and associated device-to-device variability [159]. However, planar devices offer other advantages, including a simpler fabrication process, and possibility to deposit VO_2 on a variety of substrates (indeed, obtaining a high-quality growth of VO_2 on a metallic substrate can be challenging [167]).

In the previous section, we have shown that the thermally-triggered transition in VO_2 can be modulated by a variety of stimuli. As discussed above, in two-terminal devices, the voltage-triggered transition is due to a combination of thermal effects (Joule heating) and electrical effects (field-induced carriers). Since the mechanisms inducing the transition are similar to those at play in a temperature-controlled transition, it is thus expected that the I-V curve of VO_2 two-terminal devices could also be modulated by the stimuli explored above.

This claim has been confirmed for several stimuli, including strain [148, 168] (Fig. 1.23a), optical stimuli in the visible domain (450 to 900 nm, Fig. 1.23b) [91,169] and UV domain [170,171], and temperature [156,172] (Fig. 1.23c). Regarding strain, the proposed physical mechanism is the emergence of metallic domains due to lat-

tice deformation [148]. Modulation in the UV domain is attributed to hole-trapping effect on the surface of VO_2 micro-wires, which are increased due to oxygen species under UV light [170,171]. However, these results are reported on devices which do not show sharp voltage-triggered transition, preventing their use as spiking neurons. In the visible domain, sensitivity is lower and attributed to coupled thermal and optical effects [169]. Thermal modulation is the stimuli for which the most observations were reported. It is also the most well understood as it follows directly from the strong coupling between thermal and electrical effects during the voltage-triggered IMT. Thermal sensitivity has been demonstrated from room-temperature up to the transition temperature.

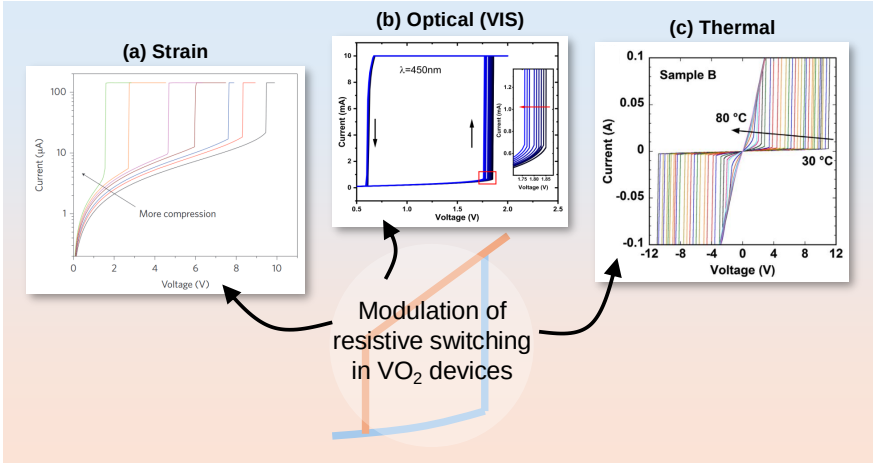


Fig. 1.23 Modulation of the voltage-triggered transition properties of VO_2 two-terminal devices under different external stimuli. **(a)** Compressive strain (up to ~ 2 %) modulates the IMT in VO_2 single crystal nanowires [148]. **(b)** Modulation of the IMT in a VO_2 polycrystalline two-terminal device under visible light at different power (wavelength is 450 nm) [169]. **(c)** Modulation of the IMT in a VO_2 polycrystalline two-terminal device in temperature range 30 - 80 °C [172].

1.3.3 VO_2 neuromorphic sensors

It is now clear that **the multi-responsiveness of VO_2 makes it an interesting candidate to design neuromorphic sensors**. These sensors can take the form of *sensitive synapses*, exploiting the dependence of VO_2 resistivity to various stimuli (see Fig. 1.20). Alternatively, VO_2 two-terminal devices can be used to design *sensitive spiking neurons*, thanks to the modulation of their voltage-triggered resistive-switching properties under certain stimuli (see Fig. 1.23). In other spiking neurons implementations, transduction is performed by an external sensor, while the VO_2 two-terminal devices solely implement the spike generation function. We will quickly review the state of the art of these different implementations.

VO_2 sensitive synapses

As introduced in Section 1.2.1, sensitive synapses based on non-volatile memristors show variation of their resistance under an applied stimulus, and implement various learning rules with respect to the properties of input spikes (duration, frequency, ...). A number of light-sensitive synapses have been realized using VO_2 devices, and operate in various wavelength ranges. These implementations are summed up in Table 1.1.

Table 1.1 State-of-the art of VO_2 sensitive synapses

Reference	Type of device	Neuron material	Wavelength	Synaptic behavior	Learning rule
Li 2022 [68]	Three-terminal device	VO_2	UV (375 nm) VIS (450 nm, 532 nm, 650 nm)	LTP (UV) STP (VIS)	SNDP (UV), SDDP (UV)
Niu 2024 [173]	Two-terminal device	VO_2	UV (275 nm, 365 nm)	LTP (both wavelengths)	/
Wang 2024 [174]	Two-terminal device	HfS_2 /VO_2 heterostructure	VIS (405 nm)	STP, LTP	PPF, SRDP, SDDP, SNDP
Yu 2024 [175]	Three-terminal device	Graphene/VO_2 heterostructure	VIS (405 nm)	STD, LTD	PPF, SRDP, SDDP
Guo 2024 [176]	Two-terminal device	VO_2 /MoO_3 heterostructure	UV (355 nm) VIS (405 nm) NIR (980 nm, 1342 nm, 1550 nm)	STP, LTP (all wavelengths)	PPF (1550 nm), SDDP (1550 nm)

Synapses based solely on VO_2 show non-volatile variation of resistance (i.e. long-term potentiation or LTP) under UV light, owing to photo-induced oxygen vacancies evoked previously [68, 173]. Volatile synaptic behavior, i.e. short-term potentiation (STP), has also been observed in VO_2 in the visible wavelengths. It has been attributed to photo-generated electron-holes pairs which recombine rapidly when light is removed [68]. Non-volatile modulation of VO_2 devices in larger wavelengths than UV is achievable when combining VO_2 to other materials within heterostructures. Indeed, devices involving HfS_2 [174], graphene [175], and MoO_3 [176] showed LTP respectively at 405 nm for the first two, and at 355, 405, 980, 1342 and 1550 nm for the latter. VO_2 can thus effectively be used to implement

light-sensitive synapses from the UV to the near-IR (NIR) spectra. Various learning rules are implemented depending on the wavelength and specific devices, including paired-pulse facilitation (PPF), spike-number dependent plasticity (SNDP), spike-rate dependent plasticity (SRDP), and spike-duration dependent plasticity (SDDP). In such sensitive synapses, voltage-pulses can also be used to trigger plasticity behavior or to reset light-induced memory states. They can be applied either in two- or three-terminal device configurations.

Additionally, a few implementations of "computational" VO_2 synapses (whose hardware does not perform stimulus transduction) are worth mentioning. They use three-terminal topologies with liquid ionic gating, where non-volatile potentiation is triggered electro-chemically through gate voltage pulses which induce proton-doping [177–179]. Finally, among other vanadium oxide stoichiometries, V_3O_5 neurons enable visible light induced non-volatile modulation [180].

VO_2 sensitive spiking neurons

Table 1.2 State-of-the art of VO_2 sensitive spiking neurons

Reference	Neuron material	Stimulus	Sensing	Number of devices	Encoding	Device topology
Rosca 2021 [91,169]	VO_2	RF (60 - 110 GHz) Optical (500 - 700 nm)	VO_2 memristor itself	3 (capacitor, transistor, memristor)	Rate coding	Planar
Duan 2022 [181]	VO_2	Temperature (284 - 306 K) Strain	VO_2 memristor itself (temperature) External piezoresistive sensor (strain)	2 for temperature (resistor, memristor) 3 for strain (resistor, memristor, piezoresistive sensor)	Rate coding	Planar
Fang 2022 [95]	VO_2	Pressure (4 - 11 kPa)	External pressure sensor	3 (capacitor, memristor, pressure sensor)	Rate coding	Vertical
Yuan 2022 [96]	VO_2	Pressure (0 - 0.98 N) Optical (broadband) Temperature (30 - 100 °C)	External sensors	5 (capacitor, two resistors, memristor, external sensor)	Rate coding	Planar
Han 2022 [92]	VO_2	Temperature (20 - 40 °C)	VO_2 memristor itself	2 (capacitor, memristor)	Rate coding	Vertical
Qaderi 2023 [93]	VO_2	RF (130 - 160 GHz)	External RF ports together with VO_2 memristor itself	5 (capacitor, transistor, memristor, two RF ports)	Rate coding	Planar

Table 1.2 sums up the state of the art of VO_2 spiking sensors. While they all take care of the spike generation function, the different works can be divided between those in which stimulus transduction is performed by the VO_2 memristor itself, or by an external sensor. In the first case, spiking sensors have been reported for RF (radio-frequency) radiations [91, 93], visible light [91, 169], and temperature [92, 181]. Transduction mechanisms are similar to those evoked in Section 1.3.2. Regarding transduction of RF radiations, the RF signal is routed towards the VO_2 two-terminal devices using RF ports and a co-planar waveguide; transduction mechanism is expected to take place through a combination of field-induced carriers and Joule heating [93]. Other implementations use external sensors for the transduction of temperature [96], pressure [95, 96] or strain [181], and broadband white light [96]. Both planar and vertical topologies are reported.

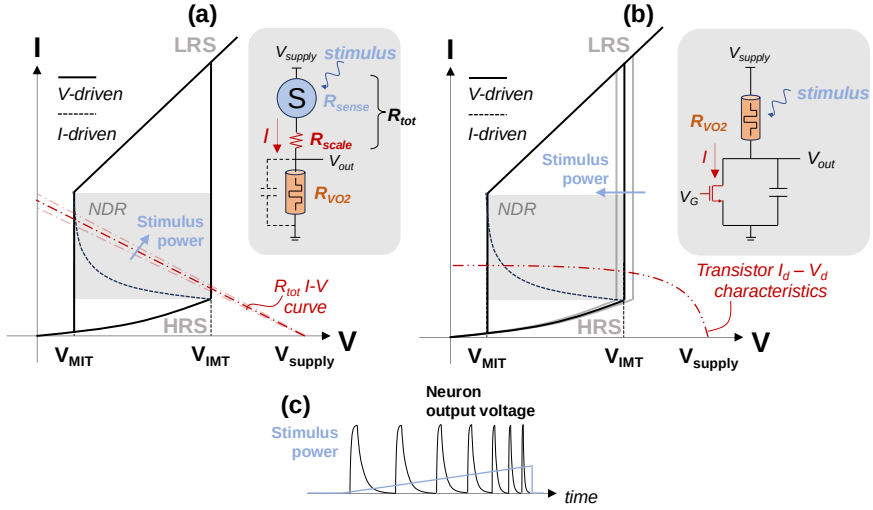


Fig. 1.24 Examples of two architectures of sensitive spiking VO_2 neurons. **(a)** Transduction is performed *via* an external resistive sensor R_{sense}. Spiking frequency is modulated by the change in the sensor output current. A scaling resistor R_{scale} is used to ensure that the current falls within the device NDR, inducing spiking behavior. The capacitor is intrinsic to the VO_2 device. **(b)** Transduction is performed by the VO_2 memristor itself. Current is fixed, and set through the gate voltage V_G of a NMOS transistor set in saturation regime, so that it falls within the device NDR. An external capacitor is added. Spiking frequency is modulated by the change in the VO_2 memristor I-V characteristics under the stimulus. **(c)** In both case, the stimulus is encoded through rate coding, as an increase in stimulus power results in spikes of smaller duration.

Regarding spike generations, all implementations follow a similar principle. To enter spiking regime, a current must be forced within the negative differential resistance range of the device (see Fig. 1.24), which induces an astable behavior in

which the VO_2 two-terminal device oscillates between its metallic and insulating state, resulting in output voltage spikes. The duration of a given spike is dictated by the circuit RC constant in the device metallic and insulating regimes. The capacitor can be either intrinsic to the VO_2 device, or external. To make sure it falls within the NDR region, the current level can be set either using an external resistor or a transistor working in saturation regime. In the latter case, the I-V curve is flatter, offering finer control. When transduction is performed by the VO_2 device itself, spiking frequency will be modulated by the variation in the VO_2 device I-V characteristics under the stimulus (as evoked previously in Section 1.3.2). Meanwhile, the current fed to the device is kept constant. When transduction is performed by an external sensor, spiking frequency is directly modulated by the sensor current which is fed to the VO_2 device. Depending on the different cases mentioned above, the number of circuit elements can vary from 2 to 5. Two examples of spiking VO_2 sensor designs are illustrated in Fig. 1.24a-b. The different characteristics of the design (intrinsic or external capacitor, transduction through the VO_2 memristor or an external sensor, current set through a scaling resistor or a transistor) are mixed and matched in the different works reported in the literature. In all cases, these spiking sensors perform *rate coding* of the input stimulus, in an *all-or-nothing* regime. Indeed, they are either completely OFF (when the current does not fall within the device NDR), or always spiking, with the stimulus being encoded into the *duration* of the consecutive spikes.

A few additional architectures are worth mentioning; while they do not directly implement sensing, they have been used as encoders in sensing scenarios such as epileptic seizure recognition [103], spatio-temporal sound source differentiation [182], and retina emulation [86]. In these cases, the stimulus is encoded into spikes using a software before feeding it to the VO_2 neuron. These implementations have more complex circuit architecture but are able to implement a larger variety of neuron behaviors, such as spike frequency adaptation, spikes with various polarity, and tonic bursting.

1.4 Key messages

Several key messages can be retained from this introductive chapter.

Spiking Neural Networks could offer energy-efficient processing of data when such data features a prominent temporal dimension. In the realm of biomimetic or neuromorphic hardware, SNNs hold theoretical promises of higher energy efficiency compared to ANNs to perform information processing in at-the-edge sensing scenarios of dynamic stimuli. This theoretical advantage relies on two main paradigm shifts: memory and computation co-location, and spike-based temporal encoding that leverages event sparsity. However, this superiority remains to be proved through benchmarking of hardware platforms using sensory data with a prominent temporal dimension, encoded in the spiking domain.

Spiking Neural Networks require neuromorphic sensors, and existing solutions have limitations. In at-the-edge sensing scenarios, SNNs need stimulus information to be encoded in the spiking domain. This encoding is currently either performed using software, ADC and DSP units, or CMOS-based event cameras. The first two existing options induce large overhead. Event cameras suffer from scalability concerns (a single pixel has a large area as it involves several transistors, which are also associated to a large static power consumption) and are currently limited to visual stimuli, as they are constrained by the transduction capabilities of silicon.

Memristive devices allow for compact implementation of spiking sensors. To overcome the above mentioned limitations, memristive devices based on various emerging materials have been proposed as an alternative to design neuromorphic sensors compatible with SNNs. They offer promises of high scalability thanks to their compactness.

- *Non-volatile memristors* have been proposed to implement *sensitive synapses*, whose resistance can be durably altered under electrical spikes and various stimuli.
- *Volatile memristors*, also referred to as *resistive-switching devices* thanks to their ability to switch reversibly from a high- to low- resistive state under electrical stimulation, have been proposed to implement *sensitive spiking neurons* enabling compact spike-encoding of stimuli.

Among these memristors, those based on Vanadium dioxide offer both neuromorphic functionalities and stimuli versatility. Indeed, vanadium dioxide (VO_2) stands out as an interesting candidate to design neuromorphic sensors, thanks to

several key properties.

- First, its reversible insulator-to-metal transition that takes place close to room-temperature, enabling the design of voltage-triggered resistive-switching devices with volatile memristive properties. This allows it to emulate complex neuron models such as LIF and H&H.
- Second, this transition can be *modulated* and even *directly triggered* through a wide variety of stimuli, including thermal, electrical, optical, magnetic, electro-chemical and strain. This opens the possibility to embed transduction of different stimuli directly within VO_2 devices. However, cross-sensibility issues could arise.

Vanadium dioxide neuromorphic sensors recently started to develop. Leveraging these two properties, several implementations of neuromorphic sensors based on VO_2 have emerged in the four previous years:

- VO_2 enabled the realization of *light-sensitive synapses* in wavelengths ranging to UV to near IR.
- VO_2 enabled the realization of *spiking neurons sensitive to temperature, visible light, strain, pressure, and RF radiations*. Among these implementations, direct embedding of sensing within the VO_2 memristor has so far been achieved for temperature and visible light.

The research field of VO_2-based neuromorphic sensors is very young, but holds promises for the realization of compact, scalable spiking neurons sensitive to a wide variety of stimuli. However, their development faces several challenges, and this thesis aims at addressing some of them. First, Chapter 2 will tackle the intrinsic stochasticity of the resistive switching phenomenon, and its impact on cycle-to-cycle variations in VO_2 two-terminal devices electrical properties. Chapter 3 will then investigate the consequence of these variation on the dynamic operation of VO_2 spiking sensors, and on their resulting resolution. In its second part, Chapter 3 will discuss the lack of sparsity in rate-coded VO_2 spiking sensors, and propose an alternative encoding technique. Finally, Chapter 4 will evaluate the direct environmental footprint of their fabrication, to propose eco-designed alternatives. Note that not all of the limitations of present VO_2 spiking neurons are addressed in these works. Some of them, such as their reliability issues and their relatively high static power consumption, are addressed in the conclusion and perspectives of this work.

2

Understanding the stochasticity of resistive switching in VO_2

We can never achieve absolute truth,
but we can live hopefully by a system of calculated probabilities.
– Agnes Meyer Driscoll

STOCHASTICITY is inherent to the operation of resistive-switching devices, and results in cycle-to-cycle variations of their electrical properties. VO_2 two-terminal devices are no exception, as the electrical characteristics of a given device vary from one insulator-to-metal transition to the next one. Depending on their specific application in the neuromorphic research field (and beyond), such variations can be either beneficial or detrimental. Since the mechanisms behind these variations are not yet fully understood, there is currently a **lack of guidelines to design VO_2 memristors with minimized or maximized cycle-to-cycle stochasticity**. In response, this chapter explores the following research question:

How can specific material properties and device design parameters impact the resulting VO_2 two-terminal devices cycle-to-cycle stochasticity?

Theories presently put forward in the literature highlight the role of cycle-to-cycle variations in individual crystallites in the device channel. Based on these theories, we make the hypothesis that increasing the number of grains within the device channel should mitigate cycle-to-cycle stochasticity. To challenge and experimentally verify this hypothesis, we fabricate VO_2 two-terminal devices with varied grain size and channel length, hence varying the number of grains inside the channel. We then study the impact of these two parameters on the cycle-to-cycle variations of the devices electrical properties.

This chapter starts from a material science investigation, detailing the synthesis of VO_2 thin films. We **identify key process parameters allowing to tune grain size, while monitoring their impact on the films transition properties**. We then present the process flow leading to the fabrication of VO_2 two-terminal devices, along with their electrical characteristics. Next, we detail the main present explanatory theories behind their stochastic behavior. To investigate the impact of process and design parameters on device cycle-to-cycle stochasticity, we pick two specific VO_2 thin films with large difference in grain size, and study the behavior of the two-terminal devices fabricated from these films. **We show that grain size and device channel length have large effects on cycle-to-cycle stochasticity**. We also highlight the presence of bi-modal distributions in the electrical properties statistics of small-length devices, a phenomenon that has been sparsely observed but never explained. By analyzing the evolution of the electrical parameters towards the insulator-to-metal transition, we show that these modes could correspond to distinct percolation paths and filaments. In conclusion, we derive key guidelines for the design of VO_2 devices with minimized or maximized cycle-to-cycle stochasticity. Results from this chapter are adapted from our *Nano Letters* publication [JP1].

2.1 VO_2 thin film synthesis

In this section, we first detail the synthesis technique for VO_2 thin films which was developed in the framework of this thesis, including the choice of substrate and the choice of deposition technique. We then explore two key parameters allowing to tune grain size: oxygen concentration and annealing temperature. We also show that oxygen concentration is a key calibration parameter to obtain the right vanadium oxide stoichiometry.

2.1.1 Choice of substrate

Synthesizing vanadium dioxide thin films starts from the choice of *substrate*, which can greatly impact the films residual stress, crystallinity, interfacial defects and resulting transition properties [122, 123, 183]. In the framework of this thesis, we chose to work with 3 inches **(100) oriented silicon substrates** on which we perform a wet thermal oxidation step to obtain a **200 nm-thick SiO_2 layer**. This choice is motivated by several factors. First, previous work showed that polycrystalline VO_2 thin films deposited on Si/SiO_2 substrates allowed to obtain a satisfying $\sim 10^3$ OFF/ON resistivity ratio through various deposition techniques [184–187], enabling the design of functional resistive-switching device. Indeed, a sufficiently high OFF/ON ratio is required for the devices to show a negative differential resistance region [188], which is itself a condition to observe self-sustained spiking

behavior [166]. Other types of substrates such as Al_2O_3 allow to obtain even higher OFF/ON ratio (up to 10^5 [120]), thanks to their lattice constants being closer to that of VO_2, allowing to grow epitaxial VO_2 thin films. However, a too high OFF/ON ratio is likely to induce irreversible damages upon transition, due to the high current density within the metallic filament [120]. In addition, Si/SiO_2 have the advantage of CMOS compatibility, which could become relevant if VO_2 neurons were to be integrated in architectures containing CMOS transistors[1].

2.1.2 Deposition technique

Another important factor impacting the properties of VO_2 thin films is the *deposition technique*. In this thesis, we chose to use **DC reactive magnetron sputtering** (illustrated below in Fig. 2.1), which is one of the most widely used technique for the synthesis of VO_2 thin films for neuromorphic applications [86, 92, 95, 182, 189].

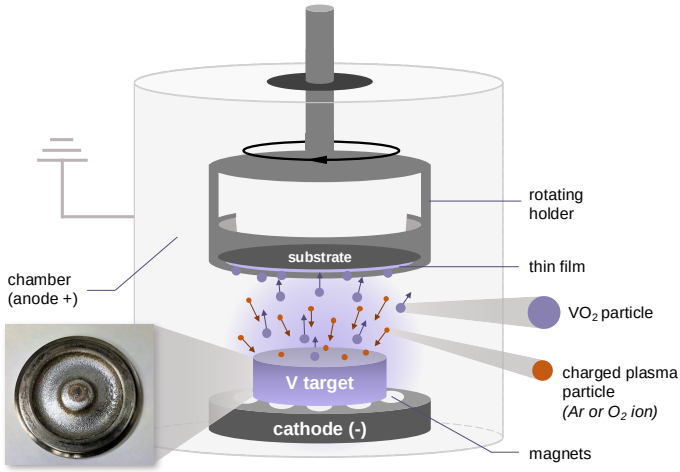


Fig. 2.1 Illustration of a reactive magnetron sputtering chamber. Image of the sputtering target adapted from [190].

It possesses several assets, including a high deposition rate, and good deposition uniformity across a large substrate area. Pulsed laser deposition

[1]However, this claim depends on the specific process flow which would be selected for CMOS co-integration. One could imagine depositing a thin Al_2O_3 buffer layer through e.g. ALD at the back-end of the CMOS process flow, which would also be CMOS-compatible. However, it would be less straightforward than using SiO_2, since the latter is already present in the CMOS process.

(PLD) is another popular technique yielding VO_2 thin films with similar properties [91, 93, 96, 168, 181], but provides less uniformity as the deposition area is typically smaller than the substrate. PLD however enables better control of the film composition, since the target stoichiometry is preserved during the deposition. In DC reactive sputtering, a plasma is created by applying a voltage between a cathode and an anode. The cathode is usually located behind the target, a small cylinder of the specific material to be deposited (**metallic vanadium in our case**). The anode corresponds to the chamber walls. The chamber is first pumped to low pressure, then filled with an inert gas such as argon, before the plasma is ignited. The high electric field accelerates the electrons present in the gas, which will collide with gas atoms and tear off their electrons through electrostatic repulsion, creating positively charged Ar^+ ions. These positively charged ions are then accelerated towards the cathode and will in turn collide with the target. The substrate is placed face down on a rotating holder. Atoms at the surface of the target will be ejected and subsequently deposited on the substrate, creating a thin film. In *reactive* sputtering, a reactive gas (**oxygen in our case**) is introduced in the chamber to contribute to the plasma. Oxygen ions will combine with vanadium atoms sputtered from the target to form vanadium oxide species. The oxygen ions will also directly react with the target and form a surface oxide. The specific stoichiometry can be controlled through the oxygen concentration in the chamber. In *magnetron* sputtering, the plasma is closely confined near the target surface through strong magnets. This techniques yields higher density plasma with higher deposition rate, and improves film quality by preventing accelerated plasma particles to collide with it and damage it. However, target will wear quicker over time, as it will thin out in its center at the place where magnets are located. We chose to use *DC* over *RF* sputtering as the former offered a more stable plasma and faster sputtering rate.

Table 2.1 Process parameters related to the sputtering of VO_2 thin films studied in this work

Parameter	Value
Target	Vanadium 99.95 % purity
Plasma power	200 W
Target cleaning under Ar	5 min
Target cleaning under Ar/O_2	5 min
Base pressure	3e-7 Torr
Deposition pressure	5e-3 Torr
Deposition duration	17 minutes 20 seconds
Total gas flux (Ar + O_2)	50 sccm
Oxygen concentration	7 to 14 %

Process parameters related to the synthesis of VO_2 thin films through DC reactive magnetron sputtering realized in this work are detailed in Table 2.1. Deposition starts with a target cleaning under argon, followed by a conditioning step

under mixed argon-oxygen atmosphere. The goal is respectively to clean the target surface from previous depositions, and to reset it back to a controlled oxidation state. During these two steps, a shutter prevents the sputtered material to reach the sample; this shutter is opened at the beginning of the actual deposition step.

Table 2.2 Process parameters related to the annealing of VO_2 thin films studied in this work (sputtering equipment is an *AJA* model).

Parameter	Value
Base pressure	2e-2 Torr
Annealing pressure	1,5e-1 Torr
Gas flux	Ar, 30 sccm
Ramp duration	30 min
Annealing duration	60 min
Annealing temperature	300 - 650 °C
Cooling	Natural cooling under argon flux

The deposition step is performed at room temperature due to equipment constraints, and thus yields amorphous VO_2 thin films. To obtain a film with sufficiently high OFF/ON ratio, it is necessary to perform an **annealing step** in order to trigger crystallization [191]. We perform such step in a tubular oven with controlled pressure and atmosphere. All parameters are detailed in Table 2.2.

2.1.3 Impact of oxygen concentration

Stoichiometry

Oxygen concentration during the sputtering step heavily impacts the resulting properties of the vanadium oxide thin film, starting with its stoichiometry. As introduced in Chapter 1, there exist many different phases of vanadium oxide. In order to obtain VO_2, a **calibration** of the sputtering equipment is necessary to determine the exact oxygen range yielding VO_2. In reactive sputtering, the deposition rate varies with the oxygen concentration in the chamber, and serves as a qualitative indicator of the thin film oxidation state. It is expressed in nm/min and measured *in situ* using a quartz piezoelectric micro-crystal. A calibration curve can be produced by continuously increasing the oxygen concentration during a test deposition. As shown if Fig. 2.2, the sputtering rate first increases at low oxygen concentration, then starts dropping at a critical concentration $[O_2]_{crit}$. This phenomenon is detailed in [192]: below $[O_2]_{crit}$, all oxygen ions of the plasma are adsorbed by the vanadium atoms sputtered from the target. Above $[O_2]_{crit}$, the oxygen flow rate becomes higher than the vanadium atoms adsorption rate, and oxygen starts to react with the target, creating a layer of vanadium oxide at its surface. Oxidized targets have a lower sputtering yield, hence the drop in deposition rate.

We identified the $[O_2]$ range yielding specific stoichiometries by sputtering vanadium oxide thin films at different oxygen concentrations (indicated on Fig. 2.2). Deposition parameters are those specified in Table 2.1. We know from [193] that the VO_2 stoichiometry typically corresponds to chamber oxygen concentrations located around the midst of the drop in deposition rate. The range in which the deposition rate rises (at low oxygen concentrations) corresponds to VO, while V_2O_3 corresponds to the beginning of the drop of deposition rate. Saturation of this rate corresponds to the most oxidized phase, namely V_2O_5.

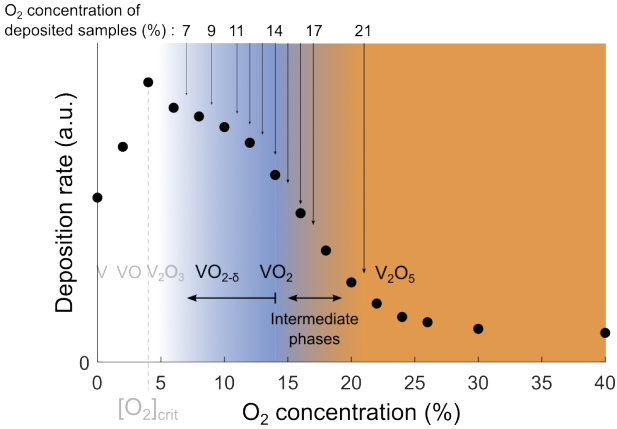


Fig. 2.2 Evolution of the sputtering rate with oxygen concentration in the chamber, measured by continuously increasing oxygen concentration during the deposition on a single dummy substrate. Individual films have latter been deposited at the specific concentrations highlighted by arrows. Stoichiometry indicated in black has been later deduced through Raman spectroscopy. Stoichiometry indicated in gray are deduced based on previous work performed on a similar equipment [193].

To assess the stoichiometries of these phases, we used **Raman spectroscopy**. Raman is a non-destructive spectroscopy technique allowing to retrieve information about the chemical structure and phase of a sample. It exploits the scattering of light (emitted by a monochromatic laser) by the sample crystal lattice. Most specifically, Raman leverages *Stokes and anti-Stokes Raman scattering*, an event upon which light is re-emitted at a higher or lower wavelength than that of the incident light. This process is inelastic: the lattice gains (resp. loses) part of the energy from (resp. to) the incident photon and becomes excited to a higher (resp. lower) vibrational level. In consequence, the photon has lost (resp. gained) energy and has a larger (resp. smaller) wavelength. Statistically, in a large group of crystal unit cells, Stokes scattering (resulting in a reflected photon with larger wavelength) is more likely. The so-called Raman shift in wavelength depends on the incident wavelength and on the different possible vibrational modes of the crystal lattice. This vibrational

mode is related to the angles and length of the crystal lattice bonds. This means each crystalline structure has its specific Raman signature with peaks of reflected light intensity located at various wavelength "shift" with respect to the laser incident wavelength (expressed in cm-1), corresponding to its different vibrational modes.

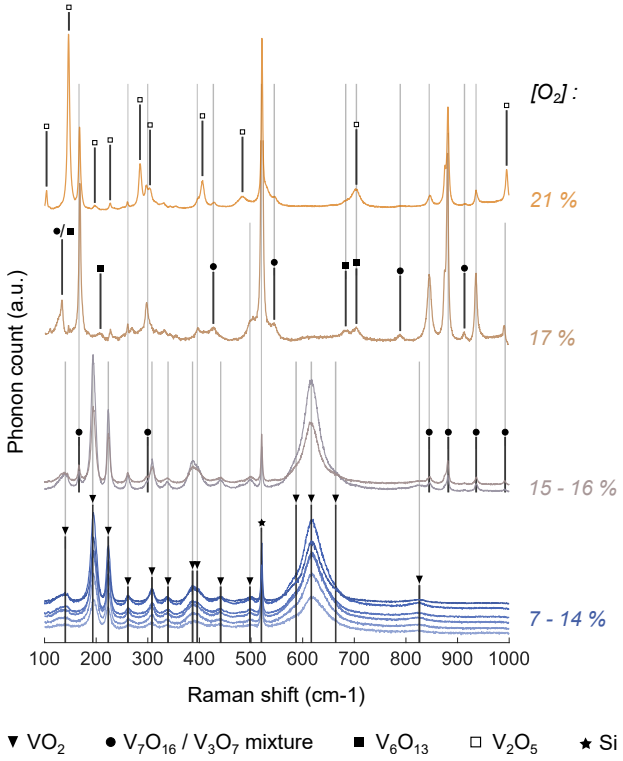


Fig. 2.3 Raman spectra of the vanadium oxide thin films sputtered at different O_2 concentrations, then annealed. The peaks attributed to the different stoichiometries are indicated. Spectra closely located on the y-axis are plotted with identical y-scale. Spectra were acquired with a blue laser of 488 nm wavelength, at 10 % of full laser power to avoid any heating effect, with 20 second acquisitions averaged over 3 measurements.

For their Raman peaks to be visible, vanadium oxide thin films need to be crystalline or polycrystalline. Indeed, molecules in amorphous films are randomly oriented, meaning the angle and length of their bonds will vary depending on their chemical interactions with their neighbours. As a result, Raman peaks will be broader in amorphous than in crystalline samples, often too weak to be detected. Performing thermal annealing of all films at 500 °C (with the process parameters detailed in Table 2.2) yields polycrystalline films due to the amorphous nature of

the SiO_2 substrate. The Raman spectra of these different films are shown in Fig. 2.3. Peak attribution has been performed by comparison to reference Raman spectra of films prepared similarly, i.e. by sputtering of a V target on a SiO_2/Si substrate [105]. In films deposited with oxygen concentration ranging from 7 to 14 %, all peaks can be attributed to the VO_2 stoichiometry. Peaks tend to become more intense and defined with respect to the background noise with increased oxygen concentration, with peak intensity achieved at 13 % [O_2]. In the 15 to 16 % range, new peaks corresponding to an intermediate phase between V_7O_{16} and V_3O_7 emerge. At 17 %, peaks corresponding to V_6O_{13} start to appear, while most peaks can still be attributed to the V_7O_{16} / V_3O_7 mixture. Finally, at 21 %, most of the peaks can be attributed to V_2O_5.

Grain size and transition properties

Oxygen content in VO_2 thin films can also heavily impact their transition properties. We thus take a closer look at the VO_2 films deposited with 7, 9, 11, 12, 13 and 14 % oxygen concentration. Top and cross-section scanning electron microscope (SEM) images of all 6 samples are shown respectively in Fig. 2.4 and 2.5. Grain size has been extracted from the top view pictures using X-ray diffraction (XRD) and the Scherrer equation [194], which allows to deduce mean crystallite size from the width of VO_2 XRD peaks:

$$t = \frac{K\lambda}{\epsilon \cdot cos\theta} \quad , \tag{2.1}$$

with K a shape factor usually equal to unity, λ the wavelength of the incident X-ray, ϵ the peak width at half of its maximum intensity (FWHM), and θ the Bragg angle. One witnesses a clear increase in grain size with oxygen concentration (Fig. 2.6).

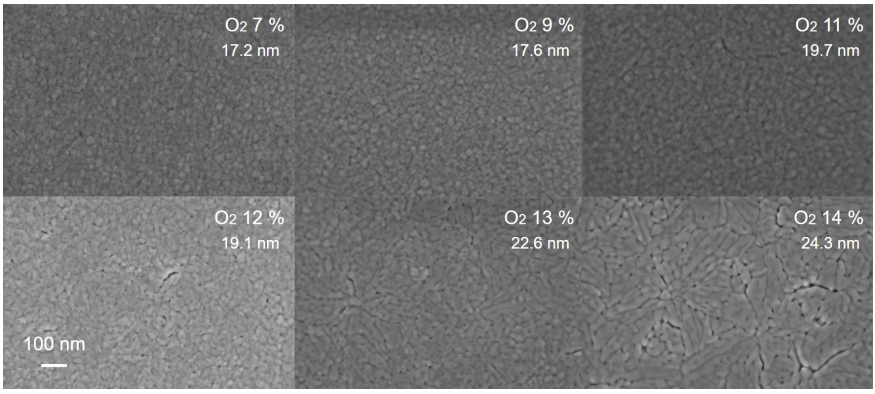


Fig. 2.4 SEM top view images showing the impact of the oxygen concentration during sputtering step on the grain size of the resulting polycrystalline VO_2 thin films (after 500 °C annealing). Grain size has been deduced through XRD. Scale bar is valid for all pictures.

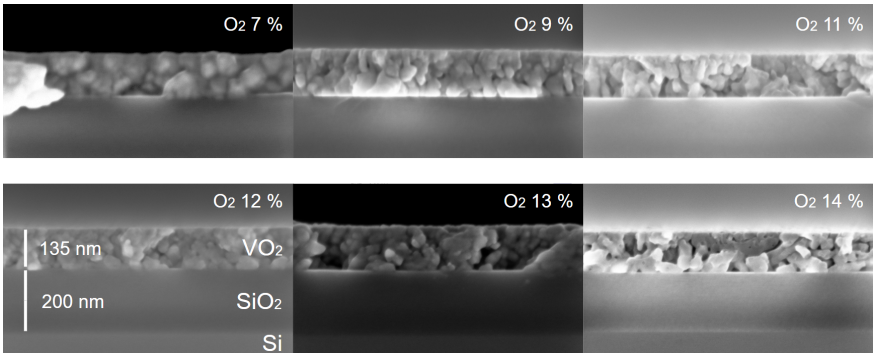


Fig. 2.5 SEM cross section images of the same samples. One can distinguish the polycrystalline VO_2 layer, on top of the SiO_2 200 nm-thick layer grown from the Si substrate. All VO_2 thin films have similar 135 nm thickness.

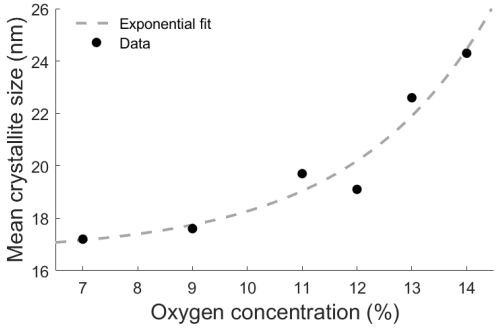


Fig. 2.6 Extracted mean grain size as function of oxygen concentration for the six deposited VO_2 thin films, along with exponential fit.

Oxygen concentration has rarely been reported as a mean to tune grain size in VO_2 thin films, as this effect is not spectacular (mean grain size remains in the same order of magnitude in the whole oxygen concentration range investigated). To understand this phenomenon, it is necessary to explain the mechanism behind thin film growth [195], illustrated in Fig. 2.7. Impinging particles ejected from the vanadium target will condense on the substrate and attach to it through *adsorption*. After losing part of their energy due to the collisions, the adsorbed atom (referred to as adatom) can travel on the substrate surface through diffusion [196]. They can then either be re-evaporated, or chemically bind with other adatoms deposited on the substrate (*chemisorption*). Successive chemisorption will create clusters of critical size (*nucleation*), which will latter grow to larger crystallites upon thermal annealing of the samples. The *supersaturation* is defined as the ratio of rate of impinging atoms R_i over the rate of re-evaporated species R_e.

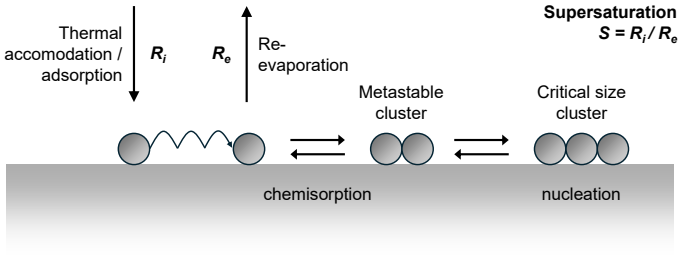


Fig. 2.7 The different steps of thin film growth. Adapted from [195].

At lower oxygen concentration, oxygen-deficient species could have lower probabilities to undergo chemisorption or chemical binding with other adatoms due to their sub-stoichiometric state. The probabilities of such particles to be re-evaporated would thus increase, reducing supersaturation and nucleus density. This hypothesis is also put forward by [127] in a study on the impact of oxygen flux in VO_2 thin films deposited through PLD. Another hypothesis could be related to the annealing step. During annealing, species will gain energy through temperature and diffuse to form larger clusters, yielding the polycrystalline structure. Oxygen-deficient species in films deposited at lower oxygen concentration could have smaller diffusion rates.

In addition to grain structure, we studied the impact of the oxygen concentration on the transition properties of the VO_2 thin films. We used the four-point measurement technique (see inset in Fig. 2.8b) to extract the films sheet resistance R_s. We evaporated gold contacts patterned using a shadow mask, then performed wire bonding on these contacts[2]. The samples were placed in a *Espec SH-42* climatic chamber to vary ambient temperature from 25 to 100 °C upon a heating and cooling cycle. Electrical measurements were performed every 2 °C, using a *Keithley 2400* source-meter an a remote-controlled program to automate electrical measurements and temperature stepping. Several trends are clearly visible (Fig. 2.8): the transition temperature, OFF/ON resistivity ratio, and OFF and ON state resistivities tend to increase with oxygen concentration, while the hysteresis width reduces. We attribute these observations to the higher concentration of *oxygen vacancies* in films prepared with lower oxygen concentration. Indeed, as introduced in Chapter 1, oxygen vacancies have been shown to modulate the thermally-triggered transition in a similar way [122, 130, 140, 197]. Oxygen vacancies are missing oxygen atoms in VO_2 crystal lattice, leaving a positively charged vacancy at their previous position (see inset in Fig. 2.8c). They are commonly found in oxygen-defficient $VO_{2-\delta}$

[2]Even to perform four-point measurements using probes, gold contacts are necessary. Indeed, unlike for silicon substrates, attempts to contact VO_2 thin films directly with probes does not yield reliable ohmic contacts.

films [198], and form preferentially at grain boundaries due to the lower formation energy of such sites [136]. They play the role of carriers as they can act as electron donors, thus lowering the resistivity of the film [199]. They serve as nucleation sites for the insulator-to-metal transition as they contribute to the destabilization of the monoclinic state, hence increasing the metallicity of the film and lowering the transition temperature [130].

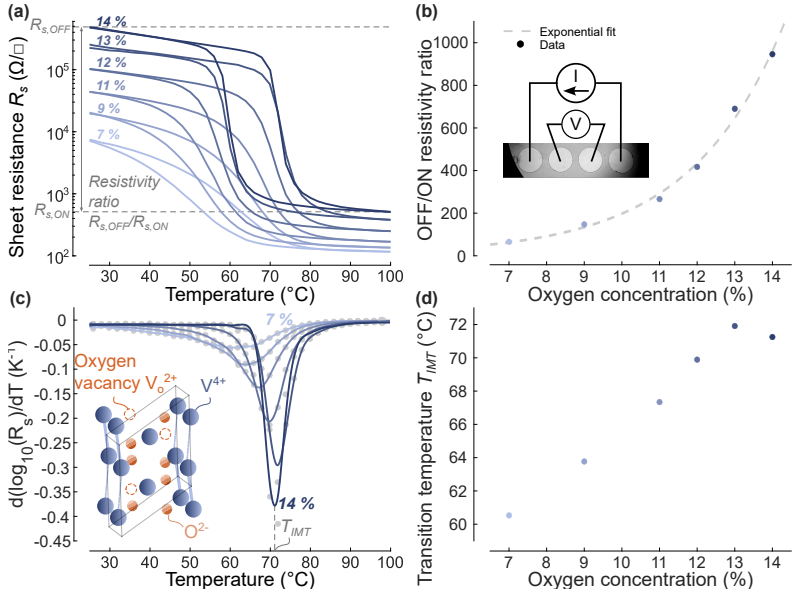


Fig. 2.8 Impact of the oxygen concentration during the sputtering step on the transition properties of the resulting VO_2 thin films. **(a)** Thermal hysteresis of their sheet resistance R_s. **(b)** Evolution of their OFF/ON resistivity ratio. Inset shows four-point setup for sheet resistance extraction. **(c)** Temperature derivative of their sheet resistance (decimal logarithm value), allowing to extract their insulator-to-metal transition temperature T_{IMT} **(d)**. Inset illustrates the definition of an oxygen vacancy.

2.1.4 Impact of annealing temperature

We have seen that oxygen concentration during the sputtering step can have large effects on transition properties, and a slight effect on grain size. However, a most common way to tune grain size is through annealing temperature. During thermal annealing, grain growth is driven by the reduction of the system energy, resulting in the reduction of grain boundary surface area. The higher the temperature, the higher the diffusion rate of species, the higher the migration rate of grain boundaries, yielding bigger grains [200]. Grain size has also been shown to increase with annealing duration, although this parameter was not investigated within this thesis [201].

To study the impact of annealing temperature on grain size, we sputtered five VO_2 thin films with identical process parameters (see Table 2.1) and 13 % oxygen concentration. We then annealed them at 400, 500, 600, 650 and 800 °C, respectively (other annealing parameters are those of Table 2.2). The SEM top views and cross sections of the resulting films are shown on Fig. 2.9 and 2.10. A one order of magnitude difference in crystallite diameter can be observed between 400 and 800 °C annealed films, proving that it can be an effective parameter to tune grain size in our specific process. Microstructure of the thin films also varies with annealing temperature, as shown in Fig. 2.10: at low annealing temperatures, small grains tend to stack up closely across the layer thickness, while starting from 650 °C, crystallites widths span across the full film thickness. The stoichiometry of all fabricated films was associated to VO_2 through Raman spectroscopy.

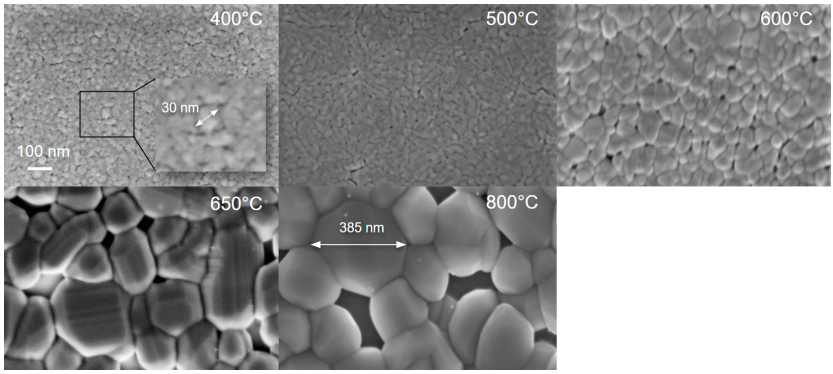


Fig. 2.9 SEM top view images showing the impact of annealing temperature on the grain size of the resulting polycrystalline VO_2 thin films. Scale bar is valid for all pictures.

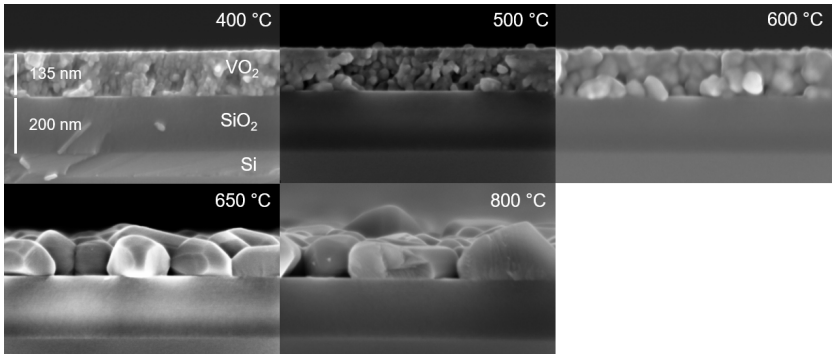


Fig. 2.10 SEM cross section images of the same samples. One can distinguish the polycrystalline VO_2 layer, on top of the SiO_2 200 nm-thick layer grown from the Si substrate. All VO_2 thin films have similar 135 nm thickness.

At larger temperatures, pinholes (void areas in between grains) start to appear, which is common when performing annealing of amorphous films [202]. This phenomenon fades out when further increasing film thickness, as shown in Fig. 2.11.

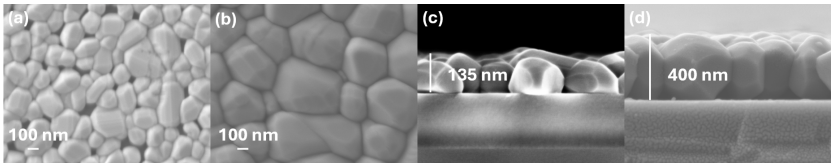


Fig. 2.11 Top view **(a),(b)** and cross-sections **(c),(d)** showing the impact of film thickness on the grains microstructure. Film shown in **(a),(c)** is 135 nm-thick, while film shown in **(b),(d)** has been deposited during 53 min 30 seconds (for about 3 times longer), yielding a 400 nm thickness. Other process parameters are equal (13 % $[O_2]$, 650 °C annealing).

The impact of the annealing temperature on the transition properties will be explored in the following section for two specific films.

2.2 Two-terminal device fabrication

Annealing temperature has thus proven to be an effective lever to tune grain size, in larger proportions compared to oxygen concentration. In the remaining of this work, we will thus fabricate two-terminal devices from two specific VO_2 thin films: the *small-grains film* annealed at 500 °C, and the *large-grains film* annealed at 650 °C, both produced at 13 % oxygen concentration[3]. All other process parameters related to the synthesis of the VO_2 thin films are identical, and have been specified previously in Table 2.1 and Table 2.2.

2.2.1 Process flow

The seven-steps process flow leading to the fabrication of VO_2 two-terminal devices is shown in Fig. 2.12. Cleaning of the silicon wafer is performed using a 3:1 mixture of H_2SO_4 and H_2O_2 in two consecutive baths heated at 100 °C for 10 minutes, each of which is followed by 10 minutes rinsing steps using DI water. The goal of these baths is to remove any organic impurities from the wafer surface. Wafers are then immersed in a diluted HF bath (2 %) for 15 seconds to remove the thin native oxide at the surface of the silicon wafer, followed by further DI water rinsing and drying. Wafers then undergo a 4-hours wet thermal oxidation at 1000 °C in a *Koyo* furnace to obtain a 200 nm-thick silicon oxide, which serves both as

[3]This concentration was selected as it maximized the OFF/ON resistivity ratio, which must be sufficiently high in order to observe a voltage-induced sharp resistive switching (see Fig. 2.17). For example, two-terminal devices fabricated with an 8 % $[O_2]$ and annealed at 500 °C did not present a sharp resistive switching behavior in their voltage-driven I-V curve.

amorphous layer for the growth of the VO_2 thin film, and as electrical insulation layer between the VO_2 thin film and the silicon substrate. Sputtering and annealing of the VO_2 thin film have been detailed previously in Section 2.1.2.

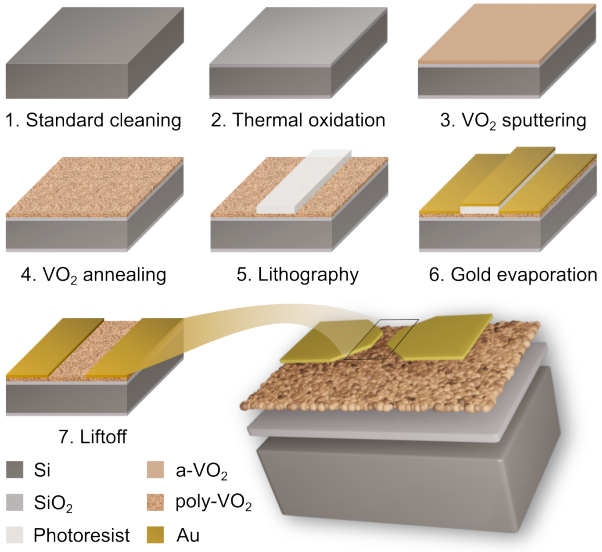


Fig. 2.12 Process flow leading to the fabrication of the VO_2 two-terminal devices

The two-terminal devices are defined by patterning electrodes on top of the unpatterned VO_2 thin film through liftoff. A lithography step is performed by spincoating a thin layer of *MicroChemicals AZnLof 5510* negative resist (dedicated to liftoff applications) at 3000 rpm for 60 seconds. A soft bake is perfomed at 90 °C for 60 seconds to dry the resist. The wafer is then exposed under a UV lamp at 120 mJ dose under a customely designed quartz photomask, followed by a post-exposure bake of 105 °C during 120 seconds. Development is performed manually during 33 seconds (this duration is a key parameters that has been optimized to achieve the target 600 nm resolution) using *MicroChemicals AZ 726 MIF* developer, before rinsing the wafer with DI water and drying it under nitrogen. A bilayer of nickel and gold[4] (respectively 5 and 150 nm thick) are then deposited through electron-beam thermal evaporation, based on previous literature work [203]. Liftoff of the resist is performed to reveal the contacts geometries, using two consecutive 5 minute-long baths of dimethyl sulfoxide (DMSO) heated at 60 °C under ultrasound vibrations (37 kHz, 100 % power, pulse mode), followed by acetone, methanol and DI water rinsing. Finally, the wafer are dried under a nitrogen flux.

[4]Gold is commonly used as material for the electrical contacts when investigating the electrical properties of emerging materials, thanks to its low resistivity and stainless properties.

The two types of two-terminal VO_2 devices studied in this work, namely the *large-grains devices* and the *small-grains devices*, are shown in Fig. 2.13. The channel dimensions are defined solely by the contact geometries, since the VO_2 layer is not patterned. Such choice can increase the power consumption of the resulting device, as current lines will exceed the channel area, inducing losses. However, it simplifies the fabrication process, removing the need for a VO_2 etching technique selective to the SiO_2 substrate, and leaving the VO_2 layer in a pristine state. Dimensions of the fabricated devices range from L = 0.6 to 6 µm, and from W = 3 to 100 µm. Such low resolution has been achieved through optimization of the lithography parameters, with the development duration playing a key role.

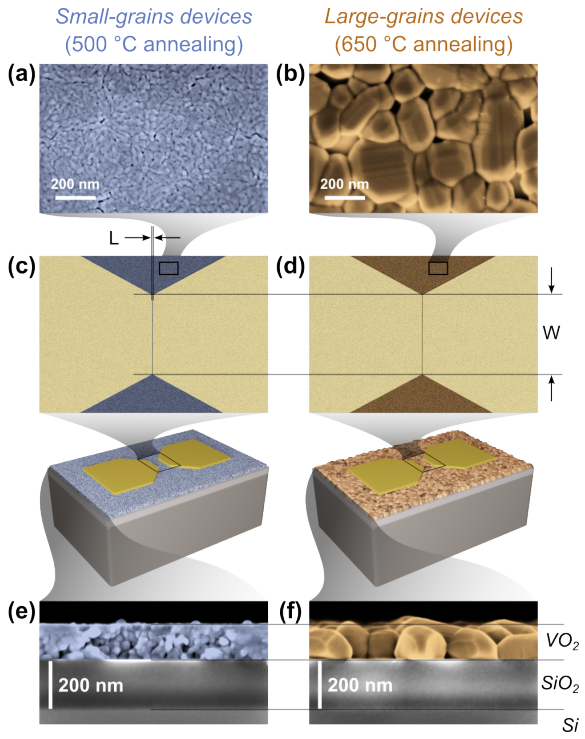


Fig. 2.13 Two-terminal devices have been fabricated from the small-grains and a large-grains films **(a)(b)** SEM top views show a ten-fold difference in mean grain size (from 32 nm mean size with 8 nm standard deviation, to 210 nm mean size with 77 nm standard deviation, deduced through image processing of the SEM pictures). **(c)(d)** Convention for device dimensions: length *L* and width *W*. **(e)(f)** Thin films cross section.

2.2.2 Material characterization of small- versus large-grains films

Similarly to what we did in Section 2.1.3 for oxygen concentration, we studied how the difference in grain size tuned through annealing temperature impacted the films transition properties. Results are shown in Fig. 2.14. Differences between the two films are similar to those due to change in oxygen concentration, despite the fact that both films have been sputtered at the same 13 % $[O_2]$. We witness a larger OFF and ON state resistivity, higher transition temperature, smaller hysteresis width and larger OFF/ON resistivity ratio in the case of the large-grains film.

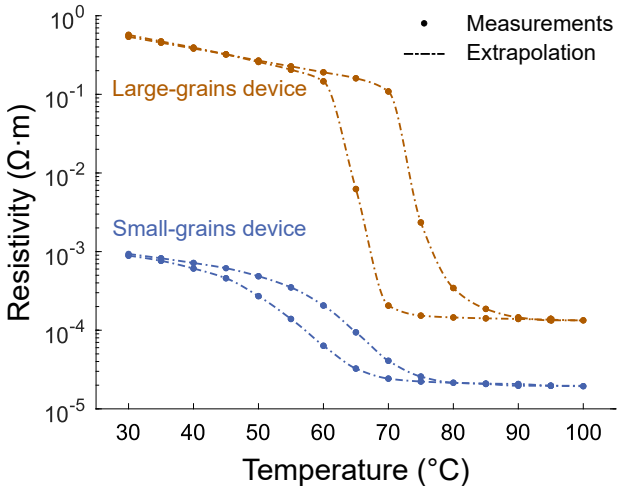


Fig. 2.14 Resistivity versus temperature hystereses measured from two two-terminal devices of identical dimensions, fabricated respectively from the small-grains film (purple) and the large-grains film (orange). Measurements have been performed on a temperature-controlled hot chuck, in the linear resistance regime to prevent self-heating effects (corresponding to the 0 – 10 mV voltage range).

Several works have observed such effects in correlation with grain size reduction [125, 204, 205]. Once again, we believe we can attribute these results to the higher concentration of oxygen vacancies in the small-grains film, in correlation to its higher grain boundaries density. Indeed, as mentioned above, oxygen vacancies tend to concentrate at grain boundaries [136]. By increasing grain size, the grain boundaries area is reduced, reducing the preferential nucleation sites of these defects. Additionally, remaining oxygen species in the annealing ambient could diffuse in higher concentration inside the VO_2 thin film when annealed at higher temperature, increasing oxygen content and filling oxygen vacancies. One can note that the small-grains film presents a much smaller OFF/ON resistivity ratio than the film prepared with similar process parameters when investigating the impact of oxygen concentration (see Fig. 2.8a), with a one order of magnitude difference.

This can be explained from the fact that several months separate the experimental campaigns related to oxygen concentration and annealing temperature, meaning the V target has aged significantly through numerous depositions, altering the transition properties of the resulting thin films. However, in each individual campaign, films have been deposited consecutively and can thus be compared.

To ensure that the difference between the two films were not due to the presence of a distinct vanadium oxide phase, their stoichiometry was assessed through both Raman spectroscopy (Fig. 2.15) and XRD (Fig. 2.16). The Raman peaks of both films were identical in terms of Raman shift, and could all be attributed to the VO_2 stoichiometry (beside the 520 cm^{-1} silicon substrate peak).

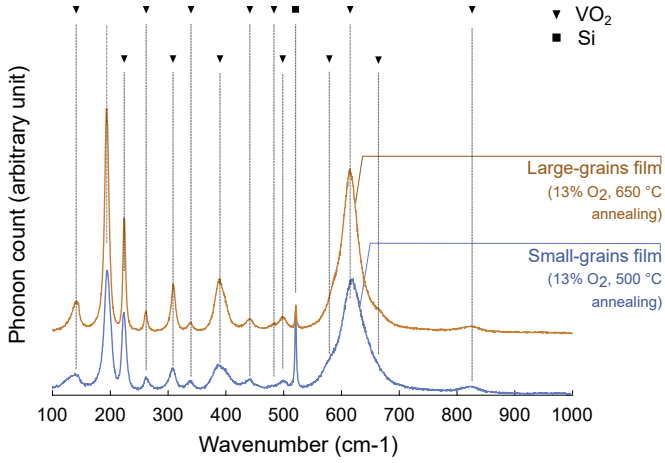


Fig. 2.15 Raman spectra of the small- and large-grains film, performed with similar parameters as those of Fig. 2.3. All peaks could be attributed to VO_2 or the Si stoichiometry.

However, Raman is not a quantitative technique, as phases present with smaller concentration might have peaks of low amplitude which would be lost in the background. XRD spectra reveal additional information. A peak can clearly be attributed to VO_2 (011) (Fig. 2.16b), hinting that both films have a preferential (011) crystalline orientation. The peak of the small-grains film being shifted right and having a larger FWHM possibly relates to its expected sub-stoichiometry ($VO_{2-\delta}$). For the two other peaks shown in 2.16c, three candidates have been identified: V_3O_7, V_2O_3 and V_7O_{13}. References used for peaks attribution are 65-2870 (Au), 44-0252 (VO_2), 71-0454 (V_3O_7), 65-9474 (V_2O_3), and 18-1449 (V_7O_{13}). No peaks corresponding to V_3O_7 or V_2O_3 were visible in the films Raman spectra; however, no Raman peaks have been reported so far for V_7O_{13}, so it could be a candidate. Whatever the nature of this peak, it is present in both films, and should thus not be the cause of the difference in transition properties.

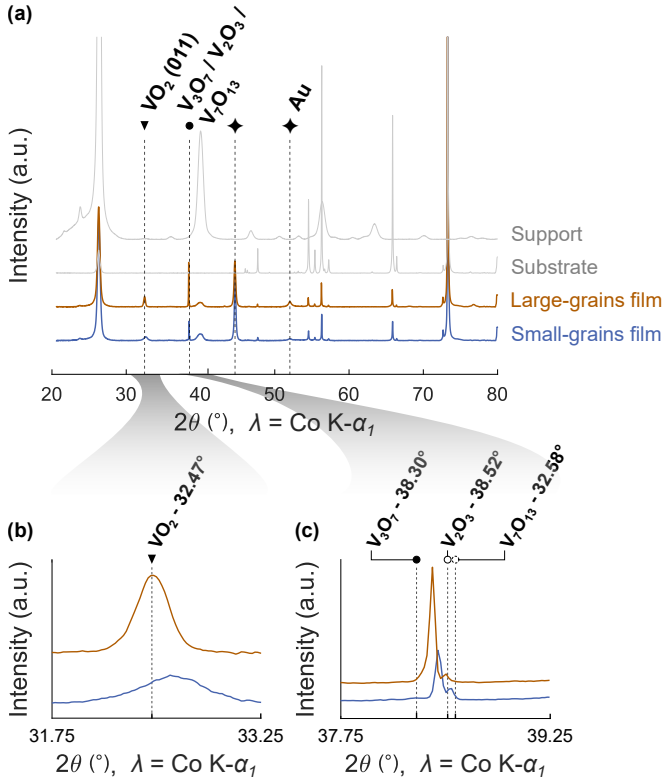


Fig. 2.16 **(a)** XRD spectra of the small-grains (purple) and large-grains (orange) film, along with the substrate and sample support references. Spectra were acquired using a Cobalt radiation source (K-α1 = 1.78897), with a 30 kV accelerating voltage. The characterization was performed at the end of the fabrication process, and peaks attributed to the gold contacts on the sample surface are indicated by a star. Figures **(b)** and **(c)** show close-up views on the peaks attributed to VO_2 and to another vanadium oxide stoichiometry, respectively.

2.2.3 Two-terminal devices electrical characteristics

The typical voltage-driven I-V characteristic of a VO_2 two-terminal device (small-grains, in this case) is presented in Fig. 2.17b. As introduced in Chapter 1, the insulator-to-metal transition can be triggered electrically through a combination of Joule heating and charge injection effects, resulting in a abrupt drop in the device resistance. A current compliance is set on the voltage source to protect the device in its metallic regime. We define 5 important electrical parameters: its insulator-to-metal and metal-to-insulator threshold voltages V_{IMT} and V_{MIT}, its voltage-dependent resistance in the insulating state ($R_{ins,0}$ close to zero bias and R_{IMT} close to the transition), and its resistance in the metallic state R_{met}. Resistances have been extracted through linear regression around a bias point. As mentioned previously, the transition takes place through the formation of a metallic filament, which we

were able to image through high-resolution microscopy (see Fig. 2.17d). The formation of this filament is an intrinsically stochastic process, resulting in cycle-to-cycle variations in the electrical characteristics of VO_2 devices, as illustrated in Fig. 2.17b.

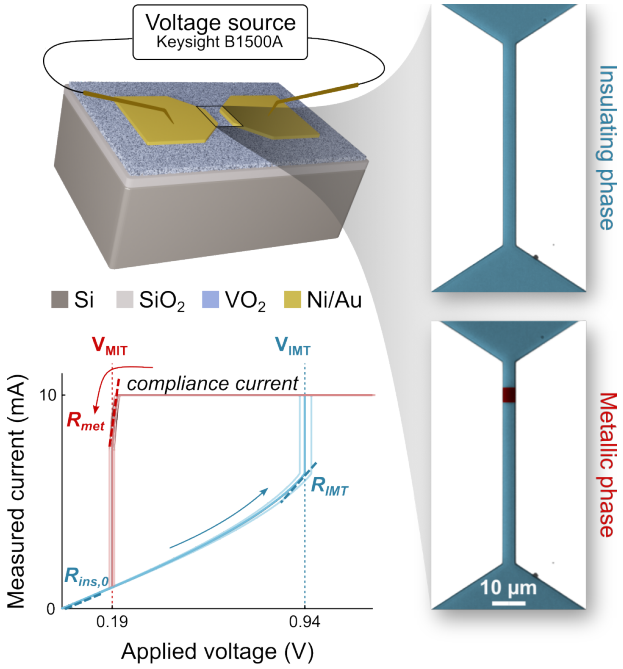


Fig. 2.17 **(a)** Structure of the VO_2 two-terminal device, with the measurement setup used to extract its I-V characteristic. **(b)** Voltage-driven I-V characteristic of a VO_2 two-terminal device, with its electrical parameters in the insulating (blue) and metallic (red) regimes. **(c)(d)** High-resolution microscopy image of a VO_2 two-terminal device in both regimes. In its low resistance state **(d)**, the metallic filament is visible through contrast thanks to the change of optical constants of the VO_2 film when transitioning. Blue (insulating) and red (metallic) colors have been added to enhance the phase distinction.

2.3 Tuning the stochasticity of resistive switching

Using the two types of two-terminal VO_2 devices fabricated with different crystallite size, we can investigate whether the number of grains in the channel can be used to tune cycle-to-cycle stochasticity. But why would it be desirable to control this inherent stochastic behavior in the first place? In certain applications of VO_2 micro-resistors, these cycle-to-cycle variations are sought for. They have been harnessed to design stochastic neuron, which implement behavior such as

phase-locked firing (also called skipping) [85], or stochastic resonance [206]. These variations have also been exploited to realize stochastic functions including probabilistic bits [207, 208], stochastic oscillators [209, 210] and random access memories [211]. Generally speaking, inherently stochastic neurons would be desirable for the implementation of Stochastic Spiking Neural Networks, which have shown benefits to escape local minima, optimally explore space of solutions [212], and realize probabilistic inference [213, 214]. On the other hand, in the case of the sensory VO_2 neurons presented in Chapter 1 where stimuli are encoded in the spike rate [91–93, 95, 96, 169, 181], these variations should be minimized as they induce jitter on the oscillator frequency. This can impede the sensor resolution, as further explored in Chapter 3.

2.3.1 On the origin of stochasticity

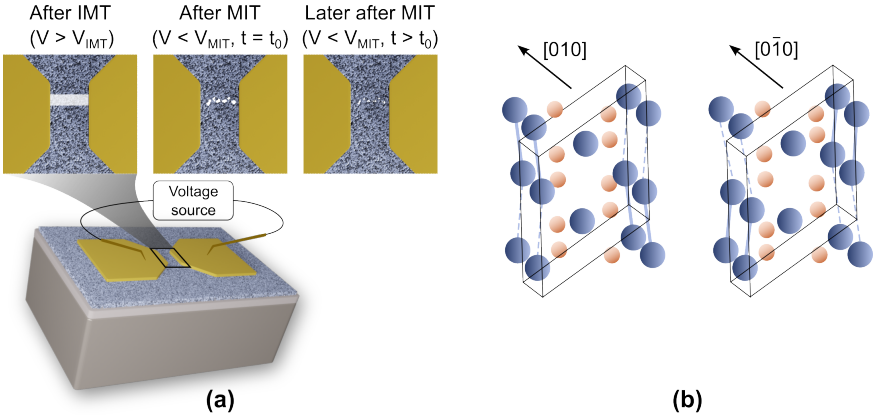


Fig. 2.18 Illustration of the two hypotheses behind cycle-to-cycle variations: **(a)** Persistent metallic islands, located randomly at the location of the metallic filament formed during the previous cycle, serve as nucleation sites for the formation of the next filament. Inspired from [208]. **(b)** Rutile VO_2 crystallites can transition into one of four symmetries of the VO_2 M1 monoclinic phase during the MIT. Two of them are illustrated here. The solid purple lines show the V-V dimers, which are located at different places of the unit cell in the two configurations. Inspired from [215].

Although the mechanisms behind these variations are not fully understood, two phenomena stand out as possible causes, which both take place during the relaxation from metal to insulator (MIT). A first possible mechanism is the presence of long-lived metallic domains [216–219] which can persist up to several milliseconds after the metallic filament resorption (once the applied voltage is removed), in the region of the device where the filament previously formed. Indeed, these metallic islands serve as preferential nucleation sites for the next transition [219], and

their expected random nature [208] could result in slightly different insulating resistances and threshold voltages from cycle to cycle. This hypothesis is illustrated in Fig. 2.18a. Presence of these metallic domains have been observed through nano-XRD [216], optical reflectivity [217] and dark-field X-ray microscopy measurements [219]. A second possible mechanism is the fact that individual grains might change crystalline orientation from cycle to cycle, impacting the device insulating resistance and threshold voltage. These changes have been imaged through transmission electron microscopy (TEM) [215], and four possible transition pathways have been computed for a crystallite undergoing MIT, corresponding to four 90 °-symmetrical crystalline orientations of the insulating phase. This hypothesis is illustrated in Fig. 2.18b. Stochasticity could also be due to a combination of the two phenomena. In both observations, variations of properties in individual grains are involved: one thus expects to witness a larger impact on the micro-resistor electrical properties variations as device dimensions get closer to crystallite size. However, such claim has never been verified, nor through simulation or measurements; and the magnitude of the impact on the stochasticity has never been quantified: this will be the focus of this work. To test whether stochasticity is dependent on the grain size/device dimensions ratio, we will successively study the impact of grain size and VO_2 micro-resistor length.

2.3.2 The impact of grain size

To understand the impact of grain size on stochasticity, we studied the I-V characteristics of two micro-resistors of identical channel dimensions (L = 600 nm, W = 70 µm) fabricated from the small- and large-grains films, respectively labelled device A0 and B0. To obtain a good estimate of the electrical parameters distributions, we measured 300 voltage-driven cycles for each device. Measurements were performed in a 3-probes configuration[5] using a *Keysight* B1500A Semiconductor Device Analyzer to extract the exact voltage drop across the micro-resistor. Compliance current was set to protect the device from over-heating during the metallic filament formation, and avoid damages to the crystalline structure [120, 220, 221]. Measurements were performed on a temperature-controlled chuck at 35 °C, to ensure that the device did not follow room-temperature fluctuations during measurements. We filled the probe station chamber with a 6 l/min nitrogen flux during measurements to prevent the VO_2 film from oxidizing. We start by investigating the distribution of the insulator-to-metal threshold voltage V_{IMT}, as it is a key parameter in most of the above-cited applications. Results are detailed in Fig. 2.19.

[5] We used 3 source measure units (SMUs): SMU1 contacts one of the device terminals, is set as a voltage source and fixed at zero volts (ground); SMU2 contacts the other terminal, is set as a voltage source and performs the voltage sweep; SMU3 contacts the same terminal, but is set in current mode and fixed at zero amperes to measure the actual voltage drop.

A first difference between the two devices is a much lower insulating resistance of the small-grains device (A0) compared to the large-grains one (B0), and its IMT taking place at a lower threshold voltage V_{IMT} and larger current levels. In consequence, we set different current compliances for the two devices. These differences are related to the large difference in resistivity and transition temperature between the two films, as discussed previously (see Fig. 2.14).

The second difference is that the large-grains device shows larger cycle-to-cycle V_{IMT} variability, confirming our hypothesis that larger stochastic behavior would be observed with crystallite size getting closer to device channel length. Threshold voltages have been extracted over 300 cycles, and their distribution fitted (Fig. 2.20g-h) after removal of a small drift over time ($\sim$0.001 %/cycle, obtained from linear regression). The difference in stochasticity is observed through two features:

- First, the type of distribution, which is purely Gaussian in the case of the small-grains device, while the large-grains device shows a mixture of two Gaussian components.
- Second, the widths of these Gaussian functions, which we study through their standard deviations $\sigma_{V_{IMT}}$ (in millivolts) and coefficients of variation $CV_{V_{IMT}}$ (standard deviation normalized by the mean, in percentage of the mean). One notes an 18-fold ratio in $\sigma_{V_{IMT}}$ and 8-fold ratio in $CV_{V_{IMT}}$ when comparing the small-grains and large-grains device (taking the Gaussian component with the smallest width for the latter).

The stochastic behavior scaling is thus confirmed both from an absolute and relative perspective.

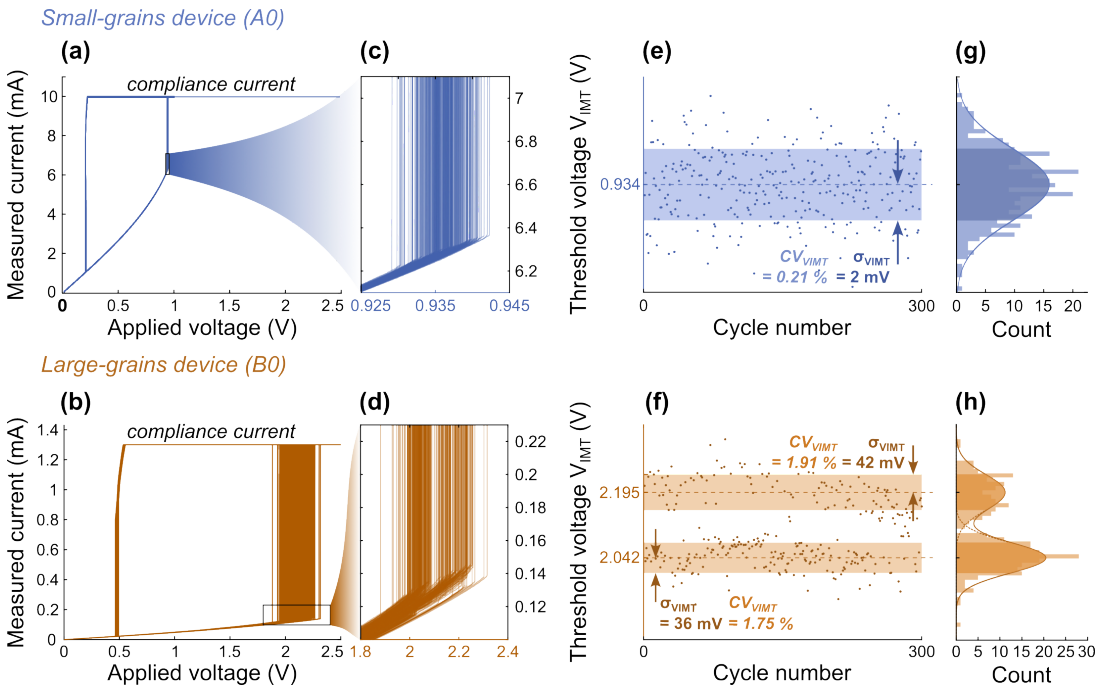


Fig. 2.19 **(a)(b)** Voltage-driven I-V characteristics of the small-grains device (A0) and the large-grains device (B0) of identical dimensions (L = 600 nm, W = 70 µm), measured over 300 cycles. **(c)(d)** Close-up views on the cycle-to-cycle variation in insulator-to-metal threshold voltage V_{IMT}. The larger variations in the case of device B0 are visible when comparing the colored x-axis of the two figures. **(e)(f)** Extraction of V_{IMT} for each cycle, after drift removal. **(g)(h)** Fit to Gaussian and Gaussian mixture distributions, respectively for the small-grains and large-grains devices. Standard deviations $\sigma_{V_{IMT}}$ and means $\mu_{V_{IMT}}$ are indicated, as well as coefficients of variation $CV_{V_{IMT}}$.

2.3.3 The impact of channel length

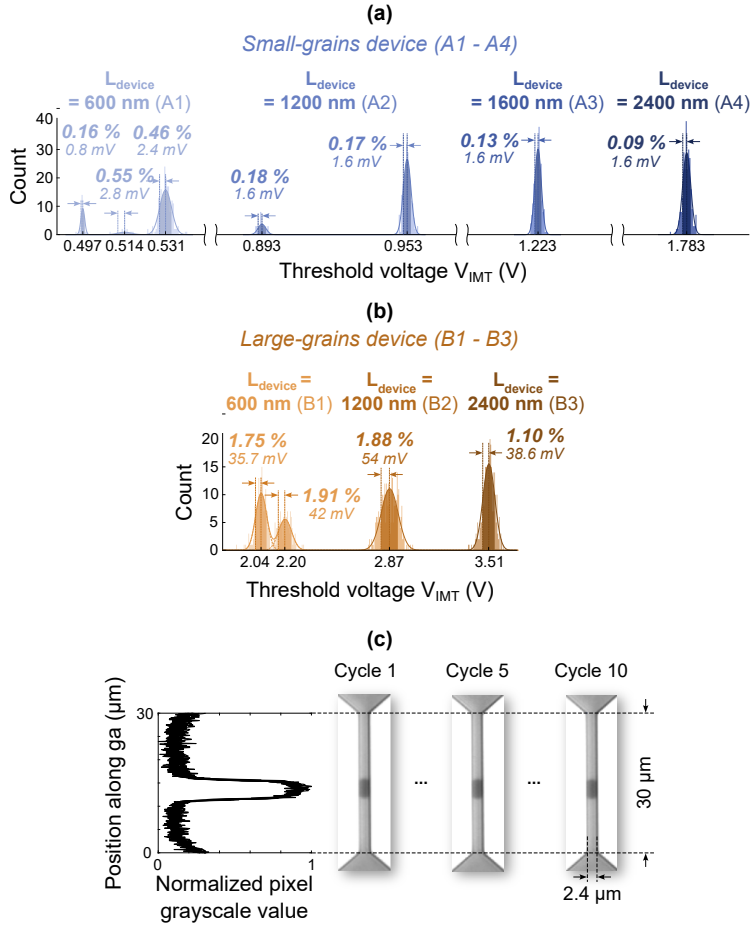


Fig. 2.20 Impact of the VO_2 device channel length on their threshold voltage distribution. **(a)** Small-grains devices, for channel lengths of 600 (A1), 1200 (A2), 1600 (A3) and 2400 nm (A4). Channel width is 6 µm in all cases. **(b)** Large-grains devices, for channel lengths of 600 (B1), 1200 (B2) and 2400 nm (B3). Channel width is 70 µm in all cases. The mean $\mu_{V_{IMT}}$ (in V), standard deviation $\sigma_{V_{IMT}}$ (in mV), and coefficient of variation $CV_{V_{IMT}}$ (in %) are displayed for each individual Gaussian distribution. **(c)** High-resolution microscopy imaging of the metallic filament across 10 cycles in a small-grains device with 2.4 µm long channel. The plot shows normalized pixel grayscale values extracted along the channel width over the 10 cycles.

To go further, we also investigated how the distribution of the threshold voltage would evolve with the micro-resistor channel length, as shown in Fig. 2.20. We extracted V_{IMT} values from 300 measured cycles for devices of varying channel

length, fabricated both from the small-grains film (devices labelled A1-A4, see Fig. 2.20a) and the large-grains film (devices labelled B1-B3, see Fig. 2.20b). It is noteworthy that for small-grains devices, channel width changed from 70 to 6 µm compared to the device studied in Fig. 2.19a. The V_{IMT} distributions of these small-grains devices with smaller channel width and length also appear to be Gaussian mixtures. The stochastic behavior magnitude tends to reduce with increased channel length, providing a second lever to tune intrinsic stochasticity in VO_2 micro-resistors. Once again, this stochasticity reduction manifests itself through two distinct aspects: (1) the Gaussian mixtures observed for devices of small channel length tend to fuse into single-Gaussian distributions, and (2) the coefficients of variation of the individual Gaussian distributions are reduced, as channel length increases. In all cases, the $CV_{V_{IMT}}$ orders of magnitude are similar to the previous case (see Fig. 2.19e-f) and maintain a $\sim$10 ratio between the large- and small-grains devices.

2.3.4 Discussion

We separately investigate the physical causes for the two aspects of our observations: reduction in coefficient of variation $CV_{V_{IMT}}$, and shift from single-Gaussian to bi-modal distributions.

Reduction in coefficient of variation

The coefficient of variation of V_{IMT} tends to decrease when reducing grain size and increasing channel length. The percolation process leading to the creation of a metallic filament during voltage-triggered IMT has been shown to nucleate from "hotspots" [164, 165, 219] where Joule heating/electrical effects are of higher intensity, and thus transition first. Once this group of crystallites has transitioned, the voltage drop re-distributes across their neighbours which transition in their turn, triggering the metallic filament. We hypothesize that V_{IMT} values belonging to the same Gaussian component correspond to closely located hotspots, yielding filaments in similar positions. Indeed, recent work showed that localized defects induced with an ion beam reduced the IMT cycle-to-cycle stochasticity by defining the filament percolation hotspot [222]. To further support our claim, we imaged filaments in individual devices across several cycles using a *Microsanj* optical microscope, acquiring images at a 530 nm wavelength. Results show that for devices with larger channel width and single-Gaussian V_{IMT} distributions, the filament always forms at the same location. Results for one of these devices is shown in Fig. 2.20c, and the analysis has been repeated for other devices with single-Gaussian V_{IMT} distributions (Fig. 2.21), showing that the filament width W_f is relatively constant with channel width W. This can be explained by the fact that the filament width is limited by the current compliance, which is kept constant at 10 mA for all measured small-grains devices.

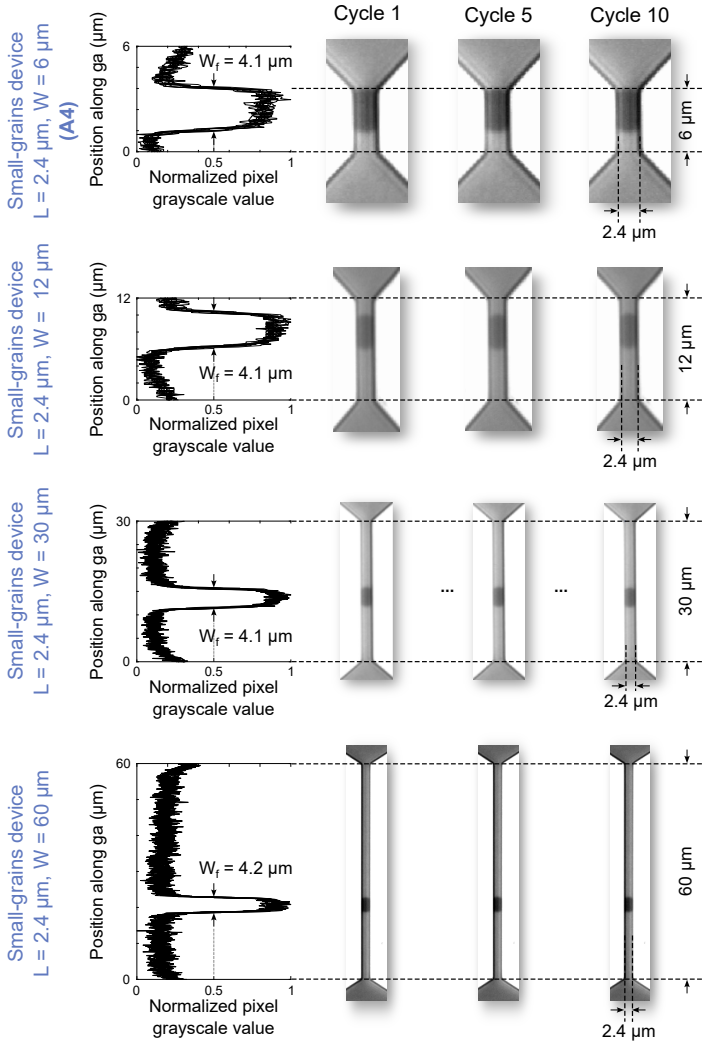


Fig. 2.21 Optical imaging of the filament over several voltage-driven cycles in 4 different small-grains devices of identical channel length (L = 2.4 µm) but different channel width (W = 6, 12, 30 and 60 µm). Left plots show normalized pixel grayscale values extracted along the channel width over the 10 cycles. The first device corresponds to device A4. One can see that filament width W_f, extracted at half the normalized intensity, is relatively constant with channel width W, given the equipment resolution. In a given device, the position of the filament is also constant from cycle-to-cycle. The same 10 mA current compliance is applied for all devices.

Although the filament always nucleates from closely localized hotspots, slight alterations of the grains resistance from cycle to cycle could affect the initial percolation path, yielding variations in V_{IMT}. As stated earlier, experimental observations that could explain these variations include the presence of persistent metallic domains [217], or changes in grains crystalline orientation [215]. Both theories involve variation in grains located in the area of the metallic filament formed in the previous cycle. In particular, the presence of residual metallic domains has been shown to facilitate filament re-triggering at the same location from cycle to cycle [219]. In both cases, as these variations involve individual grains, one expects them to have less impact on the IMT bias point if the percolation path involves a larger number of grains, or if these grains are more uniform. In relation to that, we do show that the small-grains film has a smaller grain size standard deviation, as specified in Fig. 2.13. Regarding uniformity, low cycle-to-cycle variability was recently reported [103] in VO_2 micro-resistors threshold voltages ($\sigma_{V_{IMT}}$ and $CV_{V_{IMT}}$ of 9.4 mV and 0.71 %), and attributed to the epitaxial nature of the VO_2 film yielding highly uniform crystallites. However, epitaxial growth requires specific substrates such as Al_2O_3 which are not straightforwardly CMOS compatible. In the case of our small-grains device A0 whose film was grown on Si/SiO_2 substrates, we observe both smaller $\sigma_{V_{IMT}}$ and $CV_{V_{IMT}}$ (2 mV and 0.21 %), which shows that epitaxial growth might not be necessary to minimize cycle-to-cycle variations.

The grain size variation comes with differences in the films material and electrical properties, which could also explain the stochasticity difference between the resulting devices. First, the fact that the large-grains film shows a sharper transition and a higher OFF/ON resistivity ratio (see Fig. 2.14). In this case, filament formation could be triggered from a smaller number of seed grains, as the "percolation effect" (applied voltage redistribution across the seed grains neighbours, precipitating the filament creation [165]) could be stronger. This would imply stronger stochasticity in large-grains devices, consistently with our hypothesis. Secondly, the possible higher proportion of rutile VO_2 within the small-grains film insulating state [130], and possible presence of another vanadium oxide phase (hinted by XRD measurements shown in Fig. 2.16), could impact filament formation. Thirdly, residual stresses and strain (known to have large impacts on the IMT [223] [224]) could be present with larger inhomogeneity in the large-grains film. Different degrees of crystalline orientation uniformity between the two films could also explain the observed differences, but should be confirmed using a more sensitive characterization technique such as grazing incidence XRD. In addition to these differences, dynamical effects related to the decay of persistent metallic islands could play a role. In our study, the high-resolution measurement of a single cycle lasts for about 10^2 s, while reported timescales for long-lasting metallic islands range from 10^{-2} s [217] to 10^2 s [219]. They are also more concentrated at temperatures closer to

T_{IMT} [217]. It is thus possible that small-grains devices - which have lower T_{IMT} - contain more long-lasting metallic domains, impacting stochasticity. However, the strongest argument hinting that stochasticity could be correlated to the number of grains in the channel, is the fact that stochasticity depends on device length. This effect is smaller for large grains: in this case, there is a smaller number of "possible seed grains location" along the channel length axis, which could explain this difference. Finally, the presence of an electro-forming behavior in the large-grains devices is discussed in Appendix A.

Bi-modal behavior

Let's now investigate the possible causes for the multi-modal Gaussian V_{IMT} distributions. These could correspond to distinct percolation hotspots and resulting filament locations. To our knowledge, multi-modal V_{IMT} distributions have been solely observed once in VO_2 oscillators [93], but no explanatory mechanism has been proposed so far in the literature. Our optical microscopy imaging equipment resolution did not allow us to image the filament in the devices showing multi-modal V_{IMT} distributions, due to their small channel length. However, a previous work [225] studied filament localization after voltage-triggered IMT upon thermal cycling. They found that while the filament always forms at the same spot in devices of larger channel length, its location is varies from cycle-to-cycle in devices of smaller channel length. They attributed this to the higher electric field arising across channels of smaller-length devices, triggering a purely electrical, field-induced transition. Such transition takes place when carrier density locally reaches a critical threshold, as recently observed in VO_2 nanowires [226]. Since field-induced carrier generation is an inherently stochastic process, it could happen in any individual grain of the channel, and thus trigger filament formation at different locations. Indeed, the probability of a given grain to undergo field-induced IMT increases exponentially with the local voltage drop across this grain [121]. As devices distributions show a finite number of Gaussian components, we could argue that there could be preferential sites for field-induced nucleation.

So far, most works who mention stochasticity in VO_2 focused on cycle-to-cycle variations in the threshold voltage V_{IMT} [93, 96, 207, 217, 227]. In this work, we also studied the variations in the insulating state resistance. The importance of taking R_{ins} variations into account will be highlighted in Chapter 3 for VO_2 temperature spiking sensors, showing that they contribute to spike rate stochasticity with the same proportion as V_{IMT} variations. To perform this study, we distinguish the device resistance close to zero bias $R_{ins,0}$ from the resistance R_{IMT} just before the IMT (Fig. 2.17b). Indeed, the OFF-state device resistance decreases with Joule heating, as VO_2 resistivity is temperature dependent in its insulating phase. Fig. 2.22 compares the distributions of these two resistances for the small-grains device A0 and

large-grains device B0 of identical dimensions. Full comparison of σ and CV values for the distributions of $R_{ins,0}$, R_{IMT} and V_{IMT} is available in Table 2.3. Once again, the large-grains device shows larger cycle-to-cycle variations in the two studied resistances, both from an absolute ($\sigma_{Rins,0}$ and σ_{IMT}) and relative ($CV_{Rins,0}$ and CV_{RIMT}) perspective. Interestingly, the coefficient of variation of the insulating resistance is much larger close to the IMT point than around zero bias. This could hint that the previously mentioned small variations in the resistances of individual grains (persistent metallic islands, change in crystalline orientation) are amplified when approaching transition. Indeed, prior to IMT, the resistance of a given grain in the channel follows a decreasing exponential temperature-dependent law [121].

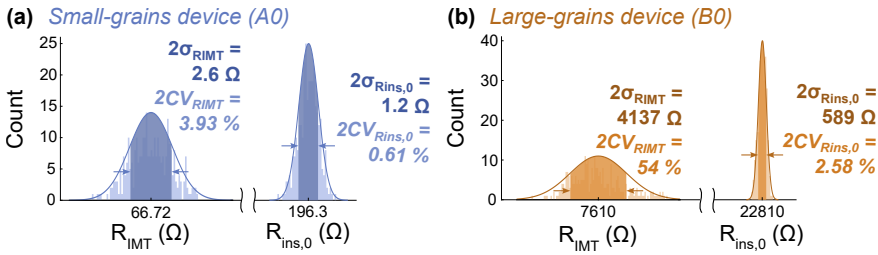


Fig. 2.22 Distributions of the devices insulating resistances close to zero bias ($R_{ins,0}$, on the right of each graph) and just before the insulator-to-metal transition (R_{IMT}, on the left of each graph), for small-grains device A0 **(a)** and large-grains device B0 (b) of identical dimensions (L = 600 nm, W = 70 µm). The resistances standard deviations $\sigma_{R_{IMT}}$, $\sigma_{R_{ins,0}}$ and coefficients of variations $CV_{R_{IMT}}$, $CV_{R_{ins,0}}$ are indicated.

Table 2.3 Comparison between the parameters of the Gaussian distribution of the three key electrical parameters of the small-grains device A0 and large-grains device B0

	Mean μ_X		**Standard deviation** σ_X		**Coefficient of variation** CV_X	
	Small- grains device A0	Large-grains device B0	Small-grains device A0	Large-grains device B0	Small-grains device A0	Large-grains device B0
$R_{ins,0}$	196.3 Ω	22812 Ω	0.60 Ω	294 Ω	0.31 %	1.29 %
R_{IMT}	66.7 Ω	7613 Ω	1.31 Ω	2069 Ω	1.97 %	27.17 %
V_{IMT}	0.934 V	2.042 V 2.195 V	2 mV	36 mV 42 mV	0.21 %	1.75 % 1.91 %

More interestingly, studying the relationship between $R_{ins,0}$, R_{IMT} and V_{IMT} gives us insight on the causes behind bi-modal V_{IMT} distributions. We perform this study for device A2, one of the small-grains devices with bi-modal behavior, using scatter plots between its $R_{ins,0}$ and V_{IMT} values (Fig. 2.23a) as well as its R_{IMT} and V_{IMT} values (Fig. 2.23b). The goal is to see whether bi-modal distributions are also observed in these two parameters: while it isn't the case for the resistance close to zero bias $R_{ins,0}$, for R_{IMT}, two clusters corresponding to the V_{IMT} modes are clearly visible (Fig. 2.23a-b). The resistance just before IMT is dominated by the resistance of the zone near the hotspot, as it will be less resistive and

concentrate current flow. These clusters hint that the two threshold voltages could correspond to distinct hotspots in the channel, while individual variations inside one cluster could be related to changes in grains of the same hotspot, as explained earlier. Such observation gives further evidence that bi-modal V_{IMT} distributions could correspond to distinct filamentary paths, as illustrated in Fig. 2.23c.

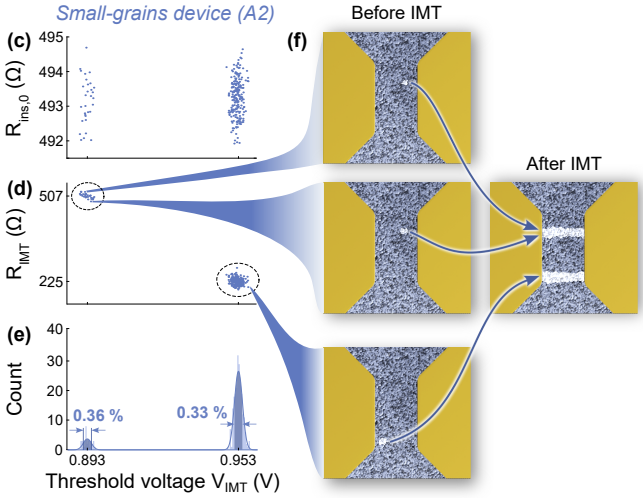


Fig. 2.23 Investigation of the relationship between $R_{ins,0}$, R_{IMT} and V_{IMT} for device A2, one of the devices showing multi-modal threshold voltage distribution. **(a),(b)** Scatter plots of $R_{ins,0}$ and R_{IMT}, respectively, against V_{IMT}. **(c)** Bi-modal distribution of the device V_{IMT} **(d)** Illustration of our hypothesis: cycles whose R_{IMT} values belong to the same cluster would yield closely located filaments, while those belonging to distinct clusters would yield distinct filaments.

2.4 Key messages

Several main conclusions can be taken away from this chapter.

There is currently a lack of design guidelines to understand how to fabricate VO_2 resistive-switching devices with minimized or maximized cycle-to-cycle stochasticity. Depending on the specific application in the neuromorphic research field, this stochasticity can be either detrimental (e.g. in sensitive spiking neurons implementing rate coding of input stimuli), or beneficial (in computational neurons used to build stochastic spiking neural networks). This diversity of targeted stochastic properties underlines the need for the mentioned design guidelines.

Cycle-to-cycle stochasticity in VO_2 two-terminal devices can be minimized by increasing the number of grains along the device channel length. This can be realized in two ways:

- By *reducing grain size* at constant channel dimensions. Reducing grain size by one order of magnitude allows to reduce relative cycle-to-cycle variations in the device threshold voltage (quantified by its coefficient of variation) by one order of magnitude as well.
- By *increasing channel length* at constant grain size. Increasing channel length by a factor of 4 yields a $\sim$6 fold and $\sim$2 fold reduction in threshold voltage coefficient of variation for small- and large-grains devices, respectively.

This scaling in stochasticity manifests itself through two main characteristics:

- Reduction in *relative* stochasticity (coefficient of variation, in percent), and even in *absolute* stochasticity (standard deviation, in Volts) in the case of grain size. These reductions are valid for three of the key electrical parameters of resistive swicthing devices: insulator-to-metal threshold voltage V_{IMT}, and insulating resistance both close to zero bias ($R_{ins,0}$) and to the transition point (R_{IMT}).
- Emergence of *bi-modal distributions* in the device electrical parameters (insulator-to-metal threshold voltage V_{IMT} and insulating resistance close to transition R_{IMT}), for devices with a smaller number of grains within the channel.

An hypothesis explaining the presence of bi-modal distributions of the device electrical parameters could be distinct filament locations. This was deduced by observing correlation between the modes of V_{IMT} and R_{IMT}. Indeed, the value of R_{IMT} is dominated by hotspots which concentrate current flow. The two correlated modes observated in both parameters thus hint for the presence of distinct

hotspots yielding to predominant filament formation paths. We consolidate this hypothesis by showing that filament always form in similar locations in devices showing single-Gaussian statistics, through high-resolution optical imaging.

Annealing temperature is an efficient lever to tune the grain size of VO_2 thin film. To a smaller extent, oxygen concentration during the sputtering step also has an impact on grain size. Both parameters have significant impact on the films transition properties, which could also explain the difference in stochasticity between the small and large grains samples. However, the scaling of the stochasticity with device channel length brings generality to our conclusions.

The impact of the devices width on their cycle-to-cycle stochasticity must still be investigated. Our work did not provide enough measurements to assess such correlation. We however believe the impact of the devices width would be smaller than that of their length: indeed, our high-resolution microscopy images showed that the filament maintained a similar width W_f, regardless of the device width W. This is due to the fact that W_f is set by the compliance current. As a consequence, the number of grains involved in a filament does not depend on the device width.

Through our work, we provide key guidelines to design planar VO_2 resistive-switching devices with minimized or maximized stochasticity. Neuromorphic designers aiming to design sensitive spiking sensors with minimum cycle-to-cycle variations should choose films with smaller grain size and larger channel length, while designer seeking to exploit these cycle-to-cycle variations should select devices with larger grain size and smaller channel length. These conclusions could theoretically extend to planar resistive-switching devices based on polycrystalline thin films of other materials in which transition also takes place through a filamenting process, such as V_3O_5 [228], V_2O_3 [229], or NbO_2 [230].

3

Temperature-sensitive stochastic spiking neurons: different approaches to encoding

The manifest unity of our perceptions
belies the indeterminacy of our sensations.
– Misha Mahowald

To perform energy-efficient processing of data, Spiking Neural Networks require stimuli to be encoded in the timing of electrical spikes. As introduced in Chapter 1, sensitive spiking neurons based on VO_2 two-terminal devices are one of the existing options to perform such encoding. However, they suffer from several drawbacks. First, the intrinsic cycle-to-cycle variations in their electrical characteristics, which we studied in details in Chapter 2 by measuring them across several DC measurements. Indeed, when VO_2 memristors are used as spiking sensors, such *static* variations have been observed concomitantly with *dynamic* cycle-to-cycle variations in spike duration, or *jitter*. Presently, it is not yet clear how these static and dynamic stochastic phenomena are related, and whether the jitter could come from other noise sources. Among existing works on VO_2 spiking sensors, such frequency jitter is seldom quantified, and its impact on the sensor Limit of Detection (LoD) is rarely investigated. In this chapter, we use the resistive-switching VO_2 devices fabricated in Chapter 2 to design temperature-sensitive spiking neurons, and investigate the next research questions:

What are the dominant causes of dynamic stochasticity in rate-coded spiking VO_2 sensors? How does this stochasticity impact the sensor Limit of Detection?

3 | Temperature-sensitive stochastic spiking neurons: different approaches to encoding

As detailed in Chapter 1, one of the advantages of Spiking Neural Networks comes from the fact that they leverage event sparsity through their encoding. However, current implementations of VO_2 sensitive spiking neurons all use rate coding, while works have shown that rate-coded SNNs are not sufficient to outperform ANNs. Additionally, all these sensors operate in an "all-or-nothing" regime. Under a certain stimulus threshold, they will be completely OFF; above this threshold, they will be constantly spiking, with no dead times in between spikes. Only the *duration* of these spikes is affected by the stimulus. This makes us ask the following question:

Can VO_2 resistive-switching devices be used to design sensitive spiking neurons with sparser encoding techniques than plain rate-coding?

To answer these two research questions, the chapter will be structured in the following way. In the first section, we introduce the working principle behind the spiking neurons developed in this work, including the conditions for VO_2 devices to display an oscillating behavior, followed by the circuit interface and associated test setup allowing to trigger such spiking regime. Using an analytical dynamic model for the spiking output, we investigate the relative impact of different noise sources, both external and intrinsic to the VO_2 device. We show that the dominant source of spike rate jitter are the cycle-to-cycle variations in two specific electrical parameters of VO_2 two terminal devices. In the second section, we show how we use our spiking VO_2 neuron to perform rate coding of temperature. We quantify how the spike rate jitter impacts the sensor LoD, illustrating a trade-off between resolution and energy consumption. We explore different operating ranges of our spiking sensor, both in terms of temperature and spike rate, and see how the resolution-energy consumption trade-off evolves across these ranges. We then benchmark our implementation against other VO_2 spiking sensors. Finally, the third section explores a new kind of encoding by operating the VO_2 device at the limit of its stable oscillation regime, resulting in the generation of spike bursts whose temporal characteristics are temperature-dependent. We thoroughly characterize how the statistics of these bursts evolve with temperature, and discuss how such sensor could enable a sparser encoding of information.

All measurements in this chapter have been performed using our small-grains devices introduced in Chapter 2, since they offer lower cycle-to-cycle stochasticity. Results are adapted from our *ESSDERC 2023* conference paper (CP1) regarding the first research question, and from our journal paper submitted at *Electron Device Letter* (JP2) for the second question.

3.1 Stochastic spiking neurons

We start by detailing the operation of VO_2 two-terminal devices as constantly spiking neurons, i.e. when they are used within *relaxation oscillators*.

3.1.1 The negative differential resistance condition

For a VO_2 device to be able to produce oscillations, its I-V characteristics must posses a voltage-current region showing a *negative differential resistance* (NDR) when measured in current-driven mode. Indeed, when biased in this region, the device operates outside of its two stable states (metallic or insulator), and thus enters an astable regime in which it will alternate between the two states. The physical origin of such NDR has been thoroughly studied. Gibson [160] theorized that current-controlled NDR can, in principle, take place in any material whose electrical conductivity has a superlinear dependence on temperature, as it is the case for VO_2. In practice, he states that this dependence must be strong enough so that the NDR takes place at a sufficiently low temperature to prevent damaging the device.

The NDR corresponds to the phenomenon of redistribution of the current within the device channel upon transition of the device from its LRS to its HRS. This phenomenon can be captured through a simple equivalent circuit of two parallel resistances [231], as illustrated in Fig. 3.1a. Indeed, when the metallic filament forms, the current redistributes between the outer high-resistance "shell" (with equivalent resistance R_{shell}) which remains in the insulating domain, and the inner low-resistance "core" (with equivalent resistance R_{core}) formed by the metallic filament. The core presents a negative differential resistance R_{NDR}, since the voltage across it will decrease when increasing the current as the metallic filament expands. In literature, two types of NDR have been experimentally reported and captured using this equivalent circuit model:

- A *snap-back* response when $R_{shell} < R_{NDR,max}$, with the latter being the maximum NDR of the core. In this case, the current flows predominantly through the shell, but at lower current density since it has a greater area. This means that the total current can at first remain constant during the redistribution, hence the horizontal snap-back behavior.
- A *S-type* response when $R_{shell} >> R_{NDR,max}$. Here, the current flows predominantly through the metallic filament.

Typically, the shell resistance R_{shell} can be tuned by varying the device dimensions [166] (length and width), while R_{NDR} will depend on the OFF/ON resistivity ratio of the film (which can be tuned through the film oxygen content and

grain size, as shown in Chapter 2). When taking into account that both the metallic and insulating regions have temperature-dependent resistivities, coumpound snap-back and S-type behaviors can be observed [231].

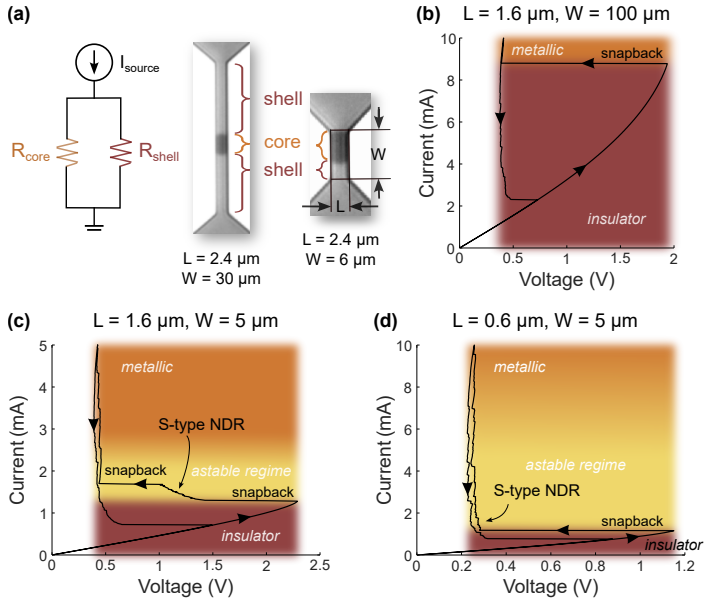


Fig. 3.1 **(a)** Parallel resistance model capturing the current distribution between the device metallic filament (core) and the insulating shell, as developed in [231]. These two components are visible through contrast between their optical constants on the high-resolution microscopy images of two of our VO_2 devices of different dimensions. **(b)-(d)** Three small-grain devices of different dimensions displaying various NDR behaviors. Insulator region is shown in red, metallic region in orange, and NDR region in yellow. **(b)** This device with dimensions L = 1.6 µm and W = 100 µm shows an NDR of pure snap-back type. **(c)** This device with dimensions L = 1.6 µm and W = 5 µm shows an S-type NDR nested between two snap-back responses. **(d)** This device with dimensions L = 0.6 µm and W = 5 µm shows snap-back NDR followed by a S-type NDR.

In the case of our small-grains devices, all dimensions present a snap-back behavior at some point in their I-V curve, as shown in Fig. 3.1. However, in device of smaller length and width (Fig. 3.1b-c), a S-type NDR also appears. This is coherent with the fact that the metallic filament (core) has a bigger area in proportion to the high-resistive shell in devices of smaller dimensions, as shown by our high-resolution microscopy images (Fig. 3.1a). The presence of a S-type NDR is required to design relaxation oscillators. Indeed, in device with a pure snap-back response, the reduction in voltage takes place at constant current, offering virtually no current range in which to bias our device so that it reaches an astable state. No matter the current fed through the device, the latter would be either in a stable insulator or metallic state.

In the remaining of this chapter, we will work exclusively with devices of smaller dimensions (L < 0.8 µm, W < 5 µm), as they present the most pronounced S-type behavior, offering a larger NDR current range in which to operate our relaxation oscillators. Devices of smaller channel area also have the advantage of displaying a spiking behavior at lower current and voltage level, reducing the system power consumption.

3.1.2 Circuit interface

Implementing a relaxation oscillator based on VO_2 two-terminal devices typically requires two additional circuit element: (1) an element controlling the current fed to the VO_2 two-terminal device, to ensure it is biased within its astable NDR region (see Fig. 3.2a) and (2) a capacitor which slows down the circuit dynamic, since the IMT takes place at ultrafast nanosecond timescales [232].

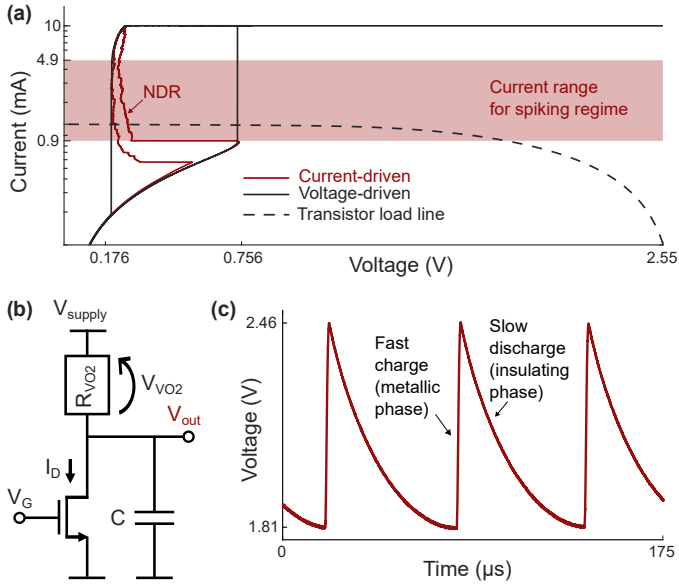


Fig. 3.2 **(a)** Voltage- and current-driven I-V characteristics of the VO_2 device (L = 0.8 µm, W = 3 µm) used within the relaxation oscillator, together with the NMOS load line, which falls in the current range for spiking regime (corresponding to the NDR). **(b)** Schematic of the spiking neuron circuit interface **(c)** Output spikes measured across the capacitor at T = 41 °C.

As mentioned in Chapter 1, the current is commonly set either using a scaling resistor or a transistor biased in its saturation regime. The latter option offers a finer tuning of the operating point, and we implement it in this work using a NMOS *ALD1106* discrete transistor. The drain current I_D is set through the gate voltage V_G. The full circuit is shown in Fig. 3.2b. For the following series of measurements,

circuit parameters are the following: $C = 33$ nF, supply voltage $V_{supply} = 2.55$ V, $V_G = 2.09$ V and $I_D = 1.025$ mA. The specific device on which measurements have been realized has dimensions L = 800 nm, W = 3 µm. The measurement of a typical output waveform is shown in Fig. 3.2c. During the abrupt charging phase, the VO_2 is in its metallic regime: the capacitor C is charged by the large current flowing through the low VO_2 resistance, increasing V_{out} and progressively decreasing the voltage drop V_{VO2} across the device. Once V_{VO2} drops to the lower threshold voltage V_{MIT}, R_{VO2} switches back to its insulating regime: the capacitor slowly discharges through the least resistive path, i.e. the NMOS transistor, and V_{out} decreases. V_{VO2} thus increases up to the higher threshold voltage V_{IMT}, triggering the transition towards the metallic phase and starting a new cycle. The dynamics of the oscillations are dictated by the circuit RC constant, which varies as the VO_2 resistance changes from the metallic to the insulator regime.

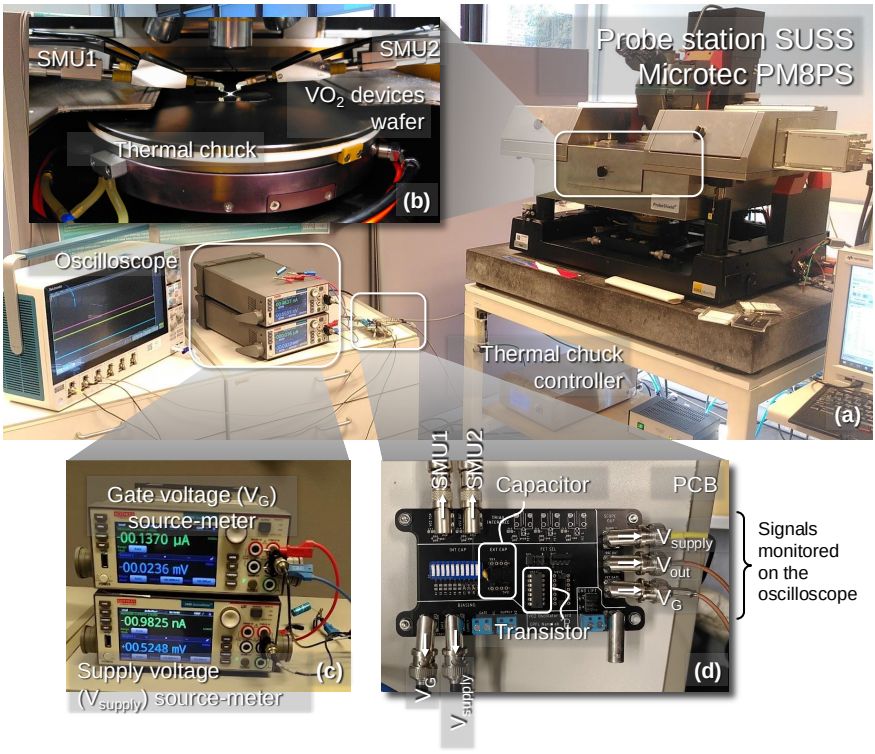


Fig. 3.3 **(a)** Full test setup allowing to measure the VO_2 relaxation oscillators under dynamic operation. **(b)** Close-up on the interior of the probe station allowing to contact the VO_2 devices directly on the wafer. **(c)** Gate and supply voltage source-meters **(d)** PCB implementation of the interface circuit, also providing coaxial output to measure the output, gate and supply voltages.

The full test setup allowing to measure these oscillations is shown in Fig. 3.3a. The VO_2 two-terminal devices are contacted directly on the fabricated wafer using an electromagnetically shielded probe station (Fig. 3.3b), allowing to test the oscillating behavior of a large variety of devices without requiring dicing and bonding steps. A thermal chuck allows to control the wafer temperature. The device is interfaced with the circuit using triaxial cables towards a PCB implementation which holds the capacitor and NMOS transistor, fabricated by our research partners at the *EPFL Nanolab* (Fig. 3.3d). Gate and supply voltages are set using two *Keithley 2450* source-meters (Fig. 3.3c) connected with coaxial cables, the second also allowing to monitor drain current. A *Tektronix MSO56* oscilloscope with sampling frequency up to the GHz allows to acquire high-resolution measurements of the oscillations.

3.1.3 Analytical dynamic model

The time-dependent equations for the output voltage of the spiking sensor can be derived from the fact the system behaves as an RC circuit. There will be a successive charge and discharge of the capacitor, bounded between the two threshold voltages of the VO_2 device. The dynamic evolution of the output voltage can thus be modeled by two exponential functions, whose time constants $\tau = R_{VO2}C$ depend on the resistance of the VO_2 device in the two phases, and whose parameters can be determined by setting the appropriate boundary conditions.

In the **charge phase**, while $V_{out} < V_{supply} - V_{MIT}$, the VO_2 resistor is in its metallic state ($R_{VO2} = R_{met}$). The charge of the capacitor can be modelled by an function of the following type:

$$A(1 - e^{-(t-t_c)/\tau}) + B \quad , \tag{3.1}$$

with t_c the time at the beginning of the charge phase, and τ the circuit RC constant. The boundary conditions at the beginning of the charge phase ($t = t_c$) and at asymptotically long times ($t = \infty$) are:

$$\begin{aligned} V_{out}(t = t_c) &= V_{supply} - V_{IMT} \\ V_{out}(t = \infty) &= V_{supply} - R_{met} I_D \end{aligned} \tag{3.2}$$

This yields the following expression for the circuit output voltage:

$$\begin{aligned} V_{out}(t) = (V_{IMT} - R_{met} I_D) \left(1 - e^{-(t-t_c)/R_{met}C}\right) \\ + (V_{supply} - V_{IMT}); \end{aligned} \tag{3.3}$$

In the **discharge phase**, while $V_{out} > V_{supply} - V_{IMT}$, the VO_2 resistor is in its insulating state ($R_{VO2} = R_{ins}$). The discharge of the capacitor can be modelled by a function of the following type:

$$Ce^{-(t-t_d)/\tau} + D \quad , \tag{3.4}$$

with t_d the time at the beginning of the charge phase, and τ the circuit RC constant. The boundary conditions at the beginning of the discharge phase ($t = t_d$) and at asymptotically long times ($t = \infty$) are:

$$\begin{aligned} V_{out}(t = t_d) &= V_{supply} - V_{MIT} \\ V_{out}(t = \infty) &= V_{supply} - R_{ins} I_D \end{aligned} \tag{3.5}$$

This yields the following expression for the circuit output voltage:

$$\begin{aligned} V_{out}(t) = (R_{ins} I_D - V_{MIT}) e^{-(t-t_d)/R_{ins}C} \\ + (V_{supply} - R_{ins} I_D). \end{aligned} \tag{3.6}$$

Similar mathematical models have been developed in previous works [91, 159, 233]; however, in all these models, the resistance in the insulating state is approximated as a constant in the $[V_{MIT}, V_{IMT}]$ voltage range, while it does in reality vary with the applied voltage through Joule heating (see Fig. 3.2a). We capture this phenomenon by including an empirical expression for the voltage dependence of R_{ins}:

$$R_{ins}(V_{VO2}) = R_{ins,0} - ae^{bV_{VO2}} \tag{3.7}$$

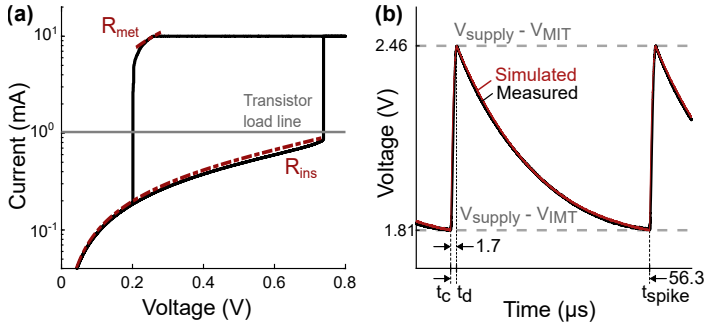


Fig. 3.4 **(a)** Measured voltage–driven DC sweep of the VO_2 device (L = 0.8 µm, W = 3 µm) at T = 41 °C (black solid line). Superimposed: transistor drain current approximated as constant (grey solid line) and fitted resistances in metallic (constant) and insulating (voltage-dependent) regimes (red dashed lines), as used in the model. **(b)** Comparison between the dynamic model of Eq. 3.3 and 3.6 (red) and the measured waveform (black).

We fit the resistances on the I-V characteristic measured at 41 °C (Fig. 3.4a), and obtain $(50\ \Omega, 2.276\ V^{-1})$ for the fitting parameters (a, b), $R_{ins,0} = 1088.53\ \Omega$, and $R_{met} = 21\ \Omega$. The other circuit parameter (C, V_{supply}, and I_D) correspond to the operating parameters specified above. The threshold voltages V_{IMT} and V_{MIT} are extracted from the average min. and max. values of the measured spikes, in coherence with the DC measurements. Our deterministic, 9-parameters model reproduces spike trains with high fidelity (Fig. 3.4b). The model neglects the Early effect and considers the transistor current I_D to be constant.

3.1.4 Dynamic stochastic analysis

To this dynamic analytical model capturing the output voltage of our relaxation oscillator, we can add various noise sources and cycle-to-cycle variations. The goal is to extract a theoretical spike frequency jitter from this new stochastic model, and to compare it with the jitter obtained from measurements of the stochastic oscillations of our device. Following this procedure, we can determine which noise sources are predominantly responsible for the observed spike rate jitter.

Cycle to cycle variations

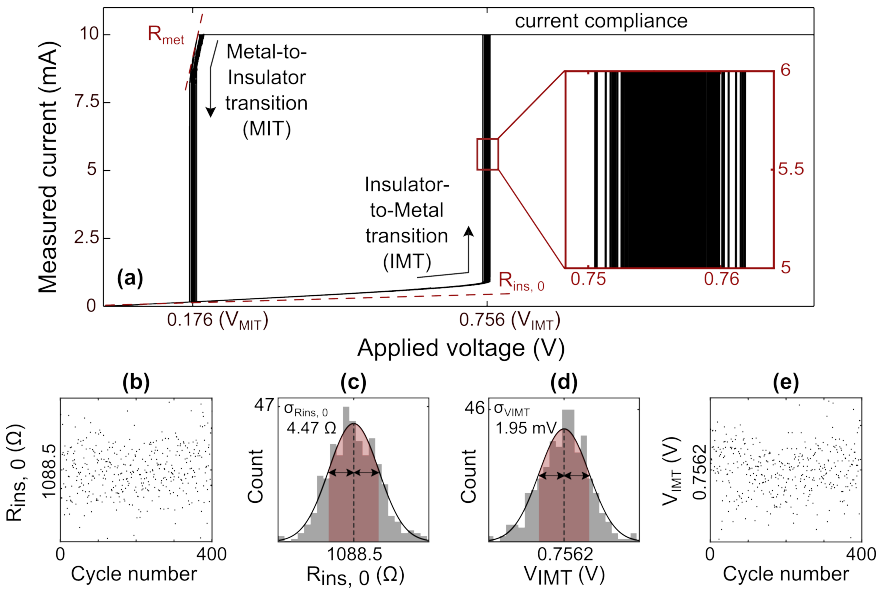


Fig. 3.5 **(a)** 400 consecutive voltage-driven DC sweeps measured on the VO_2 device (L = 0.8 µm, W = 3 µm), at T = 41 °C, with current compliance to protect the device. Close-up of the cycle-to-cycle variations in threshold voltage V_{IMT} shown in inset. **(b)-(e)** Distributions of the resistance in insulating regime $R_{ins,0}$ and the threshold voltage V_{IMT} , with their respective mean and standard deviation. Scatter plots show the absence of drift over the consecutive cycles.

We first focus on characterizing the **cycle-to-cycle variations** of the four key electrical parameters of the specific device used in our relaxation oscillator. These parameters are its two threshold voltages V_{IMT} and V_{MIT}, its resistance in the metallic regime R_{met}, and its resistance in the insulating regime close to zero bias $R_{ins,0}$. To study the statistics of these parameters, we performed 400 high-resolution voltage-driven DC measurements of the device I-V characteristics (Fig. 3.5a). The methodology is similar to the one introduced in Chapter 2. Fig. 3.5b-e show the Gaussian fitting and absence of drift for V_{IMT} and $R_{ins,0}$, which will be proved to be the dominant causes of the spike rate jitter. We obtain coefficients of variations CV_i (defined as the parameter standard deviation σ_i normalized by its mean μ_i) of 0.41, 0.042, 0.26 and 0.28 % for $R_{ins,0}$, R_{met}, V_{IMT} and V_{MIT}, respectively.

Other noise sources

In addition to the cycle-to-cycle variations of the device electrical parameters, we consider a number of other noise sources: the **transistor thermal noise**, and the **thermal noise on the VO_2 resistor**. These noise sources are considered as additive white Gaussian noises, and their root-mean square (RMS) values thus correspond to the standard deviation of their distribution. We make the hypothesis that these two noise sources are independent, and we model them as parallel current sources adding to the NMOS drain current I_D, as illustrated in Fig. 3.6. The validity of this independence hypothesis will be discussed below. The RMS value of a transistor thermal noise current when operated in saturation regime can be expressed (in amperes) by the following expression [234]:

$$I_{rms,th,NMOS} = \sigma_{th,NMOS} = \sqrt{4k_B\gamma g_m T\Delta f} \quad [A], \tag{3.8}$$

with g_m the NMOS transconductance equal to 1.8 mS from its datasheet; γ a coefficient equal to 2/3 for long-channel transistors (in the micron scale); and Δf the bandwidth set to half of the sampling frequency f_s (125 MHz in the case of our measurements, yielding a 62.5 MHz bandwidth).

The RMS value of the VO_2 resistor thermal noise can be expressed (in amperes) by the expression:

$$I_{rms,th,VO2}(t_i) = \sigma_{th,VO2}(t_i) = \sqrt{4k_B T\Delta f / R_{VO2}(t)} \quad [A] \tag{3.9}$$

This RMS value will differ between the charge and the discharge phase. During the charge phase, the VO_2 resistor is in its metallic regime, and $R_{VO2}(t) = R_{met}$. During the discharge phase, the VO_2 resistor is in its insulating regime. However, from Eq. 3.7, the resistance is modelled as voltage-dependent, and will thus vary over the course of the discharge phase, together with the noise RMS value.

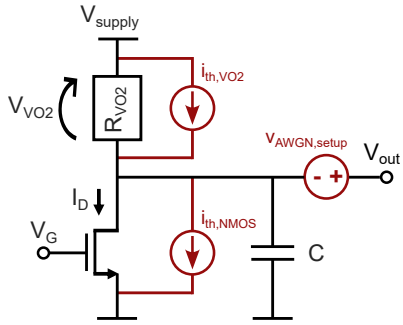


Fig. 3.6 Equivalent noise circuit with the two parallel current noise sources corresponding to the thermal noise on the NMOS transistor and on the VO_2 resistor, and one series voltage noise source modelling the setup noise.

As we considered the two noise sources as independent, the total value of the current $I_{D,noisy}$ at a given moment t in time can be included in our stochastic model by sampling values from the parallel noise sources using their Gaussian distribution, and adding them to the NMOS current I_D:

$$I_{D,noisy}(t) = I_D + i_{th,NMOS}(t) + i_{th,VO2}(t) \tag{3.10}$$

We also account for the additive white Gaussian voltage noise (AWGN) coming from the electrical setup itself, due to e.g. electromagnetic interference and power supply noise. Its RMS value is deduced empirically by measuring the output voltage of the circuit when biased just below its stable oscillation regime. It is modelled as a voltage source in series with the output, and can be included in our stochastic model by adding a value sampled from the AWGN distribution to the output voltage:

$$V_{out,noisy}(t_i) = V_{out}(t) + v_{AWGN,setup}(t_i) \tag{3.11}$$

Stochastic analytical dynamic model

Now that we have parameters describing the statistics of our different noise sources and cycle-to-cycle variations, we can include them in our dynamic analytical model for the relaxation oscillator output voltage, described in Eq. 3.3 and 3.6. To include cycle-to-cycle variations, new values of (V_{MIT}, R_{met}) and $(V_{IMT}, R_{ins,0})$ are sampled from their empirically deduced distributions (see Fig. 3.5) at the beginning of each charge and discharge cycle, respectively. The inclusion of the thermal and setup noise sources is straightforward: at each time step t, we replace the constant transistor current value I_D by the value computed from the sampling of the current thermal noise distributions, as described in Eq. 3.10. We then add the sampled value from the AWGN setup voltage noise to the resulting V_{out} expression. We thus obtain a **stochastic analytical dynamic model** for the oscillator output.

Table 3.1 Contribution of the different noise sources to the spike rate jitter at T = 41 °C

Noise source	(1) Cycle-to-cycle variations in V_{MIT}	(2) Cycle-to-cycle variations in R_{met}	(3) Cycle-to-cycle variations in V_{IMT}	(4) Cycle-to-cycle variations in $R_{ins,0}$
Statistical parameters	$\mu_{VMIT} = 0.088\ V$ $\sigma_{VMIT} = 0.24\ mV$	$\mu_{Rmet} = 21\ \Omega$ $\sigma_{Rmet} = 0.87\ \Omega$	$\mu_{VIMT} = 0.74\ V$ $\sigma_{VIMT} = 1.95\ mV$	$\mu_{Rins,0} = 1088.53\ \Omega$ $\sigma_{Rins,0} = 4.47\ \Omega$
Simulated $\sigma_{f,i}$ (Hz)	4.43	20.00	232.32	263.37

Noise source	(5) Thermal noise on R_{met} and R_{ins}	(6) Noise on transistor current	(7) Total AWGN extracted from setup	**Final model**	Measu-rements
Statistical parameters	$\sigma_{th,met} = 227\ nA$ $\sigma_{th,ins} = 32\ nA$	σ_{NMOS} $= 36\ nA$	σ_{AWGN} $= 1.56\ mV$	**All noise sources combined**	
Simulated $\sigma_{f,i}$ (Hz)	2.68	2.56	0.81	$\sigma_{f,sim}$ $= 353.64$	$\sigma_{f,meas}$ $= 344.82$

The respective contribution of each noise source to the resulting spike rate jitter can be extracted by including them individually in the dynamic model. Results are shown in Table 3.1. For each noise source, the first row specifies the mean and standard deviation of the distribution. For the thermal noise on the VO_2 resistor in the insulating state, this value is given for the mean value of $R_{ins,0}$ (around zero bias), but it will vary at each modelled time step of the discharge phase, as explained above. The second row presents the individual impact $\sigma_{f,i}$ of the noise source on the spike rate jitter. The plausibility of our stochastic dynamic model is verified when including all the different noise sources, and comparing the obtained theoretical spike rate jitter with the one deduced from a measurement of the oscillator output. There is a good match between the theoretical and observed distribution of the instantaneous spike frequency, as shown in Fig. 3.7.

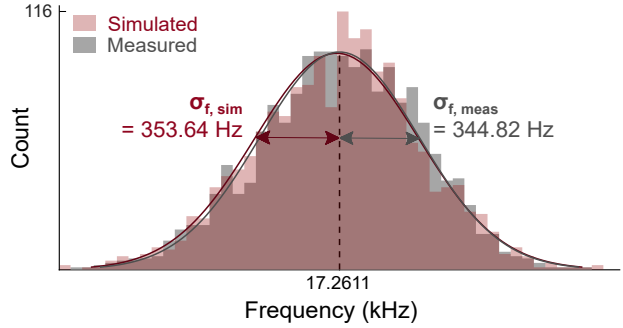


Fig. 3.7 Concordance between measured and simulated jitter on the spike rate, measured and simulated over 100 ms. The 8 ns timestepping used to generate the waveform using the dynamic stochastic model corresponds to the 125 MHz sampling frequency of the measured waveform. All the noise sources of Table 3.1 have been included in the stochastic dynamic model.

Looking at Table 3.1, it is clear that the dominant contributions to the spike rate jitter come from the cycle-to-cycle variations, and more specifically from the cycle-to-cycle variations on the threshold voltage V_{IMT} and the insulating resistance $R_{ins,0}$ (column 3 and 4). The impact of the thermal and setup noises are of lower magnitude (criteria 5, 6). These results bring additional insights to previous literature works, which usually present the cycle-to-cycle variations in V_{IMT} as the sole source of the spike rate jitter [207]. Indeed, our model is able to capture the significant contributions of the R_{ins} variations, as it accounts for its voltage-dependence. Considering it constant throughout the discharge phase would result in systematic underestimation of its variations impact. Additionally, our analysis contradicts one of the hypothesis often mentioned in the literature, which states that the cycle-to-cycle variations in R_{met} and R_{ins} could be due to thermal noise [92]. Indeed, the large difference in their effect on the spike rate proves they are distinct phenomena. The hypothesis related to the independence of the noise sources is consistent with the fact that the condition $\sigma_f \simeq \sqrt{\Sigma \sigma_{f,i}^2}$ is satisfied.

The match between our dynamic stochastic model and the measured frequency jitter does not mean that all the noise sources present in reality were captured with exactitude or comprehensiveness. However, it does prove that the cycle-to-cycle variations in the device electrical parameters dominate the spike rate jitter. The total equivalent magnitude of the remaining noise sources is estimated correctly by our model, but could possibly include other noise sources than the one considered here (i.e., flicker noise, random telegraph noise, ...).

3.2 Rate-based temperature encoding

3.2.1 Sensitivity and Limit of Detection

The spike rate is modulated by the variation of the VO_2 I-V characteristics with temperature, shown in Fig. 3.8a for $T = 41 - 47$ °C. Fig. 3.8b-c show the inversely proportional relationships of $R_{ins,0}$ and V_{IMT} to temperature, which result in an increase in spike rate (Fig. 3.8d). The modulation of the insulating resistance and insulator-to-metal threshold voltage of VO_2 two-terminal devices I-V characteristics has recently been modelled by a combination of Joule heating and field-assisted carrier generation mechanisms [92]. These field-induced carriers are believe to arise through the Poole-Frenkel effect, under which the applied electric field reduces the excitation barrier for carriers trapped within in-gap states [226, 235]. The reverse transition and associated evolution of V_{MIT} with temperature is usually captured using Newton's law of cooling [134]. The evolution of the threshold voltage V_{IMT} with temperature can give us information on the dominant type of mechanism (Joule heating or field-induced carriers).

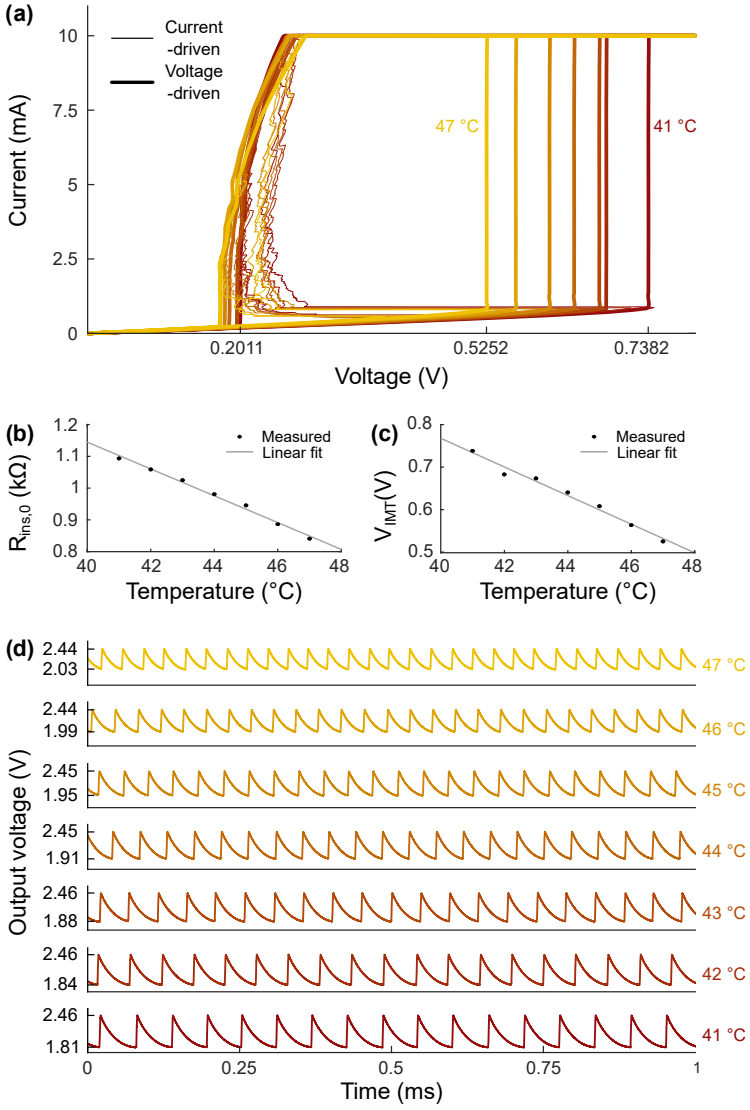


Fig. 3.8 **(a)** Voltage- and current- driven I-V characteristics of the VO_2 device for T = 41 - 47 °C. **(b)(c)** Evolution of $R_{ins,0}$ and V_{IMT} with temperature, with linear fit. **(d)** Output voltage of the spiking sensor circuit, showing spiking behavior in the 41 - 47 °C range with increase in spike frequency and reduction in spike amplitude.

In the narrow temperature range considered in the measurements, the trend seems to be linear, but should be investigated on a larger temperature range to allow to discriminate the mechanism, as it will be discussed below. As found previously using our analytical dynamic model (Eq. 3.3 and 3.6), the variations in spike rate are dominated by variations in V_{IMT} and $R_{ins,0}$. The reduction of these two parameters with temperature is thus the dominant cause for the increase in spike rate, visible in Fig. 3.8d. This increase is accompanied by a reduction in spike amplitude, due to the fact that the relaxation oscillator output is bounded between $\left[V_{supply} - V_{IMT}; V_{supply} - V_{MIT}\right]$, and that the hysteresis width $V_{IMT} - V_{MIT}$ reduces with temperature.

From the linear dependence of the spike rate with temperature, we can define a spike frequency sensitivity S_f. To obtain this sensitivity, we perform a linear fit of the average spike rate $\overline{f}_{spike}$ at each temperature. This average spike rate is obtained by taking the mean of the instantaneous frequency f_{spike} of each measured spike (defined as $1/T_{spike}$, the inverse of the spike duration), over the longest observation duration available (in our case, 100 ms). We obtain a sensitivity S_f of 1.71 $kHz/°C$ (Fig. 3.9). The extraction of the sensitivity based on the longest duration available measurement can be seen as a *calibration step*.

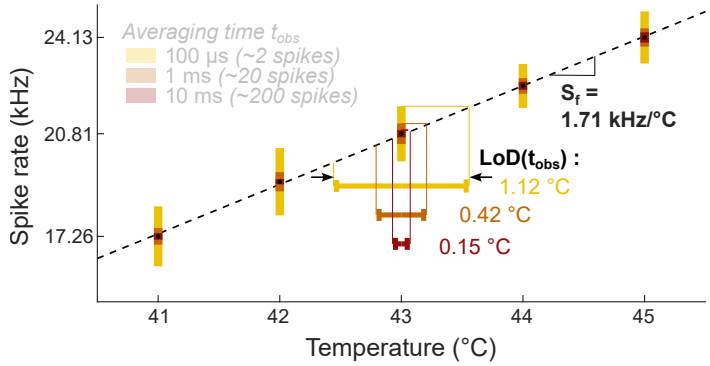


Fig. 3.9 Mean frequencies for T = 41 - 45 °C, together with their 99.7 % confidence interval, for three different averaging times (100 µs, 1 ms, 10 ms). Sensitivity is derived from a linear fit. Temperature resolution is derived for the worst case of spike rate jitter, i.e. T = 43 °C.

During the *actual operation of the sensor*, we can deduce the measured temperature using this sensitivity value. However, even assuming perfect calibration, there will be some amount of uncertainty related to the spike rate jitter. To overcome this uncertainty, the instantaneous frequency can be averaged over a certain number of cycles, corresponding to a duration of observation t_{obs}. We propose a two-step statistical estimation of spike rate uncertainty when averaged over a duration t_{obs}.

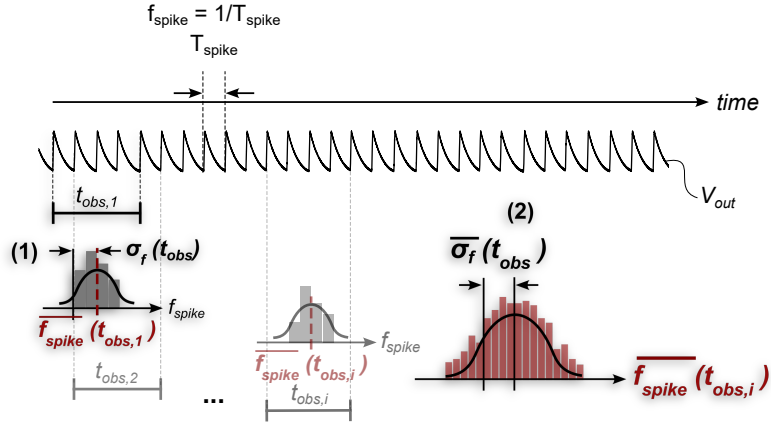


Fig. 3.10 Definition of the spike rate uncertainty associated to the averaging of the instantaneous spike frequency over a certain observation duration t_{obs}. First, the mean frequency $\overline{f_{spike}}(t_{obs,i})$ of the instantaneous spike rates is computed over each interval $t_{obs,i}$. Second, the distribution of these local mean spike rates $\overline{f_{spike}}(t_{obs,i})$ is fitted to a Gaussian distribution, allowing to extract a standard deviation $\overline{\sigma_f}(t_{obs})$ which characterizes the uncertainty of the averaging process. This parameter differs from $\sigma_f(t_{obs,i})$, the instantaneous spike rate standard deviation extracted over a given measured interval $t_{obs,i}$.

First, we take the mean $\overline{f_{spike}}(t_{obs,i})$ of the instantaneous spike rates over each interval $t_{obs,i}$ within the measured waveform; then, we fit the distribution of these local mean spike rates $\overline{f_{spike}}(t_{obs,i})$ to a Gaussian distribution. We can then obtain a standard deviation $\overline{\sigma_f}(t_{obs})$, which captures the spike rate uncertainty when performing averaging on a given observation duration[1]. This procedure is illustrated in Fig. 3.10. Knowing this uncertainty, we can define the spike rate resolution for a given observation duration using a 99.7 % (3σ) confidence interval. From the frequency sensitivity to temperature S_f, it can be converted to a temperature resolution or *Limit of Detection* (LoD):

$$LoD_{3\sigma}(t_{obs}) = 2\frac{3\overline{\sigma_f}(t_{obs})}{S_f} \quad (^\circ C) \tag{3.12}$$

Fig. 3.9 illustrates the estimation of this LoD for three different observation duration at temperature T = 43 °C: the LoD improves with the averaging time, from 1.12 °C for a 100 μs observation to 0.15 °C for a 10 ms observation. This indicates a clear resolution-energy trade-off: one order of magnitude of LoD can be gained at the cost of increasing the associated energy consumption by 100, through a 100-fold increase in observation duration.

[1]It is noteworthy that the parameter $\overline{\sigma_f}(t_{obs})$ differs from $\sigma_f(t_{obs,i})$, the standard deviation of the instantaneous spike frequency extracted on a given measured interval $t_{obs,i}$.

So far, it seems that increasing the averaging duration yields ever smaller LoD. However, distinct noise processes can take place at different timescales, and there might be a limit to this improvement. To answer this question, an interesting tool allowing to assess the frequency stability on different timescales is the *Allan variance*. The Allan variance was first introduced in 1966 by David W. Allan to study the frequency deviation of oscillators with respect to their nominal frequency [236]. It provides a prediction of how far the clock frequency is likely to drift in an interval of m periods into the future, based on an averaging of many of such intervals in the past. It does not provide a unique value, but rather depends on the observation duration t_{obs}, and is thus commonly plotted as a function of the latter.

Defining the Allan variance requires to extract the time deviation vector $x[n]$: a time series which describes, for each measured cycle of the oscillator, by how much the cycle duration deviates from an ideal reference period $T_{spike,0}$. The Allan variance σ_y^2 is then defined by averaging these time deviations over an observation duration t_{obs}:

$$\sigma_y^2(t_{obs}) = \frac{1}{2t_{obs}^2} E\left[(x[n+2] - 2x[n+1] + x[n])^2\right] \tag{3.13}$$

Performing such averaging involves a statistical expected value, integrated over an infinite number of intervals. In practice, an estimator is required to approach such expected value. A commonly used estimator for the Allan variance uses overlapped time intervals within the original measured time series [237], and is defined for an observation duration t_{obs} corresponding to m nominal periods of duration $T_{spike,0}$. The estimator uses a total of $N - 2m$ overlapped time intervals:

$$\sigma_y^2(mT_{spike,0}) = \frac{1}{2m^2T_{spike,0}^2(N-2m)} \sum_{i=1}^{N-2m} (x[i+2m] - 2[i+m] + x[i])^2 \tag{3.14}$$

The Allan deviation is obtained from the Allan variance as:

$$\sigma_y = \sqrt{\sigma_y^2} \tag{3.15}$$

The Allan deviation is unitless, but can be interpreted in terms of frequency stability by multiplying it by the oscillator nominal frequency $f_{spike,0} = 1/T_{spike,0}$. The result is shown in Fig. 3.11 for the 100 ms measurements of our VO_2 oscillator in temperature range T = 41 - 47 °C. The reference period T_{spike} is extracted as the mean of all spikes periods on the full 100 ms observation duration.

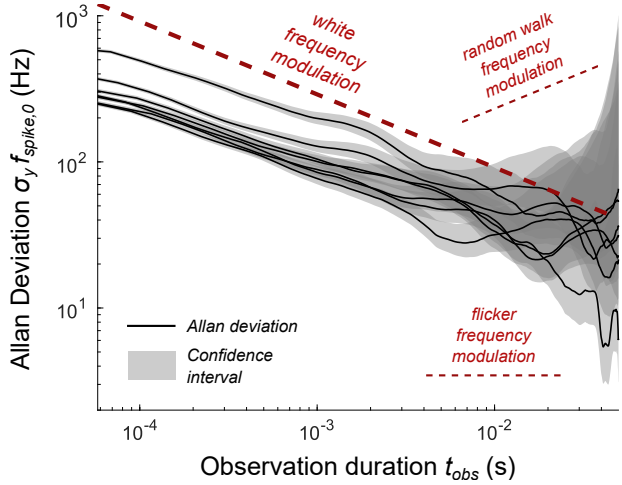


Fig. 3.11 Allan deviation of the seven 100 ms-long measured waveform of the VO_2 oscillator output in the 41 - 47 °C temperature range. No clear temperature-dependency trend has been observed, which is why the curves are all plotted without distinction. The power-law for three distinct noise sources are plotted in red. The confidence intervals on the Allan deviation estimator of each measurement are plotted in gray (see Appendix B for detailed expression).

Distinct noise sources will have different power laws when passed through the Allan deviation tool [238], which are plotted in red dashed lines on Fig. 3.11. One can see that the measured frequency variations are dominated by white frequency modulations, while flicker and random walk noises do not seem to appear within the considered timescale (up to 100 ms, the measurement duration). This validates the fact that we only included white noise sources within our stochastic dynamic model presented above. However, one can note that the uncertainty on the Allan deviation (plotted in gray) increases as the observation duration increase. Non-white noise processes could thus take place towards longer time scales. However, we can be confident that performing averaging will reduce uncertainty on the spike rate for observation duration up to ∼10 ms.

3.2.2 Extended temperature and frequency operation range

After defining the LoD and its associated energy-resolution trade-off, it is interesting to investigate how our VO_2 spiking temperature sensor behaves across a larger temperature and frequency operation range. To that end, we study the behavior of a second VO_2 relaxation oscillator in the 25 - 45 °C temperature range. This second oscillator is built from a VO_2 device of identical dimensions as the one studied above (L = 0.8 µm, W = 3 µm), but uses a smaller output capacitor of 330 pF (w.r.t. 33 nF in the previous case).

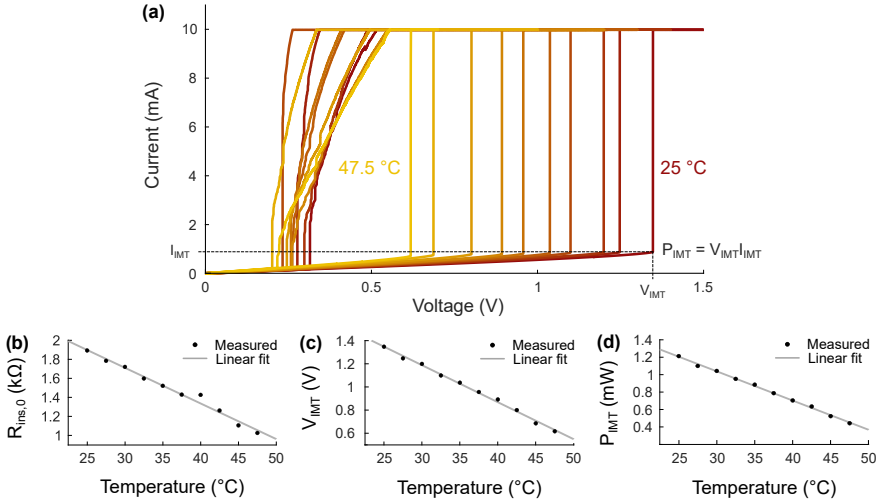


Fig. 3.12 **(a)** Voltage-driven I-V characteristics of the VO_2 device for T = 25 - 47.5 °C, with 2.5 °increments. **(b)(c)** Evolution of $R_{ins,0}$ and V_{IMT} with temperature, with linear fit. **(d)** Temperature dependency of the power P_{IMT} required to trigger the transition. All three trends are fitted with a linear curve.

The evolution of the VO_2 device I-V characteristics over the 25 - 47.5 °C temperature range is shown in Fig. 3.12. The linear dependence of the threshold voltage V_{IMT}, observed above on a narrower temperature range (Fig. 3.8c), is confirmed on this larger temperature range. This hints that the transition is not dominated by a Poole-Frenkel mechanism, otherwise a decreasing exponential behavior of the type $V_{IMT} \propto exp(-T/T_0)$ would be expected [134, 239]. On the contrary, the linear temperature dependence of the power $P_{IMT} = V_{IMT} I_{IMT}$ required to trigger the insulator-to-metal transition, shown in Fig. 3.12d, suggests that the transition is dominated by Joule heating [172, 240].

The dependency of the mean spike rate on the temperature from room temperature (25 °C) up to the maximum frequency of oscillation (45 °C) is shown in Fig. 3.13a. The mean oscillator frequency was extracted from 10 ms-long measurements acquired at 1.25 GHz, every 2.5 °C. The device stopped oscillating at T = 47.5 °C, which we attributed to the quasi extinction of the NDR range (see yellow curve in Fig. 3.12a). Whereas the spike rate temperature dependency could be considered linear when studied on a narrow temperature range (see Fig. 3.9), it becomes exponential when looking at the full temperature range. This evolution can be easily reproduced when plugging the linear temperature dependency of V_{IMT} and $R_{ins,0}$ into our dynamic analytical model presented above. Due to the smaller 330 pF output capacitor with respect to the first oscillator, the VO_2 spiking sensor operates at much higher frequencies: between 400 and 1000 kHz in the 25 - 45 °range (whereas

the first oscillator with C = 33 nF oscillated between 17 and 24 kHz in the $41 - 45$ °range). However, in order to correctly fit the spike rate temperature dependence, the dynamic analytical model assumes a larger, 1.15 nF capacitor (see Fig. 3.13a). We thus theorize that there is a large equivalent parasitic capacitance within our experimental setup, which could be due to the long triaxial cable used to contact the VO_2 device through the probe station (see the setup shown in Fig. 3.3). A parasitic capacitance of similar order of magnitude has been estimated in other VO_2 oscillators [241]. Alternatively, this "saturation" in the oscillator frequency when reducing the capacitor could also be attributed to the frequency dependence of the VO_2 device IMT point and resistance in its insulating state [242].

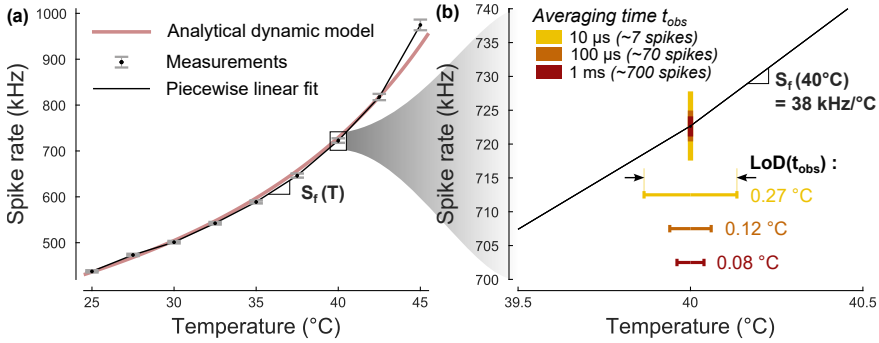


Fig. 3.13 **(a)** Spike rate sensitivity to temperature of the higher-frequency VO_2 relaxation oscillator (L = 0.8 µm, W = 3 µm, C = 330 pF), measured (black) and modelled (red). The dots correspond to the mean frequency, averaged over 10 ms-long measurements. Errorbars correspond to the standard deviation $\sigma_f(t_{obs} = 10\ ms)$ of the instantaneous frequency distribution. Sensitivity is defined in a piecewise fashion. **(b)** LoD computed at T = 40 °C for three averaging times.

A relevant question is to ask how the LoD-energy consumption trade-off evolves at such higher oscillation frequency. To answer this question, Fig. 3.13b illustrates this trade-off at a temperature of 40 °C. The sensitivity is approximated locally by performing a piecewise linear fitting between successive temperature points. Once again, increasing the observation duration allows to reduce the LoD. Interestingly, at similar temperatures, our second oscillator provides better resolution performances at equivalent energy costs: for an observation duration of 1 ms, it achieves 0.08 °C LoD, whereas the first oscillator showed a 0.42 °C resolution for the same t_{obs}. The question can also be answered by setting a LoD requirement: to achieve a 0.1 °C LoD, the first oscillator requires an averaging time of 35 ms, while the second only needs a 250 µs observation time. To explain such behavior, let's recall that the dominant source of spike rate jitter are the cycle-to-cycle variations in the VO_2 device electrical characteristics, as demonstrated above. When the oscillator is run at a higher frequency, averaging these variations over a given number of cycles requires a smaller observation time.

Table 3.2 Comparison between the characteristics of two VO_2 oscillators with LoD analysis at 43 and 40 °C, respectively, over the same number of cycles.

	(1) Device dimensions	**(2)** C	**(3)** T	**(4)** n_{cycles}	**(5)** t_{obs}	**(6)** LoD	**(7)** S_f	**(8)** μ_f	**(9)** $\frac{S_f}{\mu_f}$	**(10)** $\overline{\sigma_f}(t_{obs})$	**(11)** $\frac{\overline{\sigma_f}(t_{obs})}{\mu_f}$
Oscillator 1	L = 0.8 µm W = 3 µm	33 nF	43 °C	100	4.8 ms	0.21 °C	1.71 kHz/°C	20.8 kHz	8.2 %	60 Hz	0.3 %
Oscillator 2	L = 0.8 µm W = 3 µm	330 pF	40 °C	100	13.8 µs	0.11 °C	38 kHz/°C	722.7 kHz	5.3 %	705 Hz	0.1 %

However, this implies that the intrinsic spike rate jitter remains of the same order of magnitude when increasing the oscillation frequency of the neuron. This is confirmed by our observations. Table 3.2 compares the two oscillators along 11 numbered criteria. More specifically, it shows the LoD of both oscillators **(6)** at two close temperatures **(3)**, when extracted on an observation duration **(5)** corresponding to 100 cycles, i.e. 100 spikes. While the uncertainty $\overline{\sigma_f}(t_{obs})$ associated to the averaging is higher in absolute value in the second high-frequency oscillator **(10)**, it is actually of the same order of magnitude when normalized by the mean spike frequency **(11)**. The sensitivity **(7)**, when normalized by the mean frequency **(8),(9)**, is also of the same order of magnitude in the two cases, hence the preserved LoD. This observation seems to prove that the dynamic cycle-to-cycle stochasticity does not worsen when increasing the oscillator frequency. From our investigations detailed in Section 3.1.4, we showed that the cycle-to-cycle stochasticity in spike frequency were dominated and could be fully predicted by the cycle-to-cycle stochasticity of the device electrical characteristics, extracted in DC. Importantly, we thus conclude that for our specific device topology and VO_2 thin film, the cycle-to-cycle variations in the device electrical characteristics seem to remain of the same order of magnitude for timescales ranging from 10^2 (DC measurements) to 10^{-5} seconds (100 kHz oscillations).

3.2.3 Benchmarking

We compare our VO_2 spiking temperature sensor to other existing implementations of VO_2 spiking sensors. The full benchmarking is detailed in Tables 3.3 and 3.4, performed against 18 numbered criteria. We compare both the low- an high-frequency versions of our oscillators through several key parameters.

Compactness

First, we compare the number of components (**criterion 3** in the Table). Our implementation requires three components: the VO_2 memristor, a capacitor and a transistor. It is more compact than certain implementations [93, 96], thanks to the fact that the stimulus transduction is performed directly by the VO_2 device, whereas other rely on external sensors (which often require series resistance to scale the current within the oscillation region). The most compact implementation [92] requires only one component: the VO_2 device. To that end, it uses a current source instead

of a voltage source (removing the need for a transistor or scaling resistor to tune the current), and exploits the parasitic capacitance within the circuit to integrate the variations. Our implementation could thus be made more compact by including these two design choices; however, the transistor used in saturation regime allows a fine tuning of the current, a feature which will be exploited in the next section. As for the capacitance, a remaining question is whether it comes from the wiring between components (as most of these proof of concepts are realized using PCBs), or if it is truly intrinsic to the VO_2 device itself, as suggested in certain works [92, 181]. In the latter case, truly compact implementation of VO_2 sensors monolithically integrated with CMOS technology could be envisioned. In that perspective, compactness can also be compared by looking at the device footprint, defined by its active area $A = W \cdot L$ (**criterion 4**). The 2.4 µm² footprint of our VO_2 device doesn't beat the state of the art 0.4 µm² area of other planar devices [96,181].

Power and energy consumption

The power consumption of our spiking sensor can be divided between its static and dynamic components:

$$\begin{aligned} P_{tot} &= P_{stat} + P_{dyn} \\ &= V_{supply} I_D + \frac{C \cdot (V_{IMT} - V_{MIT})^2}{2t_{charge}} \end{aligned} \tag{3.16}$$

The expression for the dynamic power consumption is derived following a similar approach than for CMOS circuits, in which P_{dyn} corresponds to the power required to charge the internal circuit capacitances [243]. Indeed, the charge of a capacitance requires an additional current that is not fed back to the voltage source, and is rather dissipated through the ground. In our case, the charge of the output capacitance C takes place during the first half of the spike, upon a voltage step of amplitude $V_{IMT} - V_{MIT}$ (i.e. the spike amplitude), and for a duration t_{charge}. From there, one can define the energy associated to a single spike:

$$\begin{aligned} E_{spike} &= \int_0^{T_{spike}} P_{tot}\, dt \\ &= E_{stat} + E_{dyn} \\ &= V_{supply} I_D T_{spike} + \frac{C \cdot (V_{IMT} - V_{MIT})^2}{2} \end{aligned} \tag{3.17}$$

This development considers the NMOS as a current source supplying a constant current I_D, meaning it neglects its Early effect. This expression is valid for works with similar architectures [91, 93, 169]. Other architectures will have power consumption and spike energy expressions depending on $V_{out}(t)$ and $\int_0^{T_{spike}} V_{out}(t)dt$, which is why we are not able to estimate them as they require knowledge of these expressions. We however mention them when they are

estimated by the authors. For both of our oscillators as well as those with a similar architecture, the static contribution of the spike energy dominates the dynamic contribution (see **criterion 13** in Table 3.4), such that $E_{spike} \cong E_{stat}$.

Our second oscillator has a static power consumption P_{stat} similar to other oscillators with similar architecture (**criterion 11**) [91,93,169], but lower spike energy (**criterion 13**) thanks to its higher frequency (**criterion 12**). Its spike energy is also close to the ~ 3 nJ/spike state of the art [92,96]. The spike energy is dominated by the neuron static consumption, constrained by the device threshold voltage V_{IMT} (**criterion 7**) and threshold current I_{IMT} (**criterion 8**), which dictate the operation current (**criterion 9**) and supply voltage (**criterion 10**). To downscale these parameters, reduction in channel length [96] or shift toward a vertical architecture (**criterion 5**) [92] are promising tracks.

Cycle-to-cycle variations and LoD

Work on VO_2 oscillator increasingly report on the cycle-to-cycle variations of the device threshold voltage V_{IMT} (**criterion 15**). Our oscillator performs better than most implementations, which we attribute to its small grain size, as explored in Chapter 2. However, very few works analyze the impact of these variations on the frequency jitter $\overline{\sigma_f}(t_{obs})$ estimated on a certain observation duration, and the associated sensor LoD (**criteria 16 and 17**), making it hard to benchmark this dimension. One of them [96] however proposes to define the sensor SNR[2] as:

$$SNR = 10log_{10}\frac{\mu_f^2}{\sigma_f^2} \tag{3.18}$$

for which our sensor shows similar to slightly better performances, in coherence with the fact that it shows V_{IMT} cycle-to-cycle variations of a similar order of magnitude.

[2]The authors of [96] do not specify how they extract the σ_f parameter, which likely differs from the way we estimate $\overline{\sigma_f}(t_{obs})$, the uncertainty related to the frequency averaging defined in Section 3.2.1. For this comparison, we thus simply extract the standard deviation $\sigma_f(t_{obs})$ of the spike rate on the full, single waveforms we measured for both of our oscillators.

Table 3.3 Benchmarking of our VO_2 spiking temperature sensor against the SoA of VO_2 spiking sensors (Part. I)

Ref.	(1) Stimulus	(2) Sensing	(3) Number of devices	(4) Neuron area	(5) Device topology	(6) C	(7) V_{IMT}	(8) I_{IMT}	(9) I	(10) V_{supply}
Rosca 2021 [169]	Optical (450 - 800 nm, 0 - 70 mW)	VO_2 device	3	~ µm² range	Planar	150 nF	~ 1.8 V*	~ 700 µA	/	5 V
Rosca 2021 [91]	RF (60 - 110 GHz, 0 to -30dBm), Optical (500 - 700 nm, 0 - 70 mW)	VO_2 device	3	20 µm²	Planar	150 nF	~ 1 V*	~ 900 µA*	~ 1 mA*	5 V
Qaderi 2023 [93]	RF (130 - 160 GHz, 0 to -30 dBm)	External RF ports and VO_2 device	5	25 µm²	Planar	150 nF	~ 1.3 V*	~ 900 µA*	5 mA*	2.5 V
Duan 2022 [181]	Temperature (11 - 33 °C), Pressure (100 - 250g)	VO_2 device (temperature), External sensor (pressure)	2 (temp.) 3 (pressure)	0.4 µm²*	Planar	in-trinsic	~ 1.4 V	~ 500 µA*	~ 800 µA* (temp.) ~ 700 µA* (pressure, 100g)	5 V
Fang 2022 [95]	Pressure (3.8 - 10.6 kPa)	External sensor	3	314 µm²	Vertical	10 nF	~ 1.3 V*	~ 10 µA*	~60 µA* (pressure, 5.1 kPa)	4 V
Han 2022 [92]	Temperature (20 - 40 °C)	VO_2 device	1	7.9 µm²	Vertical	in -trinsic, 5.5 nF	0.66 V (20 °C)	~20 µA*	110 µA	/
Yuan 2022 [96]	Pressure (0 - 0.98 N), Optical (broadband, 0 - 1275 Lux), Temperature (30 - 100 °C)	External sensors	5	0.4 µm²	Planar	2 nF	1.33 V	~ 600 µA*	/	5 V
This work	Temperature (41 - 47 °C)	VO_2 device	3	2.4 µm²	Planar	33 nF	0.74 V (41 °C)	860 µA (41 °C)	1 mA	2.55 V
This work	Temperature (25 - 45 °C)	VO_2 device	3	2.4 µm²	Planar	330 pF	1.35 V (25 °C)	900 µA (25 °C)	1 mA	5 V

**Not explicitly mentioned by the authors, but estimated from graphs or derivated from other given numerical values.*

Table 3.4 Benchmarking of our VO_2 spiking temperature sensor against the SoA of VO_2 spiking sensors (Part. II)

Ref.	(11) P_{stat}	(12) Spiking frequency	(13) E_{spike}		(14) Sensitivity	(15) $\sigma_{V_{IMT}}$	(16) $\overline{\sigma_f}(t_{obs})$	(17) LoD	(18) SNR
			E_{stat}	E_{dyn}					
Rosca 2021 [169]	mW range	5.5 - 5.7 kHz	µJ/spike range*†	~ 92 nJ/spike*†	1 - 4.5 Hz/mW	/	/	/	/
Rosca 2021 [91]	~ 5 mW*	5.5 - 5.7 kHz (light) 3.3 - 3.5 kHz (RF)	~ 877 nJ/spike*†	~ 42 nJ/spike*†	4.64 Hz/dBm (RF) 4.23 Hz/mW (optical, 500 nm)	/	/	/	/
Qaderi 2023 [93]	~12.5 mW*	3.2 - 3.5 kHz	~ 3.6 µJ/spike*†	~ 17 nJ/spike*†	66.3 MHz/W	2.77 mV (~ 0.22 %*)	1.3 Hz at 200 ms	58 nW at 200 ms	/
Duan 2022 [181]	/	300 - 800 kHz (temperature) 450 - 950 kHz (pressure)	/		22.7 kHz/°C* (temperature) 3.4 kHz/g* (pressure)	/	/	/	/
Fang 2022 [95]	/	9.6 - 28.4 kHz	/		1.9 kHz/kPa*	~15 mV* (~1.12 %*)	/	/	/
Han 2022 [92]	/	16.8 - 60.2 kHz	3 nJ/spike†		1.7 kHz/°C	6.7 mV* (~1.08 %*)	/	/	/
Yuan 2022 [96]	/	104.8 - 253.3 kHz (pressure) 111.2 - 282.7 kHz (light) 100 - 300 kHz (temperature)	2.9 nJ/spike†		47.7 kHz/kPa (pressure) 0.1 kHz/Lux (light) 2.8 kHz/°C (temperature)	9.8 mV (0.73 %)	/	/	33.66 dB (pressure) 31.90 dB (light) 29.92 dB (temperature)
This work	2.55 mW	17.3 - 29.3 kHz	87 nJ/spike†	2 nJ/spike†	1.7 kHz/°C	1.95 mV (0.26 %)	118 Hz at 1 ms, 43 °C	0.42 °C at 1 ms, 43 °C	33.30 dB over 100 ms, 43 °C
This work	5 mW	437.6 - 974.8 kHz	5 nJ/spike†	135 pJ/spike†	14.3 kHz/°C (25 °C) 37.96 kHz/°C (40 °C)	/	496 Hz at 1 ms, 40 °C	0.08 °C at 1 ms, 40 °C	44.42 dB over 10 ms, 40 °C

**Not explicitly mentioned by the authors, but estimated from graphs or derivated from other given numerical values.*
†Computed using the maximum measured oscillation frequency.

3.3 Stochastic bursting temperature encoding

In all of the VO_2 spiking sensors presented in this benchmarking (including ours), the stimulus information is encoded solely through rate-coding (namely, the class-1 excitability among the 20 most prominent feature of biological neurons [52]). The neurons are either completely OFF in the absence of a stimulus of sufficiently high power; or, above a certain threshold, they spike constantly, with stimulus information being encoded into the duration of a given spike. This behavior is not the most bio-plausible, since biological neurons typically undergo more or less long quiescent periods in between spiking events. In addition, in a scenario in which these sensitive neurons would be interfaced with a Spiking Neural Network (SNN), such constantly-spiking behavior would impede the hardware energy efficiency, as the later relies on the sparsity of the input neurons encoding [244]. In this section, we propose to use the VO_2 spiking sensor presented above in a a novel operating regime, which exploits a phenomenon akin to stochastic resonance to produce sparse bursts of spikes whose temporal characteristics are stimulus-dependent.

3.3.1 Conditions for bursting

We start by detailing the conditions to reach such bursting regime. The circuit architecture is identical to the one used for the rate-coding sensor presented in the previous sections (see Fig. 3.14b). Measurements are performed with a VO_2 device of dimensions L = 600 nm, W = 3 µm. The new behavior is obtained by setting the transistor current I_D at the limit of the NDR region (right above the current level corresponding to the insulator-to-metal transition IMT, see Fig. 3.14a); in consequence, oscillations become unstable. The device thus emits bursts of spikes, visible in five 10 ms-long measured time series in Fig. 3.14c. Such behavior has already been demonstrated in VO_2 computational neurons by biasing them below their oscillation regime, and artificially adding noise to the input current. It has been associated to a behavior called phase-locked firing, skipping, or arrhythmic firing that is displayed by certain biological neurons [85, 245, 246]. Mammalian cold receptors [247] will for example show an irregular firing pattern, "skipping" a number of oscillation cycles. Typically, this inhibition will be periodic and tuned to the frequency of a subthreshold periodic signal.

The phenomenon that we exploit is akin to *stochastic resonance*, in which subthreshold signals are detected through the added noise, as the latter helps the signal crossing the detection threshold. Stochastic resonance has already been exploited in VO_2 neurons [206] to detect periodic voltage signals of weak amplitude, fed to the gate of the transistor. In our case, bursts occur when the noise amplitude allows the

drain current to briefly enter the stable oscillation region of the device. Such noise could come from the circuit itself (noise on the transistor current, thermal fluctuation in the VO_2 resistance; or stochastic charge injection events within the VO_2 channel) or from external interference (electromagnetic interference, power supply noise).

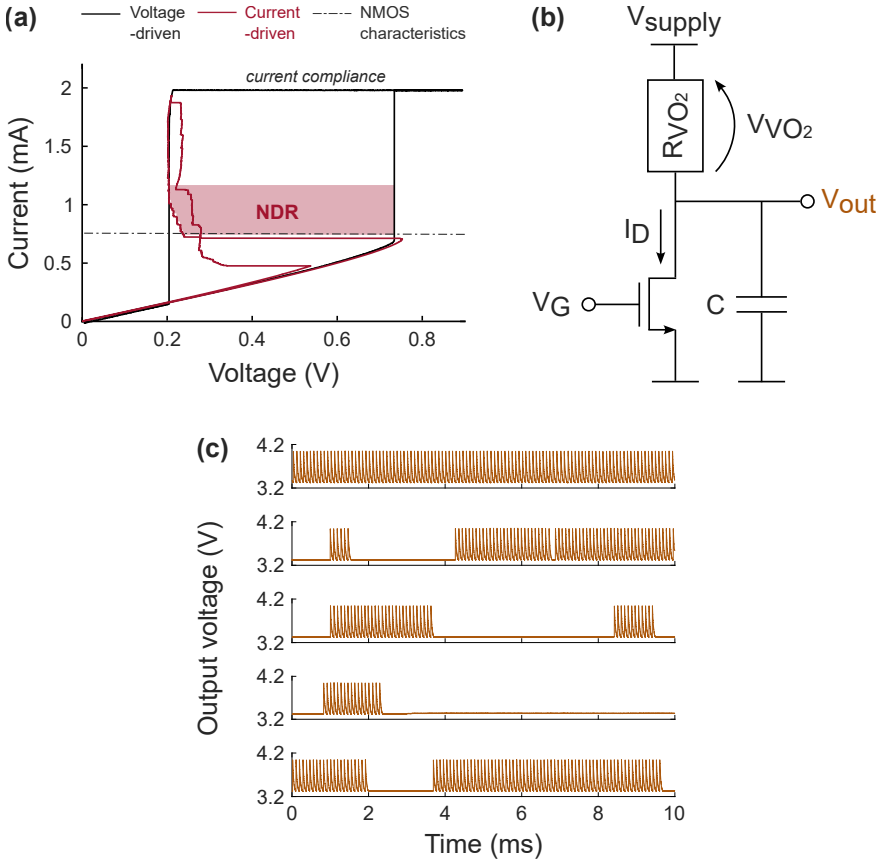


Fig. 3.14 **(a)** I-V characteristic of the specific VO_2 device studied in this work (L = 600 nm and W = 3 µm), measured in voltage-driven and current-driven modes at T = 42 °C. The NDR region is indicated in red. The transistor load line at V_G = 2.040 V is plotted against the VO_2 characteristics. **(b)** Circuit interface allowing to tune the current through the VO_2 device to bring it into bursting regime. **(c)** Five 10 ms-long samples showing the bursting output of the circuit, measured at 12.5 MHz, T = 42 °C, C = 33 nF, V_G = 2.040 V and V_{supply} = 4.5 V.

3.3.2 Temperature-dependent bursting

As for the rate-coding sensor, it is the temperature-dependence of the VO_2 device I-V curve which allows to encode temperature into the bursts temporal characteristics. Such dependence is shown experimentally in Fig. 3.15a, measured in a current-controlled way in the 41 to 47 °C temperature range.

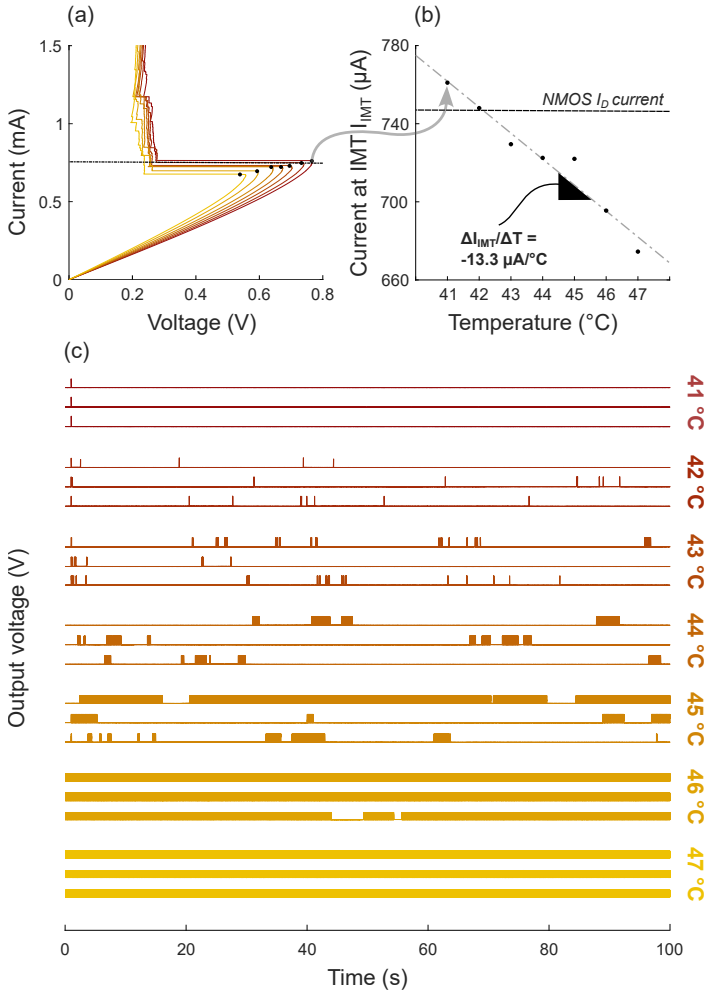


Fig. 3.15 **(a)** Evolution of the VO_2 device current-driven I-V characteristics with temperature, in the 41 °C (red) - 47 °C (yellow) temperature range. Only the forward-part the hysteresis is shown. **(b)** Extraction of the current level I_{IMT} corresponding to the insulator-to-metal transition point, plotted against temperature. **(c)** For each temperature point, three 100 s-long samples of the circuit output voltage were measured at 100 kHz.

We have seen in Section 3.2.2 that the temperature-dependence of the IMT bias point (V_{IMT},I_{IMT}) is dominated by a Joule heating mechanism. As VO_2 conductivity increases with temperature, the Joule power required to bring a portion of the VO_2 channel to the insulator-to-metal transition and trigger the filament formation reduces. This leads to a linear decrease of the corresponding insulator-to-metal transition current I_{IMT} with respect to the fixed NMOS current I_D (see Fig. 3.15b), at a rate of -13.3 µA/°C. The resulting bursting behavior of the circuit output has been measured through three 100s-long measurements at each temperature (Fig. 3.15c). At 41 °C, the set current is still below the current required to trigger the IMT, and bursts occur in a very sparse fashion. For the device to show stable and sustained oscillations, the set current must be more than tens of µA higher than I_{IMT}, as it is the case at 47 °C thanks to the evolution of the device I-V curve. In the intermediate temperature range, bursts duration tends to increase with temperature.

Past observations of bursts in resistive-switching neurons were induced by varying the amplitude of the current fed through the device [248, 249], or by artificially adding a certain amount of noise to this current [85, 206, 250]. In this work, the control current is fixed, and it is the variation of the VO_2 device I-V curve with temperature that progressively brings the system towards fully spiking regime, resulting in the observed stimuli-dependent bursting behavior. To the best of our knowledge, it is the first time that such behavior is observed in a volatile memristor.

3.3.3 Burst metrics definition

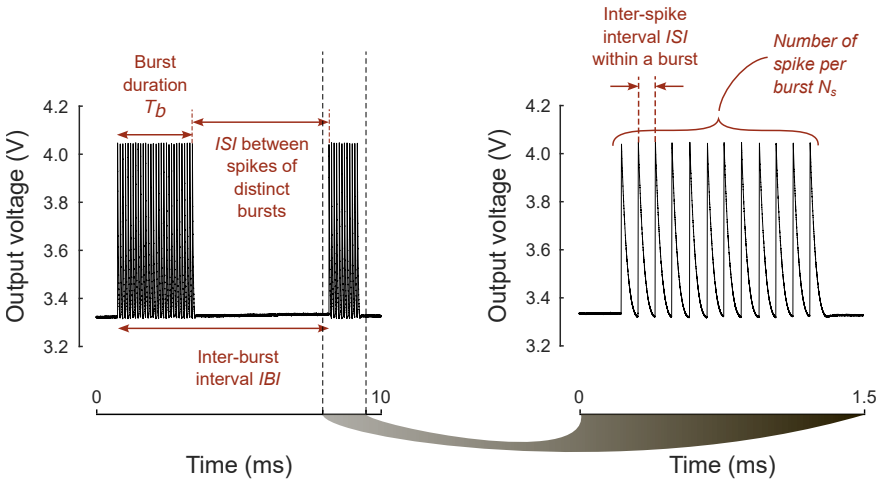


Fig. 3.16 Definition of the different burst metrics.

To study this phenomenon, we focus on the 42 - 45 °C temperature range where bursts are not too sparse nor too frequent. The goal is to understand which specific burst temporal metrics encode the information. The metrics of interest are defined in Fig. 3.16. They include the burst duration T_b and the number of spikes per burst N_s. The inter-burst duration IBI corresponds to the elapsed times between the start of two bursts. The firing rate is defined as the total number of spikes fired over the three 100-s long measured samples. Finally, we define ISI as the inter-spike interval. We differentiate the ISI of spikes belonging to different bursts, and the ISI of spikes belonging to the same burst. This raises an important methodological question: when should two consecutive spikes be considered as belonging to the same burst *vs* distinct bursts?

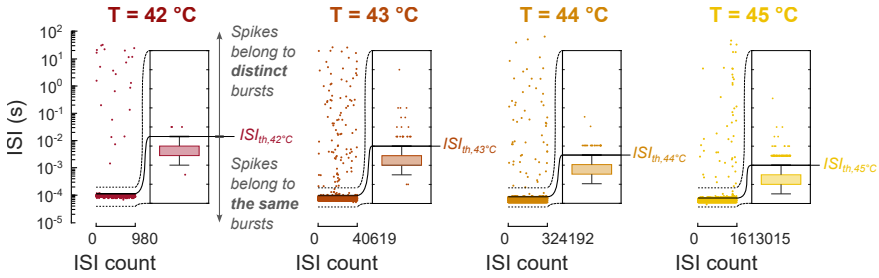


Fig. 3.17 Statistical definition of bursts by analysis of inter-spike intervals distributions at each temperature. For each temperature, the dataset corresponding to the ISIs computed for each pair of consecutive spikes is shown on the left. On the right, the box-plot of the dataset is displayed. The boxes are centered around the dataset median, and the box lower and upper boundaries correspond respectively to the first and third quartiles, $Q1$ and $Q3$. The errorbars around the boxes correspond to the outlier threshold: the upper boundary corresponds to $Q3 + 1.5 \cdot IQR$, which we set as the ISI threshold for burst definition.

To answer this question, we define a threshold ISI ISI_{th} using the Interquartile Range (IQR) method, as shown in Fig. 3.17. The IQR, or "box-plot" method, is commonly used to detect outliers within a dataset [251]. Our method relies on the assumption that in a given ISI dataset, ISIs of consecutive spikes belonging to distinct bursts are less frequent than ISIs of consecutive spikes belonging to the same burst. Quartiles are statistical metrics which divide a dataset into four intervals, such that each interval holds one fourth of the dataset samples. For each temperature, the ISIs between each pair of consecutive spikes are extracted over the three 100-s long measurements, and the three quartiles $Q1$, $Q2$ and $Q3$ dividing the ISIs dataset into these four intervals are computed. The threshold is defined as such; two consecutive spikes whose ISI is larger than:

$$ISI_{th}(T) = Q3 + 1.5 \cdot IQR \tag{3.19}$$

are considered to belong to distinct bursts. This specific threshold ISI_{th} will differ for each temperature.

3.3.4 Burst metrics statistics

Now that we defined the various burst metrics, we would like to know how they evolve with temperature. However, since the observed bursting behavior is highly stochastic, it is relevant to first study the distribution of these parameters. Fig. 3.18 and 3.19 show that both the burst duration T_b and the number of spikes per burst N_s follow an exponential distribution.

The probability density function $f_X(x)$ of an exponential distribution is represented by a single parameter λ:

$$f_X(x) = \frac{1}{\lambda} e^{\frac{-x}{\lambda}} \quad , \tag{3.20}$$

with λ being equal both to the distribution mean and standard deviation. Such bursting behavior has been observed in coupled oscillators, in which similar exponential distributions for the bursts duration were experimentally observed and modelled. This type of distribution is attributed to the "probability for a stochastic trajectory to temporarily escape from the local basin of attraction of the equilibrium, or state of rest" [252]. Modelling the bursting behavior in our spiking VO_2 neuron would require a transient model for the different noise sources.

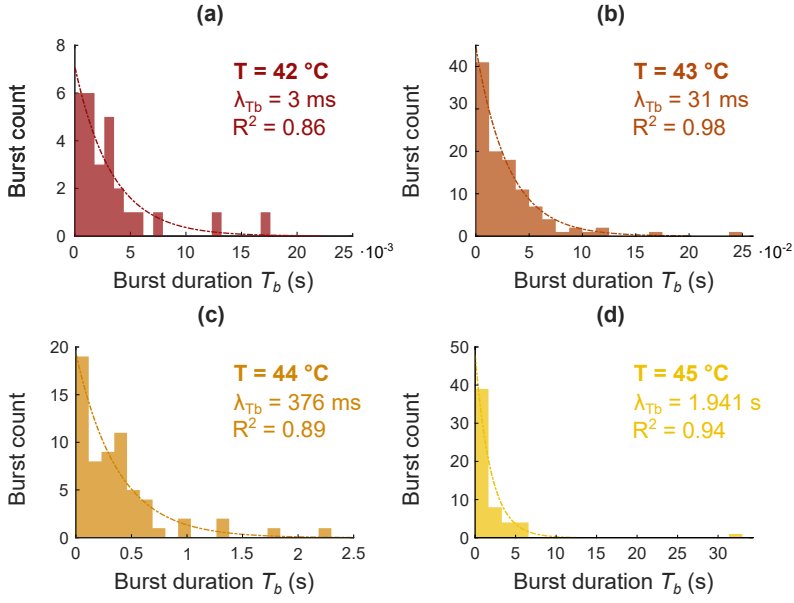


Fig. 3.18 Histograms of the burst duration at **(a)** 42 °C, **(a)** 43 °C, **(a)** 44 °C and **(a)** 45 °C, fitted to exponential distributions (dashed lines). The mean and standard deviation λ_{T_b} and the goodness of fit, corresponding to the R^2 coefficient of determination of the fit, are indicated for each temperature.

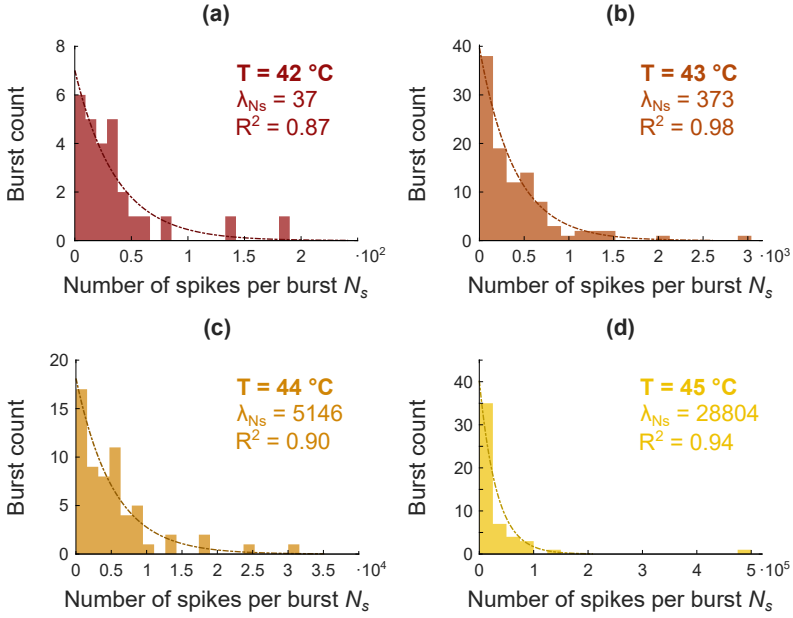


Fig. 3.19 Histograms of the number of spikes per burst at **(a)** 42 °C, **(a)** 43 °C, **(a)** 44 °C and **(a)** 45 °C, fitted to exponential distributions (dashed lines). The mean and standard deviation λ_{N_s} and the goodness of fit, corresponding to the R^2 coefficient of determination of the fit, are indicated for each temperature.

Regarding the distribution of the inter-spike intervals ISI within the bursts, they follow a Gaussian law (although not shown here), in coherence to observations of our rate-coding oscillator operating in the stable spiking region (see Fig. 3.7).

Finally, the distribution of the inter-burst intervals IBI can be studied. As shown in Fig. 3.20, it can also be fitted with an exponential distribution. To quantify the goodness of fit of approximating this distribution by an exponential law, we use the R^2 value, or so-called *coefficient of determination*. It is defined as:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad , \tag{3.21}$$

where SS_{res} is the residual sum of squares, and SS_{tot} is the total sum of squares. R^2 is bounded between 0 and 1: the closer it is to one, the better the goodness of the fit.

Studying the distribution of the IBI is especially relevant, as it can give us information on the independence of burst triggering events. If the burst occurrences were truly independent from each other (that is, if they could be modelled as a

Poisson process), the distribution of their inter-burst intervals should follow a perfect exponential law. However, the fitting shows poor goodness of fit (lower R^2 values than for the exponential fitting of metrics T_b and N_s, see Fig. 3.18 and 3.19), with a higher number of short inter-burst intervals that is not well captured by the model. This hints that the burst occurrences are not fully independent. Indeed, mechanisms such as persistent metallic islands at the location of the previously formed metallic filament after a burst, explored in Chapter 2 (see Section 2.3.4), could facilitate the occurrence of a subsequent burst, explaining this apparent lack of independence.

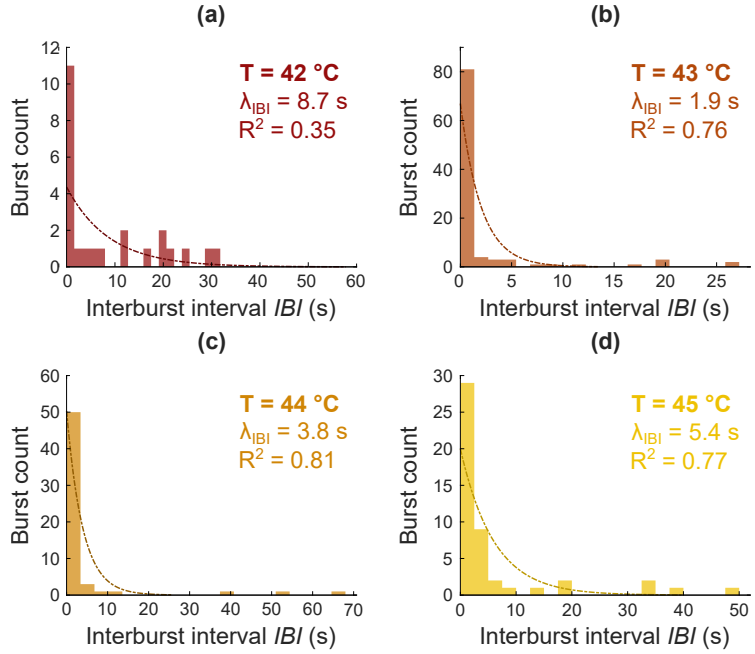


Fig. 3.20 Histograms of the inter-burst intervals at **(a)** 42 °C, **(a)** 43 °C, **(a)** 44 °C and **(a)** 45 °C, fitted to exponential distributions (dashed lines). The mean and standard deviation λ_{IBI} and the goodness of fit, corresponding to the R^2 coefficient of determination of the fit, are indicated for each temperature.

3.3.5 Burst metrics temperature dependence

Now that we have studied the statistics of the different bursts metrics, we can observe how they vary with temperature, to see how the stimulus information is actually encoded within the bursts. Fig. 3.21a-b show that the mean λ_{T_b} of the burst duration and the mean λ_{N_s} of the number of spikes per burst, both increase exponentially with temperature. The standard deviations are not shown on the plot, as

they are equal to the distributions means λ_{T_b} and λ_{N_s}. Additionally, the mean ISI within a burst decreases linearly with temperature, in coherence with the behavior observed in narrow temperature ranges in our rate-coding temperature sensor[3] (see Section 3.2). We thus show that stimulus information is encoded within the fine structure of the bursts. Finally, the firing rate increases exponentially with temperature. The inter-burst duration IBI, not shown here, doesn't show any trend in correlation to temperature. Interestingly, the evolution of the burst duration, number of spikes per burst, ISI within a burst and firing rate are similar to those observed in biological sensory neurons of certain insects in response to heat [253, 254]. The occurrence of bursts in biological neurons has been a subject of interest in the field of neuroscience: evidence show that bursts could improve the reliability of information transmission across unreliable synapses, and enhance the transmission of sensory information by improving the signal-to-noise ratio [255]. It often takes place in situation of danger, such as in noxious heat ranges [254] or in prevision of a looming collision with an object [256], both observed in insect sensitive neurons.

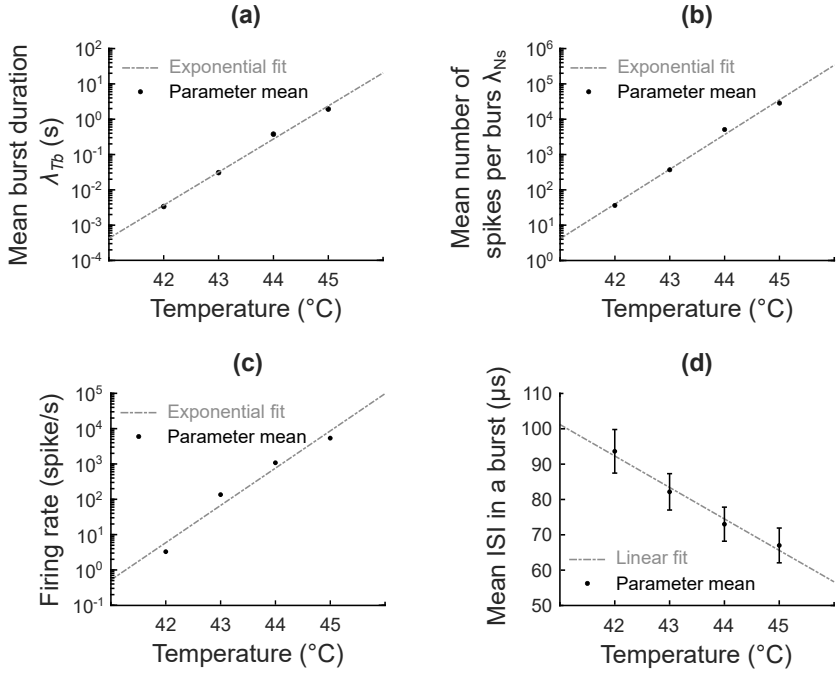


Fig. 3.21 Evolution of several burst metrics with temperature : mean burst duration **(a)**, mean number of spikes per burst **(b)**, firing rate (extracted over 300 s-long measurements), **(c)** and mean ISI in a burst **(d)**.

[3]Indeed, there is a direct correspondance between the ISI within a burst and the spiking frequency, with $ISI = 1/f_{spike}$

3.4 Key messages

We sum up here the important findings covered in this chapter.

Small-dimensions VO_2 devices are best suited to design sensitive spiking neurons. Indeed, to design VO_2 relaxation oscillators, the device needs to display a negative differential resistance, which is more prominent in device of smaller length and width.

The dominant source of the spike rate jitter are the cycle-to-cycle variations in the electrical properties of the VO_2 two-terminal devices. More specifically, the contributions of the parameters related to the device insulating phase, i.e. its threshold voltage V_{IMT} and its insulating resistance $R_{ins,0}$, dominate.

- To draw this conclusion, we used a dynamic analytical model predicting the output of the relaxation oscillator. To this model, we included the cycle-to-cycle variations in the VO_2 electrical parameters (extracted through numerous DC measurements), along with other noise sources such as the thermal noise on the VO_2 resistor and the NMOS transistor, and the empirically extracted electrical noise from the setup.
- The fact that the spike rate jitter can be predicted from these DC-extracted cycle-to-cycle variations, shows that these variations seem to have a similar order of magnitude in DC and AC (in the $\sim 10^4 Hz$ range).

The encoding of the temperature stimulus into the oscillator spike rate can be performed directly thanks to the VO_2 device, without any external sensor, thanks to the temperature-dependence of the VO_2 I-V curve.

- In the case of our specific VO_2 device, the dominating mechanism seems to be Joule heating over field-induced carrier injection.
- In narrow temperature ranges, the frequency depency to temperature can be approximated as linear, while it is rather exponential on larger temperature ranges.

Averaging the cycle-to-cycle variations across several cycles allows to reduce the uncertainty on the spike rate, inducing an energy-resolution trade-off. Our observations show that a 100-fold increase in observation duration is necessary to improve the LoD by one order of magnitude.

- Using the Allan deviation statistical tool, we can deduce that white noise sources dominate up to 10 ms timescales, confirming that averaging the spike rate on time intervals of corresponding duration will reduce uncertainty.

- Increasing the oscillator frequency by reducing the output capacitor relaxes the energy uncertainty trade-off. Indeed, it allows to screen a larger number of oscillation cycles in a shorter observation time. The trade-off relaxation is also due to the fact that the cycle-to-cycle variations in spike duration remain in the same order of magnitude at higher frequencies ($\sim 10^5 Hz$).

In terms of benchmarking, our rate-coded implementation performs well but doesn't beat the state of the art (SoA) regarding compactness, footprint and energy consumption per spike. However, quantifying the true energy efficiency of the sensor requires to analyze the spike rate jitter for a certain averaging time, and the associated LoD. Very few works report on these metrics, despite them being a crucial parameters of spike rate sensor. We call for systematically including these kind of analysis.

For the first time, we demonstrate a spiking sensor where information is encoded into the timing of self-induced bursts, rather than by plain rate coding of individual spikes. We attribute this behavior to punctual noise events allowing the system to escape its rest equilibrium.

- Similarly to the response of insect sensory neurons to noxious heat ranges, the temperature stimulus is encoded into the burst duration, number of spikes per burst, ISI within a burst, and firing rate.
- The dead times between burst are a self-induced form of sparsity. For now, this sparsity does not reduce the energy consumption of the spiking sensor itself, as the latter is dominated by the static consumption. This sparsity would become an advantage if miniaturization allowed to bring the static consumption close to or below the dynamic consumption. In addition, if encoding information into the ISI metric, the bursts should be long enough - or the spike rate jitter low enough - to provide a satisfying LoD. In all cases, such sparsity would reduce the overall system energy consumption if the sensor is to be integrated to a Spiking Neural Network, by sparsely triggering the latter.

4

Assessing the environmental impact of VO_2 neuromorphic hardware

The higher the level of a technology,
the smaller the advantages it can bring
with respect to its drawbacks.
– Simone Weil

SPIKING sensors, and neuromorphic devices in general, are often put forward as a possible solution to curb the energy consumption and associated carbon emissions of large AI models, with some claiming they could help us achieve "Green AI" [12, 257]. Beyond the fact that the energy efficiency advantage of these platforms over conventional hardware remains to be proven (as discussed in Section 1.1.2), the direct environmental impacts of their production are usually not considered when making these claims. As introduced in Chapter 1, neuromorphic spiking neurons can be implemented either in CMOS technology or using memristors based on emerging materials, such as VO_2. While several works have been conducted on the environmental footprint associated to the fabrication process of CMOS technology nodes [258–262], there has so far been no research on the environmental cost of fabricating emerging neuromorphic devices. Performing such analysis seems especially relevant in regards of the expected production volumes, if such technologies were to be industrialized. In a scenario in which emerging memristive devices would be used as hardware primitives in new generations of processors, such systems could contain up to millions of neurons [263]. If they hold their promise in terms of energy efficiency, the number of such processors could increase as well, as the amount of analog data to be processed follows an exponential growth [264]. Alternatively, if memristive devices were used as spiking sensors like

proposed in this work, deployment scenarios related to IoT applications would also imply large volumes of sensors, as the sector also follows steep growth trend [2,3]. In addition, it is recognized that performing impact assessment analysis on technologies with low Technology Readiness Levels (TRLs) allows to reorient design choices to minimize the product environmental footprint prior its market penetration, at which point changes in the process flow would come at higher costs [265]. For these different reasons, this chapter performs a cradle-to-gate Life Cycle Assessment (LCA) of the fabrication of the VO_2 two-terminal devices presented in this work, and attempts to answer the following questions:

What are the direct environmental impacts associated to the fabrication of VO_2 memristive devices? Can they be reduced through eco-design, and how do they compare to those of CMOS spiking neurons?

To address these matters, this Chapter is structured as such. First, Life Cycle Assessments and their guiding principles are introduced in Section 4.1. Information and structure of this short introduction is inspired from the reference book *Life Cycle Impact Assessment* of M. Hauschild [266], the content of the ISO norm ISO14040 [267], and the excellent work of Pirson et al. [261]. We then dive into the specific LCA of our VO_2 spiking neurons. The specific goal and scope of our LCA, and the methodology followed to perform the inventory and impact assessment are detailed in Section 4.2. We quantify the impacts on three indicators: the Cumulated Energy Demand (CED), the Global Warming Potential (GWP) and the Ultra Pure Water consumption (UPW) metrics. These impact categories are those most commonly found in LCAs of integrated circuits. The life-cycle impact assessment results are presented in Section 4.3, and allow to identify the process steps that significantly contribute to the different impact categories. The discussion touches on several questions. The robustness of our conclusions w.r.t. variations of certain parameters is analyzed in Section 4.4.1. An alternative design route to reduce the total GWP and CED of the process is discussed in Section 4.4.2. Finally, in Section 4.4.3, the VO_2 memristor direct environmental footprint and performances are compared to CMOS-based implementations. The results presented in this work have been adapted from our *Electronics Goes Green 2024* conference paper [CP2].

4.1 Basics of LCA

Life Cycle Assessment (LCA) is a tool defined by international standards. It allows to estimate the *direct*[1] environmental impacts of a product or service throughout its whole life cycle, including the extraction of raw materials needed to fabricate it, its manufacturing process, use phase, and end-of-life. It follows a systemic approach in the sense that it quantifies the product impact on several environmental indicators: beyond the sole impact on climate change, metrics such as resource depletion, water consumption, soils acidification, ecotoxicity, and many more can be considered. The specific metrics depend on the selected impact assessment methods, which are numerous. Independently of the chosen method, this multi-criteria analysis allows LCA practitioners to gain a global view and avoid problems of burden-shifting when attempting to minimize the footprint of a product or service.

4.1.1 The four stages of LCA

LCAs typically take place throughout four main stages, defined by the ISO standard ISO14040 [267] and illustrated in Fig. 4.1.

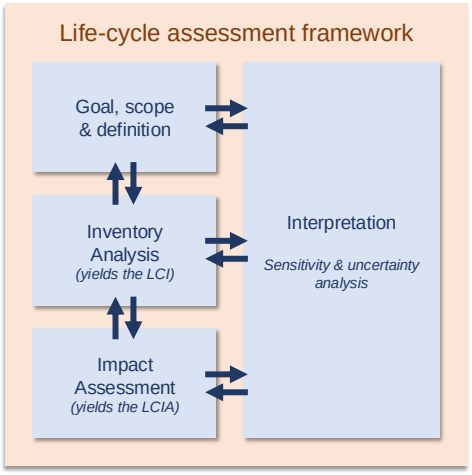


Fig. 4.1 The four main stages of an LCA, as defined by the ISO14040 standard [267]. Bi-directional arrows embody the iterative nature of LCAs. Adapted from [267] and [261].

[1] To be distinguished from *indirect* environmental impacts, which arise through the "change of production processes, products, and distribution systems [268]". Such impacts can be composed of negative contributions (associated to induction and obsolescence effects) or positive contributions (associated to substitution or optimization effects).

Goal and scope definition. The first step includes defining the *goal* of the LCA: is it performed in an eco-design approach, to minimize the product or service footprint? Is it rather intended to compare the product footprint to that of alternative options? Who will be the target audience of the LCA? Depending on the answers to these questions, different methodological choices will be made. Most of these choices are embodied within the *scope* definition, which includes several aspects. First, the *system boundaries*, i.e. the specific phases of the product life cycle that will be included in the analysis. An example of such system boundaries are so called "cradle-to-gate" analysis, in which the use and end-of-life stages are not considered. The four main categories of system boundaries are illustrated in Fig. 4.2. Secondly, the choice of the product or service *functional unit* for which the analysis is performed (e.g. the fabrication of 1 cm² of die area of CMOS logic produced in 5 nm technology node). The functional unit confers a relative nature to the LCA results.

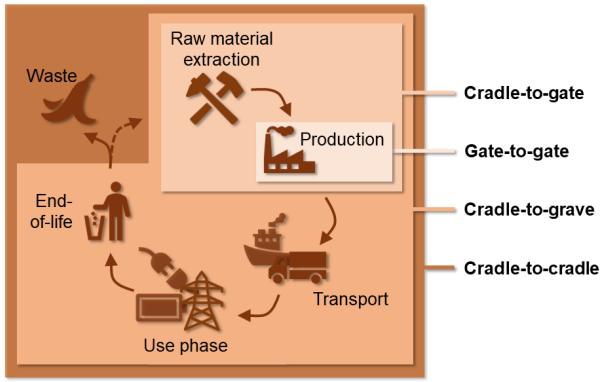


Fig. 4.2 The four main categories of system boundaries in LCAs. Adapted from [261].

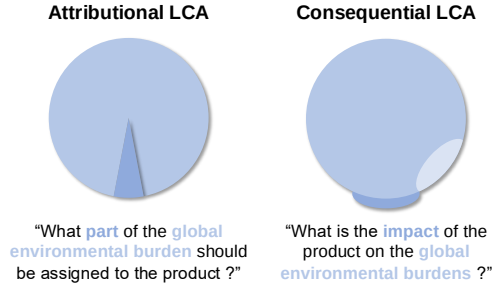


Fig. 4.3 Illustration of the two types of LCA system models. Adapted from [269].

The choice of *system model* is also made during this first stage, between *attributional* or *consequential* [269] (illustrated in Fig. 4.3). In the attributional model, one asks how much of the existing global environmental burdens should be assigned to a product. In the consequential model, one asks what is the difference in environmental burden resulting from the product being put on the market. The attributional practice is the most common. In addition, *impact categories* and associated *impact assessment methology* are also defined during this stage, and will be detailed below.

Inventory analysis. The inventory stage details the *elementary flows* attributed to the functional unit, summed up across temporal and spatial boundaries. In a cradle-to-gate study of our 1 cm² of CMOS die, these could include the masses of metals, gases and chemical products required for the fabrication; kWh of electricity and volumes of petrol used for materials extraction, transport and transformation within the fab, ... The output of the inventory analysis is the *Life Cycle Inventory* (LCI), which lists the quantities of all these elementary flows.

Impact assessment. During the impact assessment stage, one computes the contribution of LCI's elementary flows to a list of impact categories and associated indicators. Indicators can be divided between *midpoints* and *endpoints* indicators, with endpoint indicators aggregating several midpoints indicators. The specific indicators depend on the chosen *impact assessment method*. Popular methods include the *Environmental Footprint (EF)*, maintained by the European Commission, which defines 16 different midpoint impact categories and associated indicators. The *Cumulative Energy Demand (CED)* method, developed by *Ecoinvent*, assesses primary energy usage. The assessment can be performed in a *bottom-up* or *top-down* fashion, depending on whether impact data are gathered by summing the impact of each subsystem within the product, or by deducing them from macro-economic entities such as markets. Hybrid approaches can also be followed. The impact assessment stage is an iterative process, and may require reviewing the LCA goals, scope and inventory if they cannot be fulfilled with the available data. This impact assessment stage yields the *Life Cycle Impact Assessment (LCIA)*. Once each impact category result has been quantified, they can optionally be *normalized* with respect to a reference (a common normalization reference is the impact of an average human over one year on the different impact categories considered). A global score can also be defined by *weighting* the contributions of the different impact categories.

Interpretation. The interpretation stage attempts to answer the question associated to the LCA's goal, using the results of the LCIA step. Depending on the goal, it could be to determine a hotspot within the product life cycle, and attempt to find alternatives with smaller footprint. Or it could be to compare the LCIA of two products. During this step, one generally performs sensitivity, uncertainty and

scenario analysis to gain confidence on the results robustness. Indeed, uncertainty is omnipotent in LCAs, and can arise both when constructing the LCI (due to e.g. inventory errors such as imprecision in quantities, missing elementary flows, entire products sub-system left out of the analysis, ...) and during the LCIA (due to uncertainty when quantifying the impact attributed to the flows of the inventory).

4.1.2 LCA of AI and neuromorphic technologies

While researchers have started to quantify the direct and indirect environmental impacts of AI [270–272], research on the direct environmental impact of neuromorphic technologies is, to the best of our knowledge, non-existent. An obstacle to the realization of such studies is the early stage of the technology development, which makes it hard to evaluate the use phase and end-of-life of a neuromorphic system, or to project what its production would look like in industrial production lines. However, cradle-to-gate LCAs could be quite easily performed for CMOS-based neuromorphic platforms. Indeed, there is a number of work providing cradle-to-gate LCAs of CMOS technologies, both in literature [258, 259, 273, 274] and from commercial databases (*Sphera*, *EIME*, *ecoinvent*). Most of them focus on three impact categories (midpoints): the global warming potential (GWP), which quantifies all CO_2-equivalent emissions within the considered scope; the cumulated energy demand (CED), which quantifies the equivalent amount of primary energy required by the product within the considered scope; and the ultra pure water consumption (UPW) used within the cleanroom. These studies of course suffer from limitations [275], which we will detail below; however, they could provide first estimations for the direct environmental footprint of the fabrication of CMOS neuromorphic platforms.

For neuromorphic hardware primitives based on emerging memristive devices, one cannot rely on CMOS LCAs, as the process flows widely differ due to the use of new materials and the lower TRL. A cradle-to-gate LCA would thus require building a custom, detailed inventory with information granularity at the process step level. Fortunately, most of the research groups developing such devices do so in their own cleanrooms, easing the access to these informations. While these technologies are at an even lower development stage than CMOS-based neuromorphic hardware, performing a LCA can be relevant in an eco-design approach. Indeed, so-called *ex-ante LCAs*, which apply to emerging technologies which only function at lab or pilot-scale rather than full market existing systems, offer several benefits [265, 276]. These include greater inventory precision due to acute knowledge of fabrication processes; they have the potential to steer technological development towards sustainable innovation; they can prevent environmental burdens at low cost.

4.2 LCA methodology

We now dive into the specific Life Cycle Assessment of the VO_2 spiking neurons fabricated in this work.

4.2.1 Goal and scope

The first goal of this LCA is to identify the largest contributors among the different process steps in terms of the three most commonly evaluated impact categories for microelectronics components: the CED, GWP and UPW consumption. The objective is then to investigate alternatives to these process steps, to determine ecodesign guidelines and reduce the direct environmental footprint associated to the device fabrication. This analysis is aimed at researchers and designers in the neuromorphic field, eager to make design choices minimizing the direct footprint or their fabricated device at the upfront of its development. A secondary goal is to provide a rough comparison with the environmental impact of CMOS-based spiking neurons. In this comparison, we would only consider the spiking functionality of our neurons (leaving out their sensing capabilities), as very few implementations of CMOS-based sensitive neurons exist (see discussion in Chapter 1, Section 1.2.2). The functional limit chosen for this analysis corresponds to a 3-inch silicon wafer (45.6 cm² surface) on which the memristive devices are fabricated. Our model follows an attributional approach, as recommended in [266] given the context of our work. The LCA system boundaries are presented in Fig. 4.4, and correspond to a cradle-to-gate scope which can be divided into the *background* and *foreground* systems.

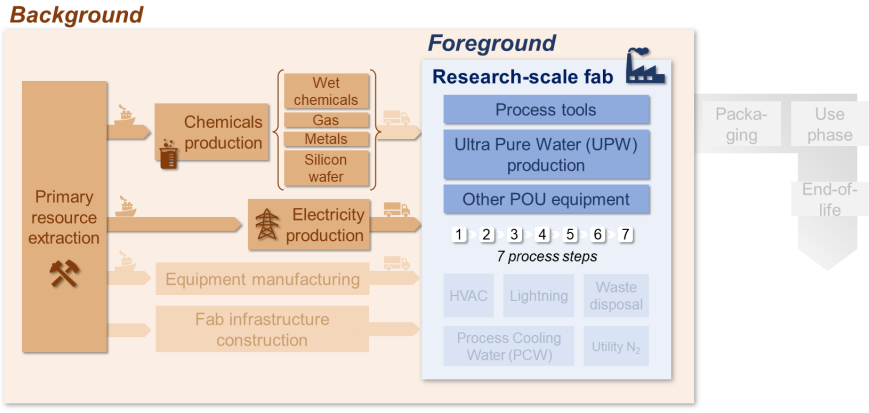


Fig. 4.4 Scope and system boundaries of the VO_2 spiking neurons LCA, with distinction between the background and foreground systems. Lighter items are not included in the scope.

The background system includes the manufacturing of upstream chemicals required for the process (wet chemicals, gas, metals), as well as the the silicon wafer production. This includes extraction of raw materials and transformation processes, but excludes transport of the input products to the fab gate, due to lack of accurate provenance data from suppliers[2]. The foreground system includes on-site processes taking place within the fab: more specifically, the electricity consumption of process tools, UPW production utilities and other point-of-use (POU) equipment[3], quantified for each of the seven process steps.

Within the fab, we exclude the elementary flows related to elements that are common to the whole fab facility, and shared by different process flows. This includes the equipment manufacturing, the construction of the fab infrastructures, and the electrical consumption of certain utilities (Heating – Ventilation and Air Conditioning or HVAC, lightning, waste disposal, utility N_2 within benches, and process cooling water). Indeed, unlike in industrial production lines, the VO_2 devices studied in this work are fabricated in a research-scale fab where a large number of different fabrication processes are conducted on the same equipments. This feature makes it difficult to allocate the shared environmental impacts in a meaningful way. In addition, the first goal pursued in this work being eco-design, such impacts are considered fixed as they currently cannot be optimized to reduce the device direct footprint. The packaging, use phase and end-of-life stages are also neglected: due to the low TRL of the device and associated uncertainty regarding its specific application, the data required to evaluate these stages of the life cycle are yet unknown.

4.2.2 Inventory

The fabrication process of the VO_2 memristor has been detailed in Chapter 2 (see Section 2.2.1). As a reminder, it is composed of seven steps: (1) standard cleaning of the silicon wafer, (2) thermal oxidation to create an 200 nm - thick electrically insulating SiO_2 layer, (3) DC reactive sputtering of a 135 nm - thick amorphous VO_2 thin film from a vanadium target, (4) thermal annealing to trigger its crystallization, (5) lithography, (6) evaporation of a 150 nm - thick gold layer (together with a thin 5 nm nickel adhesion layer) and (7) liftoff to define the electrical contacts. The inventory is completed in a bottom-up approach using data gathered directly within the fab, offering precision and a granularity at the process step level. Typical elementary flows gathered for a given process step are illustrated in Fig. 4.5.

[2]Within the ecoinvent database, this choice translates to using *production* activities instead of *market* activities.

[3]Point-of-use equipments are equipments which support a specific process tool, and are usually located in the "subfab" level (lower floor w.r.t. process tools). These include pumps, chiller, heat exchangers, ...

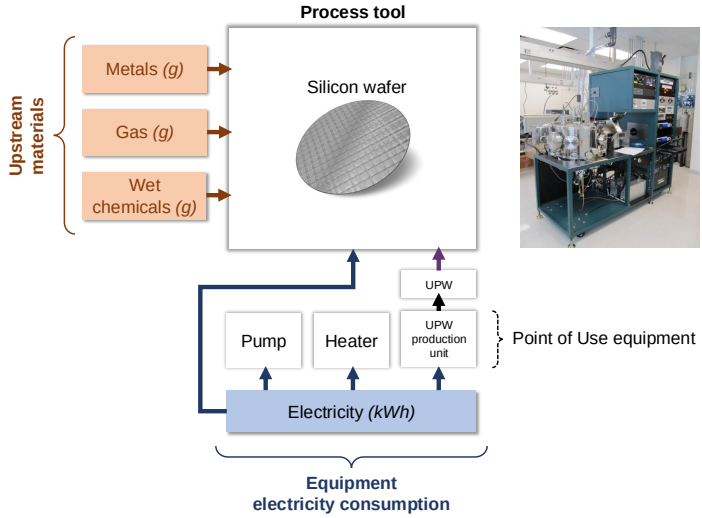


Fig. 4.5 Elementary flows of the inventory illustrated for a "typical", fictive process step. The picture shows the process tool used in the VO_2 sputtering step.

We can distinguish two key categories in the inventory of each step: input materials, and equipment electricity consumption, whose individual contribution will be highlighted below. Quantities of wet chemicals and UPW (used in the standard cleaning, photolithography and liftoff steps) are directly measured, while quantities of gas (used in the thermal oxidation, sputtering, and annealing steps) are extracted from the recipes detailed on the equipment control programs. The exact quantities of metals are either deduced through weighting metallic crucibles before and after the deposition (for gold and nickel), or by estimations based on sputtering chamber volume and film thickness (as for vanadium). Electrical consumption of process tools and POU equipments used in the different process steps are deduced from the rated power indicated on their datasheets, as well as their operation time (excluding stand-by operation). The inventory considers one run of each process step, for which each equipment has a different loading factor (i.e. number of wafers processed simultaneously), which is taken into account into the normalization towards the functional unit. The inventory is detailed for each process steps in Table 4.1, with load factor specified for each step.

Table 4.1 Detailed inventory

		(1) Wafer fabrication	(2) Standard cleaning	(3) Thermal oxidation	(4) VO2 sputtering	(5) VO2 annealing	(6) Lithography	(7) Gold evaporation	(8) Liftoff	
Equipment load	(# wafers)	1	25	25	1	30	10	4	2	
Gas										
Ar	(g/wafer)				2.18E+00	2.68E-01				†
H2	(g/wafer)			1.48E+00						†
O2	(g/wafer)			2.58E+01	2.47E-01					†
N2	(g/wafer)			3.18E+01						†
Wet chemicals										
H2SO4	(g/wafer)		1.27E+02							†
H2O2	(g/wafer)		1.19E+01							†
HF	(g/wafer)		7.96E-01							†
HMDS	(g/wafer)						9.68E-01			†
Acetone	(g/wafer)						1.58E+01		2.38E+01	†
Methanol	(g/wafer)						1.59E+01			†
Isopropanol	(g/wafer)								2.36E+01	†
DMSO	(g/wafer)								7.70E+02	†
PGMEA	(g/wafer)						1.62E+00			⋆
TMAH	(g/wafer)						2.02E-01			⋆
Metals										
V	(g/wafer)				1.87E-01					◇
Au	(g/wafer)							2.76E-01		◇
Ni	(g/wafer)							9.19E-03		◇
Others										
Silicon wafer	(m^2)	4.56E-03								†
Ultra Pure Water	(l/wafer)		8.00E+00				1.63E-01		6.43E+00	
Electricity										
Process tools & POU equipments	(kWh/wafer)		2.63E-02	2.59E+00	9.81E-01	2.43E-01	1.97E+00	3.03E+00	5.42E-02	○
UPW production	(kWh/wafer)		3.14E-01				6.38E-03		2.52E-01	○

Sources of CED and GWP data for Impact Assessment : † *ecoinvent v3.10,* ⋆ *Boyd [259],* ◇ *Nuss & Eckelmans [277],* ○ *PEF and carbon intensity values from IEA [278] [279]*

4.2.3 Impact assessment

For the majority of chemicals used in the process, the CED and GWP associated to their manufacturing are gathered from the *ecoinvent* database *v3.10* with cutoff system model, using respectively the Cumulated Energy Demand and EF v3.0 impact assessment methods. Exceptions include wet chemicals such as PGMEA and TMAH (found respectively in the composition of photoresist and developer used during lithography), for which CED and GWP values of 3.1 MJ/kg and 0.26 kgCO_2-Eq/kg were used based on the work of Kim & Overcash [280], following the approach followed in the pioneering work of Boyd on LCA of CMOS processes [259]. Since *ecoinvent* also lacks data for certain metals, we turn to Nuss & Eckelmans's work for CED and GWP values associated to the production of nickel, gold and vanadium [277]. It is noteworthy that these values are possibly underestimates of reality, as metals used in semiconductor fabrication processes such as those considered in this work are of very high purity, while Nuss & Eckelmans do not specify purity levels. All three data sources follow an attributional approach in term of system model, making them compatible on that specific matter. To compute the equivalent CED and GWP associated to the production of electricity consumed by process tools and POU equipment, we use respectively a Primary Energy Factor (PEF) for electricity[4] of 2.77 [278] and a carbon intensity for electricity of 145 gCO_2-Eq/kWh [279], two values estimated by the International Energy Agency (IEA) considering the Belgian electricity production mix. The UPW consumption is directly measured on site. The CED and GWP associated to the production of UPW only considers the electrical consumption of the UPW production unit, as the emissions and energy requirements related to the distribution of city water are negligible compared to it [259].

[4]The Primary Electricity Factor (PEF) for electricty describes the amount of primary energy required for the production of 1 kWh of electricity. It depends on the electricity production mix of the considered territory of production.

4.3 LCA results

The contribution of each individual process step to the total CED (expressed in MJ) and GWP (expressed in $kgCO_2$-Eq) associated to the fabrication of one wafer of VO_2 devices is detailed in Fig. 4.6. For both impact categories, gold evaporation stands out as the most intensive step. This is largely due to the extraction and manufacturing of gold, which represents the major part of this specific step footprint (65.51 % of its CED and 88.69 % of its GWP). The large energy and carbon intensity of gold production is generally attributed to its relatively low abundance in the earth crust [281], and to the large energy requirements and emissions of some of its refining steps, such as grinding and post-roasting gas treatments [282].

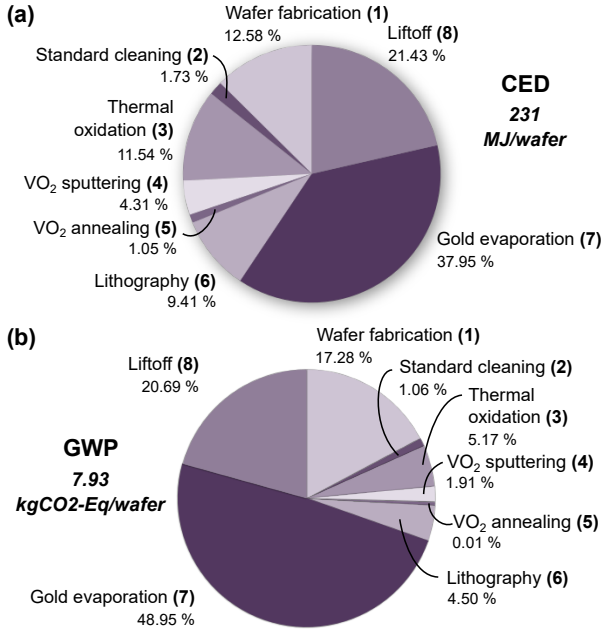


Fig. 4.6 LCA results regarding **(a)** the total CED and **(b)** GWP associated to the production of one wafer of memristive devices, with detailed contributions of individual process steps.

When looking at the total cradle-to-gate footprint, the CED and GWP impact categories are dominated by the manufacturing of the upstream materials. This trend is highlighted in Fig. 4.7, where the distinction is made between the contribution of the upstream materials manufacturing (orange tones) and the equipment electricity consumption (blue tones).

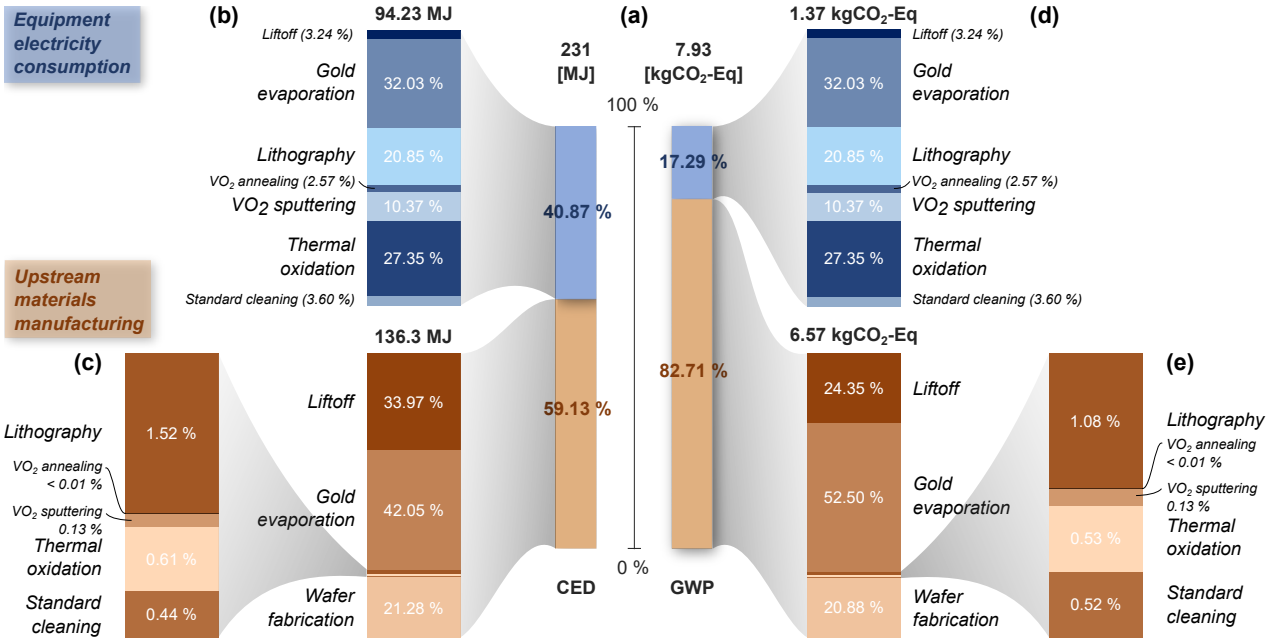


Fig. 4.7 **(a)** Decomposition of the total CED and GWP associated to the production of one wafer of VO_2 memristive devices, by making the distinction between the electrical consumption of equipment within the fab gates (blue) and the manufacturing of upstream materials (orange). In each case, contribution of individual process steps is detailed for CED **(b,c)** and GWP **(d,e)**.

The top three globally most intensive steps (Fig. 4.6) are also the top three contributors of the total CED/GWP associated to upstream materials manufacturing. Following gold evaporation, the second and third most intensive steps are the liftoff and wafer fabrication. The reasons for such high impact differ for each step. Liftoff requires dimethyl sulfoxide (DMSO, a solvent allowing to dissolve the photoresist). Its manufacturing is associated to quite standard CED/GWP values compared to other inputs of the process, but it is used in large quantities (see Table 4.1) to avoid gold particles attaching onto the samples. The fabrication of mono-crystalline silicon wafers typically involves low-yield, energy-intensive steps such as the transformation of trichlorosilane into polysilicon, the crystallization of polysilicon and the sawing of ingots into wafers [283]. Regarding the remaining steps, the dominant contribution to the CED/GWP are associated to the production of electricity required by the process tools and POU equipment. The thermal oxidation involves a high temperature furnace (1000 °C) with 17 kW rated power running for almost 4 hours. The lithography step requires a UV lamp consuming up to 2.6 kW, and the gold evaporation step requires pumping a large-volume chamber to low pressure during several hours, further adding to the contribution of this step.

Finally, the consumption of Ultra Pure Water (UPW) in the different process steps is shown in Fig. 4.8. Only three steps consume UPW: lithography, standard cleaning and liftoff. Standard cleaning and liftoff consume the largest amount of UPW, as large quantities of water are required to rinse remains of acids between cleaning baths during standard cleaning, and remove solvent, resist and gold particles during liftoff.

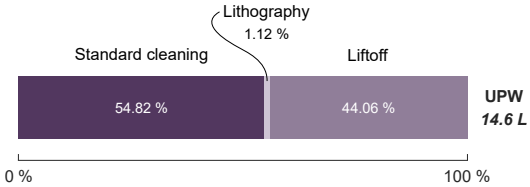


Fig. 4.8 Decomposition of UPW consumption among the different process steps, for the fabrication of one wafer of VO_2 devices.

4.4 Interpretation

4.4.1 Uncertainty & scenario analysis

To test the reliability of our conclusions, we discuss some of the main sources of uncertainty and vary the value of certain key parameters in different scenarios.

Uncertainty

In this section, we do not provide a full uncertainty or sensitivity analysis, but we discuss some important sources of uncertainty. We will discuss two main types of uncertainties put forward by Hauschild et al. [266]: first, the *parameter uncertainty*, which arises from the construction of the inventory (error on the exact quantities of the elementary flows and possible missing elements with respect to the system boundaries), and second, *model uncertainty* which relates to the impact assessment (error on the impact attributed to the element of the inventory on the different indicators).

Let us first address the parameter uncertainty related to the inventory. The construction of the inventory can be divided in two categories as introduced in Fig. 4.5: input materials, and equipment electricity consumption. As mentioned above, most of the input materials quantities were measured directly within the lab, offering great precision for this part of the inventory. However, a significant amount of uncertainty can be related to the method used to estimate the electricity consumption of the process steps equipment, which likely yields overestimation. Indeed, this consumption was extracted based on datasheet rated power values rather than actual measured instantaneous power consumption. To get insight on the impact of this hypothesis, we performed such measurements during the thermal oxidation step using a power-energy logger. We obtain a capacity factor (actual electricity consumption over rated consumption) of 0.23, confirming our hypothesis of overestimation. Applying such capacity factor uniformly to the electricity consumption of all equipment would reduce the total CED and GWP associated to wafer fabrication by respectively 29.87 % and 12.48 %, and reduce the contribution of equipment electricity consumption to 16.05 % and 5.47 % of the new total footprint (without affecting the order of the top three contributing steps). Of course, in reality, each process tool and POU equipment has its own capacity factor, which would require performing real-time power consumption measurement for each of them.

Secondly, let us address the model uncertainties related to the impact assessment stage. In our case, most of the impact quantification was done using the *ecoinvent* database, which does not systematically provides uncertainty data for all of its entries, and does so with qualitative methods. We thus deem it unrelevant to attempt assessing the individual uncertainties of each inventory flow. However, a clear source of uncertainty comes from the energy and carbon intensity of PGMEA and TMAH[5], extracted from Kim & Overcash work [280] and estimated to vary between -75 % to +25 %. Indeed, they were obtained through Economic Input-Output (EIO) LCA considering different markets than those associated to the semiconduc-

[5]Used respectively in the photoresist and resist developer of the lithography step

tor industry [280]. However, since these two chemicals are used in very small volumes (see Table 4.1), such variations would have a marginal effect on the total CED and GWP of the lithography step, and wouldn't change the ranking of this step in the top contributors.

Scenarios analysis

To address the *spatial variability* of our results [266], we analyze the evolution of the contribution of the equipment electricity consumption to the total process CED and GWP in different scenarios of electricity production mixes (see Fig. 4.9).

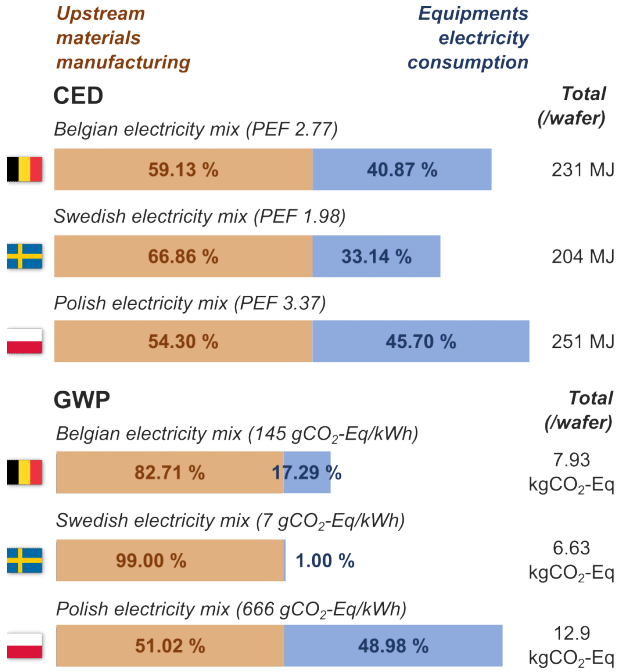


Fig. 4.9 Evolution of the relative contribution of the upstream materials manufacturing (orange) and the equipment electricity consumption (blue) to the total process CED and GWP, considering different electricity production mixes.

Indeed, as mentioned above, all previously presented results consider the electricity production mix of Belgium for the cleanroom consumption. In the case of Sweden, which has the less energy and carbon intensive mix among the European Union, this contribution would be reduced to 33.14 % and 1 % of the total process CED and GWP, but the top three most intensive steps would remain identical. In the case of Poland, which has the most energy intensive and carbonated mix in EU, the contribution of the equipment consumption would almost equal that of the materials manufacturing. The thermal oxidation step would climb to the third

most intensive step (above wafer fabrication), due to the furnace high electricity demand. All things considered, in each of the considered scenarios, the gold evaporation step remains the most intensive in terms of CED and GWP.

4.4.2 Eco-design perspectives

From this scenario analysis and the discussion on the main sources of uncertainty, there is now clear evidence that the dominant contribution to the CED and GWP associated to the VO_2 memristors fabrication process is related to the use of gold for their electrical contacts. This conclusion allows us to meet the primary goal of our LCA, which was to find a hotspot within the fabrication process: we can now discuss possible eco-designed alternatives. Gold and noble metals in general (which have similarly high CED and GWP) are often preferred to contact memristors based on materials such as Vanadium Dioxide [104], due to their high workfunction, low resistivity and to the fact that they do not oxidize over time or at high temperatures. Indeed, during its transition from high-resistive state (HRS) to low-resistive state (LRS), a VO_2 memristor can reach significantly high temperatures[6] [120]. However, alternative to noble metals for contacts have barely been investigated so far.

Among materials with small energy- and carbon-intensive manufacturing, one finds aluminium, with an energy and carbon intensity of respectively 131 MJ/kg and 8.2 $kgCO_2$-Eq/kg (w.r.t. 208 000 MJ/kg and 12500 $kgCO_2$-Eq/kg for gold [277]). Considering a constant deposited mass, replacing gold contacts by aluminium ones in the fabrication process studied in this work would result in a decrease of the total process CED and GWP by respectively -25.11 % and -43.38 %, down to 173 MJ/wafer and 4.49 $kgCO_2$-Eq/wafer. Moreover, aluminium - on the opposite of gold - is a CMOS compatible material, which would be desirable in the case of CMOS integration of VO_2 devices.

To test if such design choice would impact the device performances, we fabricated two batches of VO_2 memristors wafers with identical process parameters, only differing by their contact material (gold or aluminium). Their electrical characteristics are shown in Fig. 4.10. The contact resistance of the memristive devices was extracted using the transfer length method (TLM), in which the resistances of devices with identical electrode dimensions but varying electrode spacing L are measured. Extrapolating towards zero electrode spacing using a linear regression yields an estimation for the contact resistance $R_{contact}$, from the equation:

$$R_{tot}(L) = 2R_{contact} + \frac{\rho}{t}\frac{L}{W} \quad , \tag{4.1}$$

[6] Most of the works that report this heating do so in DC measurements; when operated dynamically, the device could be prevented from reaching such high temperatures, as it only spends very short periods of time in its metallic regime (depending on the spiking frequency).

with ρ the resistivity of the VO_2 film, t its thickness, L and W the VO_2 device dimensions, and R_{tot} the total measured VO_2 device resistance. Resistances are measured in the insulating regime of the VO_2 devices, close to zero bias.

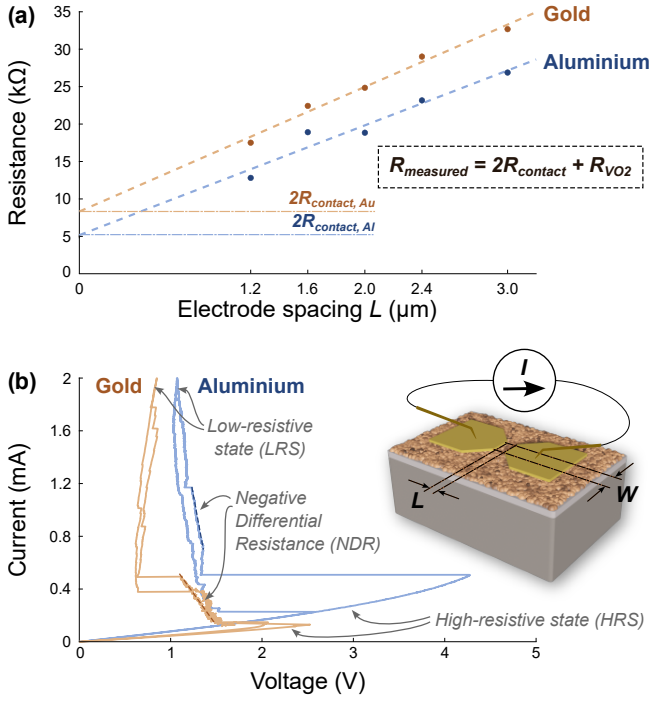


Fig. 4.10 Comparison of the electrical characteristics of VO_2 memristors fabricated with gold (orange) and aluminium (blue) electrical contacts: **(a)** extraction of the device contact resistance through the transfer length method (TLM), at T = 35 °C (the VO_2 film being in its high-resistive state) **(b)** current-driven I-V characteristic of two memristors of identical dimensions (L = 1.2 µm, W = 15 µm). Inset shows the measurement configuration and the dimension notations (since the VO_2 layer is not patterned, device dimensions are defined by electrode dimensions).

The study shows that despite the fact that aluminium typically has a higher resistivity than gold, the devices with aluminium contacts present smaller contact resistances. This difference could be attributed to the resistance of the metal/semiconductor junction itself, and/or to additional phenomena taking place at the Al/VO_2 interface, such as diffusion of aluminium ions into the VO_2. Fig. 4.10b shows that the resistive switching properties are preserved for both types of device: a reversible transition between a high-resistive state and a low-resistive state takes place. Both types of devices present a region of negative-differential resistance (NDR), a necessary condition for the devices to exhibit spiking behavior [160] as discussed in Section 3.1.1. Although not shown here, both devices have

been proven to be able to oscillate, with endurance measured up to 10^6 cycles in both cases.

4.4.3 Comparison with CMOS-based neurons

Beyond eco-design concerns, the second goal of our LCA is to estimate whether the fabrication of VO_2 neurons has a larger direct environmental footprint than that of CMOS-based spiking neurons[7]. As mentioned during the goal definition stage, for this comparison, we only consider the spiking functionality and leave out the sensing functions. Table 4.2 thus compares CMOS spiking neurons to VO_2 ones along 14 criteria. CED, GWP and UPW values associated to the CMOS nodes fabrication (with cradle-to-gate scope, as in our LCA), have been extracted from the work of Boyd[8] [259]. They must however be taken with caution, as large variability exists between distinct LCAs of CMOS processes [261], mostly due to confidentiality concerns on process details, and to scope mismatch between studies. Energy and power consumptions of the VO_2 neuron correspond to those of our high-frequency VO_2 oscillator described in Chapter 3, Section 3.2.3. The impact indicators are normalized for 1 cm² of die area.

A first conclusion is that the CED, GWP and UPW values associated to the fabrication of the VO_2 technology (normalized for 1 cm²) seem to be one order of magnitude below those of CMOS implementations (criteria 2-4). Indeed, CMOS process flows involve larger numbers of steps (up to 206 for a processor in 130 nm node [258]). For the GWP indicator in particular, chemical vapour deposition and dry etching steps, which are not present in our process, involve emissions of green house gases such as fluorinated gases and N_2O which have high global warming potentials. These steps also have large electricity consumption. However, several factors could explain the larger CED and GWP values within Boyd's work.

[7] A comparison with conventional CMOS-based, non-neuromorphic hardware would also be interesting to perform. However, VO_2 neurons are still at low TRL, and the characteristics of the functional unit required to make a comparison with a CMOS conventional processor (density, yield, number of devices needed to reach equivalent functionalities) are not yet known at this stage.

[8] Chemicals manufacturing, fabrication of the silicon wafer, and impact of the equipment operation have been included. The electricity mix corresponds to the world average in 2002 (10.7 MJ/kWh). No abatement is considered. Impacts related to the fab infrastructure, transportation, and use phase have been left out to match the scope of our study, although available in Boyd's work [259]. For UPW, we assume a yield of 0.5 liter UPW for 1 liter feedwater.

Table 4.2 Comparison between CMOS- and VO_2-based spiking neuron implentations

Reference	(1) Technology node	(2) CED/cm² (MJ)	(3) GWP/cm² (kgCO2-Eq)	(4) UPW/cm² (l)	(5) Die area for one neuron (cm²)	(6) Active area (µm²)
Indiveri [284]	CMOS 350 nm	6,07E+01	6,22E+00	1,73E+01	/	2,57E+03
Cruz-Albrecht [285]	CMOS 90 nm	3,29E+01	3,00E+00	1,21E+01	/	4,42E+02
Sourikopoulos [286]	CMOS 65 nm	3,41E+01	3,43E+00	8,57E+00	1,13E-03	3,50E+01
Nalliboyina [287]	CMOS 45 nm	3,53E+01	3,00E+00	3,57E+00	/	4,17E+01
This work	VO_2	5,07E+00	1,74E-01	3,20-01	6,60E-04	2,40E+00
Yuan [96]	VO_2	/	/	/	/	4,00E-01
Han [92]	VO_2	/	/	/	/	1,96E+01
Yi [85]	VO_2	/	/	/	/	3,00E+00

Reference	(7) CED/die (MJ)	(8) GWP/die (kgCO2-Eq)	(9) UPW/die (l)	(10) CED/neuron (MJ)	(11) GWP/neuron (kgCO2-Eq)	(12) UPW/neuron (l)	(13) Energy consumption	(14) Static power consumption	
Indiveri [284]	/	/	/	1,56E-03	1,60E-04	4,46E-04	900 pJ/spike	/	
Cruz-Albrecht [285]	/	/	/	1,46E-04	1,33E-05	5,37E-05	0.4 pJ/spike	40 pW	
Sourikopoulos [286]	3,84E-02	3,86E-03	9,64E-03	1,20E-05	1,20E-06	3,00E-06	4 fJ/spike	100 pW	
Nalliboyina [287]	/	/	/	1,47E-05	1,25E-06	1,49E-06	1.32 fJ/spike	23 pW	
This work	3,34E-03	1,15E-04	2,11E-04	1,22E-07	4,17E-09	7,68E-09	5 nJ/spike	5 mW	
Yuan [96]	/	/	/	/	/	/	2.9 nJ/spike	5 mW	
Han [92]	/	/	/	/	/	/	3 nJ/spike	100 µW	
Yi [85]	/	/	/	/	/	/	100 fJ/spike*	11 µW*	*simulated

First, the fact that the CED metrics are computed using a Primary Energy Factor for electricity production reflecting the world's average electricity mix in 2002, almost four times larger than the Belgian PEF considered in our estimation. Second, the fact that the GWP values do not consider any abatement of greenhouse gases[9]: when considering ideal abatement, they go down by a factor ∼2. Furthermore, the yield of our process is considered to be 100 % as it could not be characterized, while non-ideal yields are accounted for at several steps in the estimations of Boyd.

In addition, it is worth noting that the VO_2 devices analyzed within this LCA are currently produced within a lab-scale fab. It is hard to predict how the equivalent footprint of 1 cm² of die area of VO_2 memristors would evolve with industrialization. Some factors could decrease this footprint: equipment loading factors could be optimized (in our case, very few wafers are processed simultaneously during evaporation and sputtering steps); wafer size could be increased; power consumption of the equipment could be optimized; re-flow and filtering of wet chemicals could reduce their required quantities. However, some other factors could worsen the footprint: to remove impurities and increase the yield, larger quantities of chemicals could be used in resist stripping and wafer cleaning steps[10]; miniaturization could increase the number of lithography steps, possibly introducing dry-etching steps including fluorinated gases, ...

Despite the difference in impact indicators values when normalized for a die area of 1 cm², VO_2 neurons also offer a smaller die area and active device area compared to CMOS spiking neurons (see criteria 5, 6), since they only require a single device to implement neuron behavior. This results in a smaller equivalent footprint when normalized for the die area (criteria 7-9) and active neuron area (criteria 10-12). Here, we exclude the external transistor and capacitor used in our neuron circuit; indeed, we have seen that it is possible to implement spiking behavior using solely the VO_2 device [92] (see benchmarking in Chapter 3, Section 3.2.3). It is however clear that VO_2 neurons still have a much higher static power consumption and energy consumption per spike compared to CMOS-based neurons (criteria 13, 14). Optimization is thus still required for VO_2 neurons to become competitive regarding their use phase. It is however noteworthy that CMOS and VO_2 neurons might not be used for the same purpose. Indeed, CMOS neurons would be good candidates for computational neurons in large-scale spiking neural networks, while VO_2 neurons could solely be used for the sensitive entry layer, hence tolerating a slightly higher power consumption.

[9] Abatement systems are dedicated equipment which treat the exhaust of process tools that emit greenhouse gases at their output. It can be performed through various techniques, including combustion, plasma and UV [288].

[10] The work of Krishnan [258] provides a detailed inventory for specific steps of CMOS processes. Cleaning and resist stripping steps respectively use about 10 times the amount of H_2O_2 and TMAH used in the corresponding steps of our fabrication process

4.5 Key messages

The main conclusions of this chapter are summarized here.

Process Life Cycle Assessment of an emerging technology, such as the one performed in this chapter, allows to build high-precision inventories, reducing uncertainties related to this specific stage of the LCA. A few limitations however remain:

- While quantities of input materials can be measured with high accuracy, obtaining a similar precision on the estimation of process tools and point-of-use equipment electricity consumption would require systematic *in situ* measurements using power logger. Indeed, estimation based on rated power seem to overestimate real time consumption.
- Elementary flows associated to costs shared by the whole fab, such as the fab infrastructure construction, process tools fabrication, fab utilities electricity consumption (HVAC, water treatment, ...) have been discarded in our analysis, due to the difficulty to allocate them to our specific process. However, they can represent significant proportion of the cradle-to-gate footprint. In lab-scale fab where several process flows are run concurrently on shared equipment and facilities, relevant allocation rules should be defined to account for such costs.

In the fabrication process of the VO_2 spiking neurons studied in this work, the dominant impact on the Cumulated Energy Demand, Global Warming Potential and Ultra Pure Water consumption insicators are associated to the use of gold for the electrical contacts. This is principally due to the extraction and refining of gold, which is associated to high energy requirements and emissions.

- Replacing gold by aluminium would result in a 25 and 43 % decrease in CED and GWP, respectively.
- Firt samples have demonstrated that the transition and spiking functionalities of VO_2 devices are preserved when using aluminium instead of gold for contacts. Further tests should be perfomed to compare endurance properties, especially since aluminium can oxidize over time.
- As gold and other precious metals with similar impacts are almost always used for electrical contacts in VO_2 memristive device, this alternative could reduce the overall footprint of the technology upon industrialization. It also offers reduced costs and CMOS compatibility.

Beyond the replacement of gold, further optimization could be realized to reduce the footprint of the VO_2 devices fabrication process.

- Since the liftoff step is the second most impactful in terms of CED and GWP, mainly due to the large quantities of dimethyl sulfoxide resist stripper used, reflow and filtering of this specific chemical could be put in place. This is likely to be the case in an industrialization scenario.
- Fabrication of the silicon wafer and thermal oxidation steps, performed in order to obtain a SiO_2 buffer layer for electrical insulation and for the growth of the VO_2 thin film, also rank third and fourth in terms of CED contribution. Swapping the Si/SiO_2 substrate by glass or a flexible substrate [92, 95] could be an interesting, lower impact swap, at the cost of loss of CMOS-compatibility.

Compared to CMOS spiking neurons, the fabrication of VO_2 spiking sensors *seem* to be associated to smaller impacts in terms of CED, GWP and UPW. This is both due to the lower footprint of the fabrication process, and the smaller area of VO_2 neurons. Two important elements nuance this claim:

- While CMOS process are technologically mature, process flows of VO_2 devices have not been industrialized yet. **This assessment should thus be re-performed by projecting the evolution of the VO_2 process with industrialization, as lots of factors could vary compared to the current lab-scale prototype. This could heavily impact our conclusions.**
- VO_2 spiking neurons still have much lower energy efficiencies than CMOS spiking neurons, with 6 orders of magnitude difference in spike energy, showing that further optimization is required for them to be competitive during their use phase. Optimization such as miniaturization are likely to increase the process complexity, and the associated environmental impacts.

In this chapter, we performed a so-called *ex-ante LCA*, i.e. an LCA of an emerging technology which is still at lab-scale. To the best of our knowledge, it is the first time such approach is applied to the field of neuromorphic technologies. Our main message is that this method looks useful to identify hotspot within the fabrication process and possible lower-impact alternatives. However, using such tool to compare emerging technologies to more mature ones must be done with caution. Models to predict how certain parameters evolve upon industrialization of a microfabrication process in the semiconductor industry (such as equipment load factors, yield, equipment up time, quantities of input materials) would be desirable, in order to truly determine the potential environmental benefits or drawbacks of an emerging technology against the state of the art.

Conclusions and Perspectives

99 % of my training is falling.
– Janja Garnbret

In recent years, a large body of work started to develop on the subject of memristive devices, and more recently on their application as **spiking sensory neurons** to be interfaced as the entry layers of Spiking Neural Networks. Taking inspiration from the brain to implement two paradigm shifts (**1. memory co-location** and **2. sparse, spike-based encoding of information**), such neuromorphic sensor-and-processing systems hold theoretical promises of higher energy efficiency, which however remain to be proven. Beyond energy-efficiency, memristive spiking sensors have great potential in terms of scalability, as they allow to transduce stimuli and emulate neuronal behaviors using a very small number of components. Among the myriad of emerging materials that can be used to realize spiking neurons, VO_2 shows two main advantages. The first relates to its **volatile resistive-switching properties**, thanks to its reversible insulator-to-metal transition, and its ability to enter a **self-sustained oscillation regime**, thanks to its negative differential resistance. This allows VO_2 memristors to emulate complex behaviors such as LIF and H&H neurons. The second advantages is owed to the fact that the electrically-induced insulator-to-metal transition can be modulated by a variety of stimuli, offering the possibility to **perform stimulus transduction directly within the VO_2 memristor**. However, VO_2 spiking sensors are still at the beginning of their development, with the earliest research on the subject dating from 2021 [91, 169]. The goal of this thesis was to tackle the limitations that currently hinder the development of such sensors, focusing on four of them:

Limitation #1 The inherent cycle-to-cycle variations in their electrical properties, which are not yet well understood nor controlled (Chapter 2).

Limitation #2 The impact of these cycle-to-cycle variations on the performances of VO_2 spiking sensors, which deteriorate their Limit of Detection and introduce an energy-resolution trade-off (Chapter 3, Section 3.1 and 3.2).

Limitation #3 The limited sparsity of rate-based encoding, dominantly used in present implementations (Chapter 3, Section 3.3).

Limitation #4 The lack of information regarding the direct environmental impacts associated to their fabrication process, despite promises of applications for "Green AI" purposes (Chapter 4).

We re-state below the four main research questions associated to these limitations, and the specific findings brought by this thesis in response to them. We also provide a more complete benchmarking of VO_2 spiking sensors against other memristor-based implementations, CMOS event cameras, and CMOS neurons. We discuss the main limitations that still further hinder the development of VO_2 devices, which were not discussed in detail in this work. Finally, we highlight the main following research tracks that we deem relevant to pursue in future work, and give some final thoughts regarding the possible applications of these devices.

Specific findings

Chapter 2

In this chapter, we delved into **Limitation #1** and highlighted the present lack of guidelines to design VO_2 memristors with minimized or maximized cycle-to-cycle stochasticity, due the poor understanding of the random nature of resistive switching. Such stochasticity can play a beneficial or detrimental role depending on the specific device application, underlying the need for the mentioned design guidelines. We thus asked the following question:

How can specific material properties and device design parameters impact the resulting VO_2 two-terminal devices cycle-to-cycle stochasticity?

To answer this question, we fabricated resistive-switching VO_2 devices of various dimensions and grain sizes, and extracted the cycle-to-cycle variations of their electrical characteristics through hundreds of DC measurements. To the best of our knowledge, this is the first time a comparative approach is followed to determine the parameters that impact cycle-to-cycle variations in memristive devices. This allowed us to formulate the following claim:

Claim #1 The cycle-to-cycle stochasticity in VO_2 two-terminal devices can be mitigated by increasing the number of grains along the device channel length. This can be performed either by reducing the thin film grain size (which we tuned by varying the film annealing temperature), or increasing the device channel

length. More specifically, reducing grain size by one order of magnitude allows to reduce the cycle-to-cycle variations in the device electrical characteristics in the same proportions. Previous works who achieved low cycle-to-cycle variations in VO_2 device (<1 % variations in threshold voltage) turned to epitaxial growth to obtain highly regular crystallites. This technique however requires specific substrates such as Al_2O_3, which are not straightforwardly CMOS-compatible. In our work, we proved that we could obtain similar to slightly lower variations simply by tuning the grain size and device dimensions using polycrystalline VO_2 films grown on CMOS-compatible Si/SiO_2 substrates.

Chapter 3

In the first part of this chapter, we discussed **Limitation #2** and tried to understand the impact of the cycle-to-cycle variations studied in Chapter 2 on the spike rate jitter, when VO_2 memristors are operated in self-sustained oscillators. We then studied the potential of such oscillators to be used as spiking temperature sensors, quantifying how these variations would impact their performance. Our central questions were:

What are the dominant causes of dynamic stochasticity in rate-coded spiking VO_2 sensors? How does this stochasticity impact the sensor Limit of Detection?

To answer them, we built a relaxation oscillator from the memristive VO_2 device fabricated in Chapter 2. We used a stochastic analytical dynamic model to predict the oscillator output, to which we included various noise sources, among which the experimentally extracted cycle-to-cycle variations of the VO_2 device electrical properties. We then analyzed the sensitivity of our VO_2 oscillator to temperature, and studied how these variations impacted its resolution at various frequencies. From this analysis, we express the following conclusions:

Claim #2 The cycle-to-cycle variations in the VO_2 device electrical properties in its insulating regime are dominantly responsible for the spike rate jitter, above variations of the parameters in its metallic regime, or other white noise sources such as transistor noise or thermal noise on the VO_2 resistor.

Claim #3 The spike rate jitter directly affects the Limit of Detection, which can be improved by averaging the spike rate over a certain number of cycles. The associated energy-resolution trade-off can be relaxed by operating the oscillator at higher frequency. This allows to observe an equivalent number of cycles at a smaller energy cost, and relies on the fact that the cycle-to-cycle variations seem to remain constant at higher frequencies.

In the second part of this chapter, we investigated **Limitation #3** which states that present implementations of VO_2 spiking sensors do not provide sparse encoding of stimuli, as they perform rate-coding without real silent periods between spike events. We thus attempted to answer the question:

Can VO_2 resistive-switching devices be used to design sensitive spiking neurons with sparser encoding techniques than plain rate-coding?

By exploring a new operating region of our VO_2 relaxation oscillator, we showed that it could emit bursts of spikes separated by quiescent time intervals. We investigated the evolution of these bursts with temperature, and showed for the first time that a stimulus could be encoded into the temporal characteristics of these bursts, akin to the response of insect sensitive neurons to noxious heat. Based on these results, we formulate the following claim:

Claim #4 Beyond plain rate coding, VO_2 spiking sensors can encode stimulus information into the temporal characteristics of spike bursts. This new kind of self-induced sparsity in a spiking sensor could pave the way for more energy-efficient neuromorphic systems, if integrated with a SNN. In order to offer energy-efficiency gains at the device level, the static energy consumption of VO_2 spiking sensors should however be reduced below their dynamic spike energy, through e.g. miniaturization.

Chapter 4

With this chapter, we wanted to challenge the common narrative of literature which regularly promotes memristive applications as a possible mean to curb the energy consumption of large AI models. Indeed, up to this day, no research attempted to quantify the environmental footprint related to the fabrication of such devices, which should also be taken into balance when making these claims. We thus ask the following questions:

What are the direct environmental impacts associated to the fabrication of VO_2 memristive devices? Can they be reduced through eco-design, and how do they compare to those of CMOS spiking neurons?

To address this matter, we performed a cradle-to-gate Life Cycle Assessment of our VO_2 device. We constructed a highly precise inventory of the elementary flows involved in the process (quantity of upstream materials, ultra-pure water, and equipment electricity consumption). We then evaluated the impact of the full process flow on three metrics classically evaluated in the production of integrated

circuits: GHG, CED and UPW. To the best of our knowledge, this is the first time an LCA of a neuromorphic technology is performed. Our main conclusions are summed up in this claim:

Claim #5 At lab-scale, the direct environmental impacts associated to the fabrication of VO_2 memristive devices seem to be lower than those of CMOS neurons. This conclusion is however very likely to evolve upon industrialization. **This footprint can be significantly reduced by replacing the use of gold for their contacts by other, non-noble metals such as aluminium.** More specifically, the associated emissions (GWP) and primary energy demand (CED) are reduced by 25 and 43 %, respectively. First characterizations of device fabricated with aluminium contacts show that resistive switching and self-sustained spiking behavior are preserved. This conclusion is especially relevant as all currently reported VO_2 memristive devices have been fabricated using precious metals. If adopted, this eco-design guideline could thus steer the further development of the technology in a lower-impact direction.

Final benchmarking

How do the specific VO_2 spiking sensors developed in this work[11] compare to other VO_2 neurons, and to other technological implementations of sensitive neurons? We answer this question in terms of energy efficiency (energy per spike E_{spike}) and neuron area (Fig. 6.1), taking into account spike rate (Fig. 6.2).

VO_2 spiking sensors are still at the very early stage of their development: beside our work, only two implementations demonstrate temperature sensing embedded directly within the VO_2 device [92, 181]. We thus include VO_2 spiking sensors sensitive to other stimuli, with some of them relying on external sensors for the transduction. Our device performs close to the 3 nJ/spike SoA [96] in terms of spike energy, but 6 times worse than the 0.4 µm² area SoA [96] in terms of neuron area. However, the latter relies on an external sensor which is not included in the area estimation. In addition, very few works report on the spike rate jitter and associated LoD (introduced in Chapter 3), which are required to provide true energy consumption comparison at equivalent resolution performances: we thus call for a more systematic evaluation of such metrics.

[11] We consider the low- and high-frequency versions of our rate-coding sensor, introduced respectively in Section 3.2 and 3.2.2. We do not include our bursting sensor.

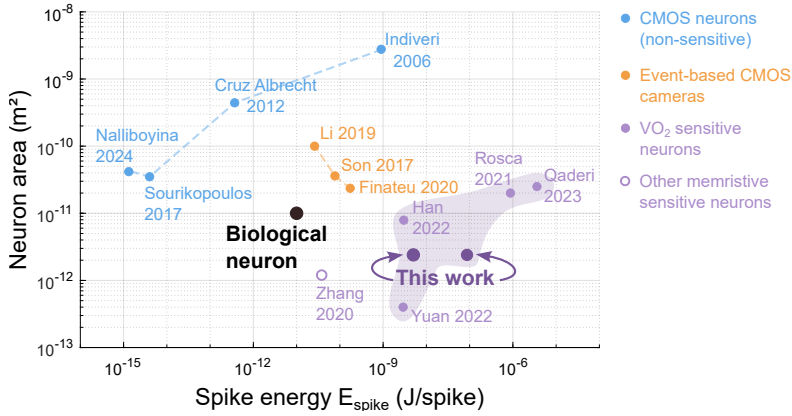


Fig. 6.1 Benchmarking of our work against different sensitive spiking neuron implementations, in terms of energy per spike E_{spike} and neuron active area. Several families of implementations are represented. **(1) Analog memristive spiking sensors *(in purple)***, including VO_2 implementations [91–93, 96] and one NbO_x implementation [88], sensitive to different stimuli (with transduction performed by the memristive device or by an external sensor). **(2) Mixed-signal event-based CMOS cameras *(in orange)***, operating in the visible light domain using photodiodes [6, 289, 290]. **(3) Analog sub-threshold CMOS computational spiking neurons *(in blue)***, which do not include sensing functionalities, but exhibit LIF [284, 285] and more complex neuron behaviors [286, 287]. Our specific contributions (low- and high-frequency versions of our rate-coding sensor) and an average biological neuron reference [291] are highlighted in bold. In each case, the spike energy is computed at the maximum measured spike rate. In the case of memristive neurons, neuron area only includes the memristor active area.

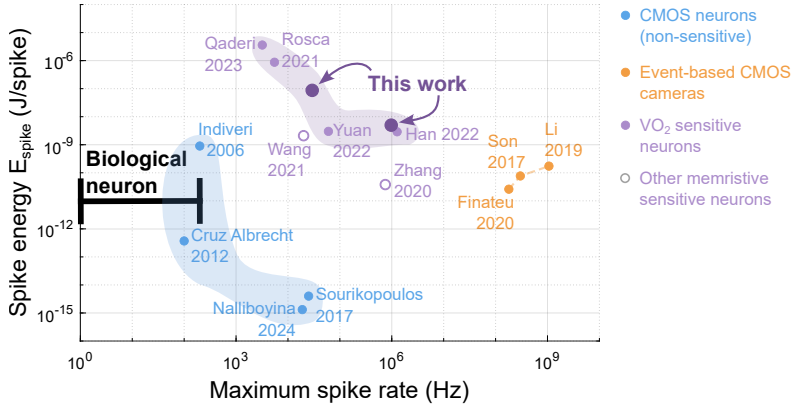


Fig. 6.2 Benchmarking of of our work against different sensitive spiking neuron implementations, in terms of energy per spike E_{spike} and spike rate. Implementation families are the same as in Fig. 6.1. The work of Wang et al., based on metal ions migration, is additionally mapped within the memristive sensitive neuron family [99]. The spike rate of biological neurons spans from 1 to 200 Hz [52].

Regarding other memristive sensitive spiking neurons, few implementations provide spike energy estimation. As introduced in Chapter 1, the two main other memristive technologies used to realize sensitive spiking neurons are NbO_x memristors [87–90] and metal-ions migration memristors [64, 82, 98, 100–102, 292]. The SoA in terms of spike energy belongs to a NbO_x memristor [88], down to 38 pJ/spike (which approaches the 10 pJ/spike estimation of biological neurons), at a ten-times smaller area ($\sim$ 1 µm² compared to $\sim$ 10 µm² for biological neurons).

Event-based CMOS cameras [6, 289, 290], which rely on silicon photodiodes to detect visible light and include mixed-signals interfaces for spike generation, exhibit lower spike energy than VO_2 sensitive neurons. This is in part due to their higher operating frequencies, as shown in Fig. 6.2; and partially thanks to their lower static power consumptions (in the nW range, *vs* µW to mW for VO_2). Event-based cameras however have larger pixel area than VO_2 neurons, as they involve complex circuits with tens of transistors. On the other hand, they implement more complex functions, with polarized spikes based on contrast which could be more energy efficient if evaluated together with a SNN.

Other types of CMOS spiking neurons do not yet demonstrate sensing functionality [284–287], but could *in theory* be integrated with the output of an analog dedicated sensor, as conceptually proposed recently [71–73]. These implementations rely on analog sub-threshold CMOS technology, and achieve lower spike energy than VO_2 memristors, event-based cameras, and even biological neurons (down to the fJ/spike), with however larger areas (which are progressively scaling down with Moore's law). However, competitive integration of sensing functionalities must still be demonstrated.

In conclusion, VO_2 memristive spiking sensors are still outperformed by other memristive devices, CMOS event-based cameras, and possibly future implementations of CMOS spiking neurons in terms of spike energy. Like other implementations, they typically operate at accelerated timescales compared to biological spike rate. While biological neuron time constants are the product of evolution and are thus optimized to process real-life input stimuli, some specific applications could benefit from accelerated timescales (e.g. obstacle avoidance in autonomous cars, real-time digital signal processing of video inputs in media applications, ...). The main advantage of VO_2 spiking neurons lies in their small area (below biological ones[12]), which offers promising scalability perspectives. Compared to other memristive implementations, their main asset relates to the possibility of embedding transduction of various stimuli within the VO_2 device, thanks to the multi-stimuli responsiveness of its transition.

[12]The $\sim$10 µm² span of biological neurons was derived by Poon et. al [291] based on the average neuron density and cortical surface in mammalian species.

Limitations and perspectives

We now discuss the limitations related to this work and to VO_2 spiking sensors in general, and identify several key tracks on which research could be deepened.

Improve the understanding of the stochastic behavior. Much could be gained in the understanding of the physical mechanisms behind the intrinsic stochasticity of resistive switching in VO_2. Ideally, (ultra-fast) high-resolution imaging of the filament formation [219, 225, 229], which was so far not performed across a large number of cycles, could help us grasp such mechanisms. It would be especially relevant to image consecutive filament formations in devices showing bi-modal distributions in their threshold voltage, to confirm our hypothesis that they correspond to distinct filament locations. Once the proper cause of stochasticity is identified through such imaging, our claim regarding the impact of grain size and channel length on stochasticity could be confirmed using resistance matrix models[13] [121, 162, 165, 232, 293]. In addition, the possible role of the films crystalline orientation on the device stochasticity should be systematically investigated by thoroughly characterizing the degree and direction of crystalline ordering in each film, using grazing incidence XRD or electron backscatter diffraction (EBSD). Finally, our study focuses on the impact of the devices length on their cycle-to-cycle stochasticity: the impact of their width should also be investigated through additional measurements.

Extend our conclusions to other materials and stimuli. To assess their reach and generality, conclusions of Chapter 2 regarding the ways of mitigating stochasticity should be verified on other resistive switching materials with similar filamentary process, such as NbO_x or V_2O_3. Similarly, the rate-coding and, more importantly, the burst-coding behavior demonstrated in Chapter 3 could be extended to other stimuli than temperature, given the multi-stimuli responsiveness of VO_2 highlighted in Chapter 1. Strain, most specifically, holds potential to be detected directly within VO_2 neurons [148], although other device architectures such as nanowires could be needed to reach sufficient strain.

Investigate endurance and device-to-device variability. A main limitation of VO_2 devices which was not fully investigated in this thesis is related to the electro-forming step that takes place during the initial cycling of certain devices [166, 216, 221]. This phenomenon is related to the high current density that arises locally during filament formation. It will be more pronounced in devices

[13] Our preliminary tests show that if large number of cycles must be stimulated and thermal effects are taken into account, such simulations can require a considerable amount of computational resources, possibly prescribing the need of dedicated servers.

of larger length, and in those built from films with larger OFF/ON resistivity ratios [120]. In some cases, it can profoundly alter the grain structure: although transition properties can be preserved after this electro-forming step, it often induces large device-to-device variations [159]. From our characterizations, we suspect such phenomenon can also take place in subsequent cycles, limiting the device endurance. In our work, no electro-forming steps were observed in the small-grains devices, which we used to fabricate spiking neurons. They were however present with limited damages in the large-grains devices studied in Chapter 2, and could possibly nuance our conclusions (see Appendix A). Fabricating films with a smaller OFF/ON resistivity ratio (by using lower annealing temperature) thus seem to be a way to prevent this phenomenon, although it will be necessary to go to small dimensions to observe an NDR. Beyond this electro-forming behavior, other mechanisms could limit the device endurance (fatigue due to repeated thermal expansion upon resistive switching, progressive oxidation) and should be further investigated.

Optimize the energy consumption. Several routes come forward when attempting to optimize the spike energy. A first obvious idea is to increase the spiking frequency, which can be done straightforwardly by **reducing (or completely removing [92, 181]) the output capacitor**. As mentioned in Claim # 3, this also allows to relax the energy-resolution trade-off. Some works however hint there could be intrinsic limits to how fast a VO_2 device can oscillate, possibly related to the heat transfer dynamics [294]. Another track is to attempt to reduce the device IMT bias point, which seems to be the case in **vertical topologies** [92, 95] or in miniaturized planar topologies (i.e. **devices with smaller length and width**). Luckily, we have seen in Chapter 3 that small-dimensions devices are also those who are the more likely to present an NDR, which is a condition to observe a self-sustained spiking behavior. Miniaturization could also reduce the device thermal capacitance, hence increase the heat transfer dynamics and the spiking frequency limit. Finally, maintaining low cycle-to-cycle variations is essential to preserve the energy-resolution trade-off, which can be done by **using films with smaller grains** *or* increasing device length. However, the latter strategy will also increase the IMT bias point, introducing another trade-off. The impact of the film thickness could also impact the device power consumption and cycle-to-cycle stochasticity (especially in vertical topologies), and should be further investigated.

Extend the bursting neuron regime. The bursting operating regime of our VO_2 sensor is currently limited to a very narrow range of stimulus power (in our case, temperature). Indeed, the specific bias range in which this behavior can be observed is much narrower than for rate-coding, making burst coding way less robust. Integrating a feedback mechanism could allow us to remain in this regime independently of the stimulus power. Encoding would then be performed by the

ISI, but with added self-induced sparsity. Additionally, the absence of a silent period between individual spikes - characteristic of most present VO_2 neurons, as they are based on an oscillator implementation - makes them less bio-plausible. Integrating a feedback mechanism [103] could also allow to include refractory periods in between spikes, further increasing their sparsity.

Deepen LCA and eco-design. Prospective LCA considering the scale-up of the technology if fabricated in an industrial production line should be performed to challenge the conclusions of Chapter 4. In addition, more extensive characterization of the devices fabricated with aluminium contacts should be performed to confirm the reliability and performances of this eco-designed alternative.

Beyond rate-coding and integration with a Spiking Neural Network. Research suggested that rate-coding might not be the most energy efficient encoding when it comes to SNN performances. However, sensory neurons must not necessarily have the same encoding as the first layer of the SNN: sparsity could be implemented by the latter. The true energy efficiency of a spiking sensor should be assessed in a systemic approach, when integrated with a Spiking Neural Network.

Final thoughts

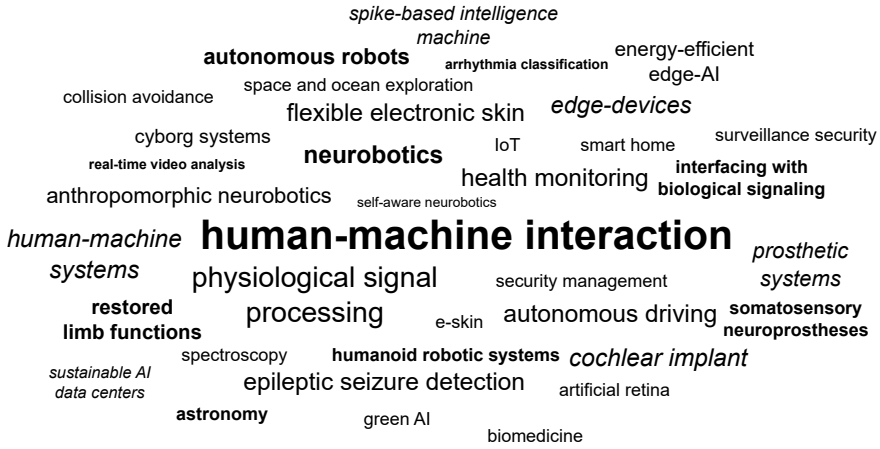


Fig. 6.3 Word cloud summing up some of the frequent words found in the abstracts, introductions and conclusions of research publications related to artificial spiking neurons, when authors suggest possible applications. Word size does not correlate to the occurrence frequency.

To conclude this thesis, we wanted to take a step back and reflect on the possible applications of artificial spiking neurons, even though they are still at their early development stage. To embark the reader on this reflection, we gathered a few of the most commonly put forward vocabulary found in the abstracts, introductions and conclusions of research papers related to artificial spiking neurons and neuromorphic hardware, when authors attempt to describe possible applications. The applications put forward by researchers are quite diverse, ranging from the treatment of medical conditions, to smart prosthetics, energy-efficient AI and autonomous cars. Recurring themes include the concept of brain-machine interfaces and augmented robots. A common point is that very few of these works demonstrate an actual proof-of-concept related to the evoked application case.

Researchers do not have the power to decide for which specific application should a technology be developed. However, this shows us that researchers are not neutral regarding these choices. Each author has its own biases, which will translate more or less subtly in its *storytelling*, or the way he decides to showcase its findings and their potential. Such storytelling, also called "promising communication" in sociology [295], can have a strong impact on the course of technological development [296]. Indeed, these prospective discourses can influence funding of research programs, private and public technology roadmaps, underlying the responsibility we have as scientists in the way we choose to present our results.

Ethical risks related to the development of brain-computer interfaces and AI, two applications commonly put forward by researchers, are increasingly discussed in the scientific literature [297–299]. However, these subjects are still rarely addressed within the neuromorphic community (although certain counter-examples exist [300]). While the reason for the absence of such discussions might be the ideal of scientific neutrality shared by a lot of engineers, Fig. 6.3 shows that some of this neutrality has already being lost along the way, as we show partiality in the applications we chose to put forward. In our opinion, low TRL research - prior to industrialization and general public market opening - is the right place and time to open up these debates. We believe it would be crucial to discuss the potential risks and benefits of these technologies within the neuromorphic community, throughout workshops during conferences and within research groups, ideally integrating researchers from other disciplines than engineering and neuroscience such as sociologists, ethicists, medical doctors, psychologists and philosophers. Researchers in the field of neuroethics share this opinion, as they believe integrating ethical concerns early into the technology design would maximize benefit and minimize potential harm [298]. In addition, we believe vulgarizing such research to the general public will open a much needed democratic debate on the desirability of these technologies, and on the specific applications that we, as society (and not just us, engineers), want to see developed.

Bibliography

[1] L. Eeckhout, "Is Moore's Law Slowing Down? What's Next?" *IEEE Micro*, vol. 37, no. 4, pp. 4–5, 2017.

[2] "Cisco Annual Internet Report (2018–2023) White Paper," Cisco, Accessed on 2024-07-2. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[3] L. Strous and V. G. Cerf, *Internet of Things. Information Processing in an Increasingly Connected World*. Springer International Publishing, 2019, vol. 548.

[4] D. Bol, T. Pirson, and R. Dekimpe, "Moore's Law and ICT Innovation in the Anthropocene," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, Feb. 2021, pp. 19–24.

[5] W. Maass, "To Spike or Not to Spike: That Is the Question," *Proceedings of the IEEE*, vol. 103, no. 12, pp. 2219–2224, 2015.

[6] T. Finateu, A. Niwa, D. Matolin, K. Tsuchimoto, A. Mascheroni, E. Reynaud, P. Mostafalu, F. Brady, L. Chotard, F. LeGoff, H. Takahashi, H. Wakabayashi, Y. Oike, and C. Posch, "5.10 A 1280×720 Back-Illuminated Stacked Temporal Contrast Event-Based Vision Sensor with 4.86µm Pixels, 1.066GEPS Readout, Programmable Event-Rate Controller and Compressive Data-Formatting Pipeline," in *2020 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2020, pp. 112–114.

[7] P. Lichtsteiner, C. Posch, and T. Delbruck, "A 128\times 128 120 dB 15 us Latency Asynchronous Temporal Contrast Vision Sensor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 2, pp. 566–576, 2008.

[8] G. Zhou, Z. Wang, B. Sun, F. Zhou, L. Sun, H. Zhao, X. Hu, X. Peng, J. Yan, H. Wang, W. Wang, J. Li, B. Yan, D. Kuang, Y. Wang, L. Wang, and S. Duan, "Volatile and Nonvolatile Memristive Devices for Neuromorphic Computing," *Advanced Electronic Materials*, vol. 8, no. 7, p. 2101127, 2022.

[9] J. J. Yang, D. B. Strukov, and D. R. Stewart, "Memristive devices for computing," *Nature nanotechnology*, vol. 8, no. 1, pp. 13–24, 2013.

[10] G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. Le Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, "Neuromorphic computing using non-volatile memory," *Advances in Physics: X*, vol. 2, no. 1, pp. 89–124, 2017.

[11] B. Vogginger, A. Rostami, V. Jain, S. Arfa, A. Hantsch, D. Kappel, M. Schäfer, U. Faltings, H. A. Gonzalez, C. Liu, and C. Mayr, "Neuromorphic hardware for sustainable AI data centers," 2024.

[12] T. Luo, W.-F. Wong, R. S. M. Goh, A. T. Do, Z. Chen, H. Li, W. Jiang, and W. Yau, "Achieving Green AI with Energy-Efficient Deep Learning Using Neuromorphic Computing," *Communications of the ACM*, vol. 66, no. 7, pp. 52–57, 2023.

[13] K. Yamamoto, A. Uno, H. Murai, T. Tsukamoto, F. Shoji, S. Matsui, R. Sekizawa, F. Sueyasu, H. Uchiyama, M. Okamoto, N. Ohgushi, K. Takashina, D. Wakabayashi, Y. Taguchi, and M. Yokokawa, "The K computer Operations: Experiences and Statistics," *Procedia Computer Science*, vol. 29, pp. 576–585, 2014.

[14] W. Maass, "Networks of spiking neurons: The third generation of neural network models," *Neural Networks*, vol. 10, no. 9, pp. 1659–1671, 1997.

[15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[16] I. I. Arikpo, F. U. Ogban, and I. E. Eteng, "Von neumann architecture and modern computers," *Global Journal of Mathematical Sciences*, vol. 6, no. 2, pp. 97–103, 2007.

[17] W. Kautz, "Cellular Logic-in-Memory Arrays," *IEEE Transactions on Computers*, vol. C-18, no. 8, pp. 719–727, 1969.

[18] N. Verma, H. Jia, H. Valavi, Y. Tang, M. Ozatay, L.-Y. Chen, B. Zhang, and P. Deaville, "In-Memory Computing: Advances and Prospects," *IEEE Solid-State Circuits Magazine*, vol. 11, no. 3, pp. 43–55, 2019.

[19] W. Jiang, P. Houshmand, M. Verhelst, and W. Dehaene, "A 16nm 128kB high-density fully digital In Memory Compute macro with reverse SRAM precharge achieving 0.36TOPs/mm2, 256kB/mm2 and 23. 8TOPs/W," in *ESSCIRC 2023- IEEE 49th European Solid State Circuits Conference (ESSCIRC)*, 2023, pp. 409–412.

[20] S. Yin, Z. Jiang, J.-S. Seo, and M. Seok, "XNOR-SRAM: In-Memory Computing SRAM Macro for Binary/Ternary Deep Neural Networks," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 6, pp. 1733–1743, 2020.

[21] J. Zhang, Z. Wang, and N. Verma, "In-Memory Computation of a Machine-Learning Classifier in a Standard 6T SRAM Array," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 4, pp. 915–924, 2017.

[22] Q. Xia and J. J. Yang, "Memristive crossbar arrays for brain-inspired computing," *Nature Materials*, vol. 18, no. 4, pp. 309–323, 2019.

[23] J. L. Lobo, J. Del Ser, A. Bifet, and N. Kasabov, "Spiking Neural Networks and online learning: An overview and perspectives," *Neural Networks*, vol. 121, pp. 88–100, 2020.

[24] J. V. Arthur and K. Boahen, "Learning in Silicon: Timing is Everything," in *Proc. of Advances in Neural Information Processing Systems 18*, vol. 18. MIT Press, 2005.

[25] D. O. Hebb, *The organization of behavior; a neuropsychological theory*. Oxford, England: Wiley, 1949.

[26] N. Qiao, C. Bartolozzi, and G. Indiveri, "An Ultralow Leakage Synaptic Scaling Homeostatic Plasticity Circuit With Configurable Time Scales up to 100 ks," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 11, no. 6, pp. 1271–1277, 2017.

[27] F. J. Pelayo, B. Pino, P. A, J. Ortega, and F. J. Fernandez, "Cmos implementation of synapse matrices with programmable analog weights," in *Artificial Neural Networks*. Berlin, Heidelberg: Springer, 1991, pp. 307–314.

[28] L. Chua, "Memristor - The missing circuit element," *IEEE Transactions on Circuit Theory*, vol. 18, no. 5, pp. 507–519, 1971.

[29] D. B. Strukov, G. S. Snider, D. R. Stewart, and R. S. Williams, "The missing memristor found," *Nature*, vol. 453, no. 7191, pp. 80–83, 2008.

[30] N. Brunel and M. C. W. van Rossum, "Quantitative investigations of electrical nerve excitation treated as polarization," *Biological Cybernetics*, vol. 97, no. 5, pp. 341–349, 2007.

[31] N. Burkitt, "A Review of the Integrate-and-fire Neuron Model: I. Homogeneous Synaptic Input," *Biological Cybernetics*, vol. 95, no. 1, pp. 1–19, 2006.

[32] N. Fourcaud-Trocmé, D. Hansel, C. v. Vreeswijk, and N. Brunel, "How Spike Generation Mechanisms Determine the Neuronal Response to Fluctuating Inputs," *Journal of Neuroscience*, vol. 23, no. 37, pp. 11 628–11 640, 2003.

[33] R. Brette and W. Gerstner, "Adaptive Exponential Integrate-and-Fire Model as an Effective Description of Neuronal Activity," *Journal of Neurophysiology*, vol. 94, no. 5, pp. 3637–3642, 2005.

[34] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of Physiology*, vol. 117, no. 4, pp. 500–544, 1952.

[35] Y. v. d. Burgt and P. Gkoupidenis, "Organic materials and devices for brain-inspired computing: From artificial implementation to biophysical realism," *MRS Bulletin*, vol. 45, no. 8, pp. 631–640, 2020.

[36] C. Frenkel, J. Legat, and D. Bol, "A compact phenomenological digital neuron implementing the 20 Izhikevich behaviors," in *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2017, pp. 1–4.

[37] C. Mead, "Analog VLSI implementations of neural systems." New York NY, US: Springer, 1989.

[38] M. Mahowald and R. Douglas, "A silicon neuron," *Nature*, vol. 354, no. 6354, pp. 515–518, 1991.

[39] Z. Li, W. Tang, B. Zhang, R. Yang, and X. Miao, "Emerging memristive neurons for neuromorphic computing and sensing," *Science and Technology of Advanced Materials*, vol. 24, no. 1, p. 2188878, 2023.

[40] S. Moradi, N. Qiao, F. Stefanini, and G. Indiveri, "A Scalable Multicore Architecture With Heterogeneous Memory Structures for Dynamic Neuromorphic Asynchronous Processors (DYNAPs)," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 106–122, 2018.

[41] N. Qiao, H. Mostafa, F. Corradi, M. Osswald, F. Stefanini, D. Sumislawska, and G. Indiveri, "A reconfigurable on-line learning spiking neuromorphic processor comprising 256 neurons and 128K synapses," *Frontiers in Neuroscience*, vol. 9, 2015.

[42] J. Schemmel, D. Brüderle, A. Grübl, M. Hock, K. Meier, and S. Millner, "A wafer-scale neuromorphic hardware system for large-scale neural modeling," in *2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2010, pp. 1947–1950.

[43] B. V. Benjamin, P. Gao, E. McQuinn, S. Choudhary, A. R. Chandrasekaran, J.-M. Bussat, R. Alvarez-Icaza, J. V. Arthur, P. A. Merolla, and K. Boahen, "Neurogrid: A Mixed-Analog-Digital Multichip System for Large-Scale Neural Simulations," *Proceedings of the IEEE*, vol. 102, no. 5, pp. 699–716, 2014.

[44] F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G.-J. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. Jackson, and D. S. Modha, "TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 34, no. 10, pp. 1537–1557, 2015.

[45] M. Davies, "Benchmarks for progress in neuromorphic computing," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 386–388, 2019.

[46] C. Frenkel, M. Lefebvre, J.-D. Legat, and D. Bol, "A 0.086-mm2 12.7-pJ/SOP 64k-Synapse 256-Neuron Online-Learning Digital Spiking Neuromorphic Processor in 28-nm CMOS," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 1, pp. 145–158, 2019.

[47] C. Frenkel, J.-D. Legat, and D. Bol, "MorphIC: A 65-nm 738k-Synapse/mm^2 Quad-Core Binary-Weight Digital Neuromorphic Processor With Stochastic Spike-Driven Online Learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 999–1010, 2019, conference Name: IEEE Transactions on Biomedical Circuits and Systems.

[48] C. Frenkel, D. Bol, and G. Indiveri, "Bottom-Up and Top-Down Approaches for the Design of Neuromorphic Processing Systems: Tradeoffs and Synergies Between Natural and Artificial Intelligence," *Proceedings of the IEEE*, vol. 111, no. 6, pp. 623–652, 2023.

[49] P. Zhou, D.-U. Choi, J. K. Eshraghian, and S.-M. Kang, "A Fully Memristive Spiking Neural Network with Unsupervised Learning," 2022.

[50] Q. Duan, Z. Jing, X. Zou, Y. Wang, K. Yang, T. Zhang, S. Wu, R. Huang, and Y. Yang, "Spiking neurons with spatiotemporal dynamics and gain modulation for monolithically integrated memristive neural networks," *Nature Communications*, vol. 11, no. 1, p. 3399, 2020.

[51] M. Payvand, M. V. Nair, L. K. Müller, and G. Indiveri, "A neuromorphic systems approach to in-memory computing with non-ideal memristive devices: from mitigation to exploitation," *Faraday Discussions*, vol. 213, no. 0, pp. 487–510, 2019.

[52] E. Izhikevich, "Which Model to Use for Cortical Spiking Neurons?" *IEEE Transactions on Neural Networks*, vol. 15, no. 5, pp. 1063–1070, 2004.

[53] M. Dampfhoffer, T. Mesquida, A. Valentian, and L. Anghel, "Are SNNs Really More Energy-Efficient Than ANNs? an In-Depth Hardware-Aware Study," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 3, pp. 731–741, 2023.

[54] E. Lemaire, L. Cordone, A. Castagnetti, P.-E. Novac, J. Courtois, and B. Miramond, "An Analytical Estimation of Spiking Neural Networks Energy Efficiency," 2023, vol. 13623, pp. 574–587.

[55] A. Bhattacharjee, R. Yin, A. Moitra, and P. Panda, "Are SNNs Truly Energy-efficient? — A Hardware Perspective," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2024, pp. 13 311–13 315.

[56] S. Davidson and S. B. Furber, "Comparison of Artificial and Spiking Neural Networks on Digital Hardware," *Frontiers in Neuroscience*, vol. 15, p. 651141, 2021.

[57] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting Static Image Datasets to Spiking Neuromorphic Datasets Using Saccades," *Frontiers in Neuroscience*, vol. 9, p. 437, 2015.

[58] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[59] A. Krizhevsky and G. Hinton, "Learning Multiple Layers of Features from Tiny Images," University of Toronto, Tech. Rep., 2009.

[60] B. Cramer, Y. Stradmann, J. Schemmel, and F. Zenke, "The Heidelberg Spiking Data Sets for the Systematic Evaluation of Spiking Neural Networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 7, pp. 2744–2757, 2022.

[61] F. Zhou and Y. Chai, "Near-sensor and in-sensor computing," *Nature Electronics*, vol. 3, no. 11, pp. 664–671, 2020.

[62] L. Liu, W. Xu, Y. Ni, Z. Xu, B. Cui, J. Liu, H. Wei, and W. Xu, "Stretchable Neuromorphic Transistor That Combines Multisensing and Information Processing for Epidermal Gesture Recognition," *ACS Nano*, vol. 16, no. 2, pp. 2282–2291, 2022.

[63] J. Yu, X. Yang, G. Gao, Y. Xiong, Y. Wang, J. Han, Y. Chen, H. Zhang, Q. Sun, and Z. L. Wang, "Bioinspired mechano-photonic artificial synapse based on graphene/MoS2 heterostructure," *Science Advances*, vol. 7, no. 12, p. eabd9117, 2021.

[64] B. Mu, L. Guo, J. Liao, P. Xie, G. Ding, Z. Lv, Y. Zhou, S.-T. Han, and Y. Yan, "Near-Infrared Artificial Synapses for Artificial Sensory Neuron System," *Small*, vol. 17, no. 38, p. 2103837, 2021.

[65] M.-Z. Li, L.-C. Guo, G.-L. Ding, K. Zhou, Z.-Y. Xiong, S.-T. Han, and Y. Zhou, "Inorganic Perovskite Quantum Dot-Based Strain Sensors for Data Storage and In-Sensor Computing," *ACS Applied Materials & Interfaces*, vol. 13, no. 26, pp. 30 861–30 873, 2021.

[66] H. R. Lee, D. Lee, and J. H. Oh, "A Hippocampus-Inspired Dual-Gated Organic Artificial Synapse for Simultaneous Sensing of a Neurotransmitter and Light," *Advanced Materials*, vol. 33, no. 17, p. 2100119, 2021.

[67] Y. Chen, G. Gao, J. Zhao, H. Zhang, J. Yu, X. Yang, Q. Zhang, W. Zhang, S. Xu, J. Sun, Y. Meng, and Q. Sun, "Piezotronic Graphene Artificial Sensory Synapse," *Advanced Functional Materials*, vol. 29, no. 41, p. 1900959, 2019.

[68] G. Li, D. Xie, H. Zhong, Z. Zhang, X. Fu, Q. Zhou, Q. Li, H. Ni, J. Wang, E.-j. Guo, M. He, C. Wang, G. Yang, K. Jin, and C. Ge, "Photo-induced non-volatile VO2 phase transition for neuromorphic ultraviolet sensors," *Nature Communications*, vol. 13, no. 1, p. 1729, 2022.

[69] M. Almatrafi and K. Hirakawa, "DAViS Camera Optical Flow," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 396–407, 2020.

[70] M. A. Mahowald and C. Mead, "The Silicon Retina," *Scientific American*, vol. 264, no. 5, pp. 76–83, 1991.

[71] H. Abd and A. König, "Adaptive Spiking Sensor System Based on CMOS Memristors Emulating Long and Short-Term Plasticity of Biological Synapses for Industry 4.0 Applications," *Technisches Messen*, vol. 88, no. s1, pp. s114–s119, 2021.

[72] J. Singh and G. Kapur, "Design of CMOS based Neuromorphic Sensor Circuit Using Floating Gate MOSFET," in *2020 6th International Conference on Signal Processing and Communication (ICSC)*, 2020, pp. 248–251.

[73] T. Birkoben, H. Winterfeld, S. Fichtner, A. Petraru, and H. Kohlstedt, "A spiking and adapting tactile sensor for neuromorphic applications," *Scientific Reports*, vol. 10, no. 1, p. 17260, 2020.

[74] E. Beyne, "Chiplet Interconnect Technology: Piecing Together the Next Generation of Chips," 2024, accessed on 2024-07-02. [Online]. Available: https://www.3dincites.com/2024/07/chiplet-interconnect-technology-piecing-together-the-next-generation-of-chips/

[75] J. J. Wang, S. G. Hu, X. T. Zhan, Q. Yu, Z. Liu, T. P. Chen, Y. Yin, S. Hosaka, and Y. Liu, "Handwritten-Digit Recognition by Hybrid Convolutional Neural Network based on HfO2 Memristive Spiking-Neuron," *Scientific Reports*, vol. 8, no. 1, p. 12546, 2018.

[76] S. Lashkare, S. Chouhan, T. Chavan, A. Bhat, P. Kumbhare, and U. Ganguly, "PCMO RRAM for Integrate-and-Fire Neuron in Spiking Neural Networks," *IEEE Electron Device Letters*, vol. 39, no. 4, pp. 484–487, 2018.

[77] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou, "Stochastic phase-change neurons," *Nature Nanotechnology*, vol. 11, no. 8, pp. 693–699, 2016.

[78] C. D. Wright, P. Hosseini, and J. A. V. Diosdado, "Beyond von-Neumann Computing with Nanoscale Phase-Change Memory Devices," *Advanced Functional Materials*, vol. 23, no. 18, pp. 2248–2254, 2013.

[79] R. A. Cobley, H. Hayat, and C. D. Wright, "A self-resetting spiking phase-change neuron," *Nanotechnology*, vol. 29, no. 19, p. 195202, 2018.

[80] Z. Wang, M. Rao, J.-W. Han, J. Zhang, P. Lin, Y. Li, C. Li, W. Song, S. Asapu, R. Midya, Y. Zhuo, H. Jiang, J. H. Yoon, N. K. Upadhyay, S. Joshi, M. Hu, J. P. Strachan, M. Barnell, Q. Wu, H. Wu, Q. Qiu, R. S. Williams, Q. Xia, and J. J. Yang, "Capacitive neural network with neuro-transistors," *Nature Communications*, vol. 9, no. 1, p. 3208, 2018.

[81] X. Zhang, W. Wang, Q. Liu, X. Zhao, J. Wei, R. Cao, Z. Yao, X. Zhu, F. Zhang, H. Lv, S. Long, and M. Liu, "An Artificial Neuron Based on a Threshold Switching Memristor," *IEEE Electron Device Letters*, vol. 39, no. 2, pp. 308–311, 2018.

[82] Z. Wang, S. Joshi, S. E. Savel'ev, H. Jiang, R. Midya, P. Lin, M. Hu, N. Ge, J. P. Strachan, Z. Li, Q. Wu, M. Barnell, G.-L. Li, H. L. Xin, R. S. Williams, Q. Xia, and J. J. Yang, "Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing," *Nature Materials*, vol. 16, no. 1, pp. 101–108, 2017.

[83] B. Sengupta, M. Stemmler, S. B. Laughlin, and J. E. Niven, "Action Potential Energy Efficiency Varies Among Neuron Types in Vertebrates and Invertebrates," *PLOS Computational Biology*, vol. 6, no. 7, p. e1000840, 2010.

[84] M. D. Pickett, G. Medeiros-Ribeiro, and R. S. Williams, "A scalable neuristor built with Mott memristors," *Nature Materials*, vol. 12, no. 2, pp. 114–117, 2013.

[85] W. Yi, K. K. Tsang, S. K. Lam, X. Bai, J. A. Crowell, and E. A. Flores, "Biological plausibility and stochasticity in scalable VO2 active memristor neurons," *Nature Communications*, vol. 9, p. 4661, 2018.

[86] Y. Xu, S. Gao, Z. Li, R. Yang, and X. Miao, "Adaptive Hodgkin–Huxley Neuron for Retina-Inspired Perception," *Advanced Intelligent Systems*, vol. 4, no. 12, p. 2200210, 2022.

[87] Q. Wu, B. Dang, C. Lu, G. Xu, G. Yang, J. Wang, X. Chuai, N. Lu, D. Geng, H. Wang, and L. Li, "Spike Encoding with Optic Sensory Neurons Enable a Pulse Coupled Neural Network for Ultraviolet Image Segmentation," *Nano Letters*, vol. 20, no. 11, pp. 8015–8023, 2020.

[88] X. Zhang, Y. Zhuo, Q. Luo, Z. Wu, R. Midya, Z. Wang, W. Song, R. Wang, N. K. Upadhyay, Y. Fang, F. Kiani, M. Rao, Y. Yang, Q. Xia, Q. Liu, M. Liu, and J. J. Yang, "An artificial spiking afferent nerve based on Mott memristors for neurorobotics," *Nature Communications*, vol. 11, no. 1, p. 51, 2020.

[89] F. Li, R. Wang, C. Song, M. Zhao, H. Ren, S. Wang, K. Liang, D. Li, X. Ma, B. Zhu, H. Wang, and Y. Hao, "A Skin-Inspired Artificial Mechanoreceptor for Tactile Enhancement and Integration," *ACS Nano*, vol. 15, no. 10, pp. 16 422–16 431, 2021.

[90] J. Zhu, X. Zhang, R. Wang, M. Wang, P. Chen, L. Cheng, Z. Wu, Y. Wang, Q. Liu, and M. Liu, "A Heterogeneously Integrated Spiking Neuron Array for Multimode-Fused Perception and Object Classification," *Advanced Materials*, vol. 34, no. 24, p. 2200481, 2022.

[91] T. Rosca, F. Qaderi, and A. M. Ionescu, "High Tuning Range Spiking 1R-1T VO2 Voltage-Controlled Oscillator for Integrated RF and Optical Sensing," in *ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC)*, 2021, pp. 183–186.

[92] C. Y. Han, Z. R. Han, S. L. Fang, S. Q. Fan, J. Q. Yin, W. H. Liu, X. Li, S. Q. Yang, G. H. Zhang, X. L. Wang, and L. Geng, "Characterization and Modelling of Flexible VO2 Mott Memristor for the Artificial Spiking Warm Receptor," *Advanced Materials Interfaces*, vol. 9, no. 19, p. 2200394, 2022.

[93] F. Qaderi, T. Rosca, M. Burla, J. Leuthold, D. Flandre, and A. M. Ionescu, "Millimeter-wave to near-terahertz sensors based on reversible insulator-to-metal transition in VO2," *Communications Materials*, vol. 4, no. 1, pp. 1–12, 2023.

[94] A. Bonagiri, S. K. Das, C. V. Marquez, A. Rúa, E. Puyoo, S. K. Nath, D. Albertini, N. Baboux, M. Uenuma, R. G. Elliman, and S. K. Nandi, "Biorealistic Neuronal Temperature-Sensitive Dynamics within Threshold Switching Memristors: Toward Neuromorphic Thermosensation," *ACS Applied Materials & Interfaces*, vol. 16, no. 24, p. 31283–31293, 2024.

[95] S. L. Fang, C. Y. Han, W. Liu, Z. R. Han, B. Ma, Y. L. Cui, S. Q. Fan, X. Li, X. L. Wang, G. H. Zhang, J. Q. Yin, X. D. Huang, and L. Geng, "A bioinspired flexible artificial mechanoreceptor based on VO2 insulator-metal transition memristor," *Journal of Alloys and Compounds*, vol. 911, p. 165096, 2022.

[96] R. Yuan, Q. Duan, P. J. Tiw, G. Li, Z. Xiao, Z. Jing, K. Yang, C. Liu, C. Ge, R. Huang, and Y. Yang, "A calibratable sensory neuron based on epitaxial VO2 for spike-based neuromorphic multisensory system," *Nature Communications*, vol. 13, no. 1, p. 3973, 2022.

[97] R. Yuan, P. J. Tiw, L. Cai, Z. Yang, C. Liu, T. Zhang, C. Ge, R. Huang, and Y. Yang, "A neuromorphic physiological signal processing system based on VO2 memristor for next-generation human-machine interface," *Nature Communications*, vol. 14, no. 1, p. 3695, 2023.

[98] Y. Pei, L. Yan, Z. Wu, J. Lu, J. Zhao, J. Chen, Q. Liu, and X. Yan, "Artificial Visual Perception Nervous System Based on Low-Dimensional Material Photoelectric Memristors," *ACS Nano*, vol. 15, no. 11, pp. 17 319–17 326, 2021.

[99] Y. Wang, Y. Gong, S. Huang, X. Xing, Z. Lv, J. Wang, J.-Q. Yang, G. Zhang, Y. Zhou, and S.-T. Han, "Memristor-based biomimetic compound eye for real-time collision detection," *Nature Communications*, vol. 12, no. 1, p. 5979, 2021.

[100] K. Shi, S. Heng, X. Wang, S. Liu, H. Cui, C. Chen, Y. Zhu, W. Xu, C. Wan, and Q. Wan, "An Oxide Based Spiking Thermoreceptor for Low-Power Thermography Edge Detection," *IEEE Electron Device Letters*, vol. 43, no. 12, pp. 2196–2199, 2022.

[101] T. Fu, X. Liu, S. Fu, T. Woodard, H. Gao, D. R. Lovley, and J. Yao, "Self-sustained green neuromorphic interfaces," *Nature Communications*, vol. 12, no. 1, p. 3351, 2021.

[102] C. Chen, Y. He, H. Mao, L. Zhu, X. Wang, Y. Zhu, Y. Zhu, Y. Shi, C. Wan, and Q. Wan, "A Photoelectric Spiking Neuron for Visual Depth Perception," *Advanced Materials*, vol. 34, no. 20, p. 2201895, 2022.

[103] R. Yuan, P. J. Tiw, L. Cai, Z. Yang, C. Liu, T. Zhang, C. Ge, R. Huang, and Y. Yang, "A neuromorphic physiological signal processing system based on VO2 memristor for next-generation human-machine interface," *Nature Communications*, vol. 14, no. 1, p. 3695, 2023.

[104] Y. Ke, S. Wang, G. Liu, M. Li, T. J. White, and Y. Long, "Vanadium Dioxide: The Multistimuli Responsive Material and Its Applications," *Small*, vol. 14, no. 39, p. 1802025, 2018.

[105] P. Shvets, O. Dikaya, K. Maksimova, and A. Goikhman, "A review of Raman spectroscopy of vanadium oxides," *Journal of Raman Spectroscopy*, vol. 50, no. 8, pp. 1226–1244, 2019.

[106] P. Hu, P. Hu, T. D. Vu, M. Li, S. Wang, Y. Ke, X. Zeng, L. Mai, and Y. Long, "Vanadium Oxide: Phase Diagrams, Structures, Synthesis, and Applications," *Chemical Reviews*, vol. 123, no. 8, pp. 4353–4415, 2023.

[107] F. J. Morin, "Oxides Which Show a Metal-to-Insulator Transition at the Neel Temperature," *Physical Review Letters*, vol. 3, no. 1, pp. 34–36, 1959.

[108] A. Heidemann, K. Kosuge, Y. Ueda, and S. Kachi, "Hyperfine interaction in V3O7," *physica status solidi (a)*, vol. 39, no. 1, pp. K37–K40, 1977.

[109] U. Schwingenschlögl and V. Eyert, "The vanadium Magnéli phases VO2-1," *Annalen der Physik*, vol. 516, no. 9, pp. 475–510, 2004.

[110] G. S. Nadkarni and V. S. Shirodkar, "Experiment and theory for switching in Al/V2O5/Al devices," *Thin Solid Films*, vol. 105, no. 2, pp. 115–129, 1983.

[111] H. Kappert, N. Kordas, S. Dreiner, U. Paschen, and R. Kokozinski, "High temperature SOI CMOS technology and circuit realization for applications up to 300°C," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1162–1165.

[112] D. Wegkamp and J. Stähler, "Ultrafast dynamics during the photoinduced phase transition in VO2," *Progress in Surface Science*, vol. 90, no. 4, pp. 464–502, 2015.

[113] Z. Shao, X. Cao, H. Luo, and P. Jin, "Recent progress in the phase-transition mechanism and modulation of vanadium dioxide materials," *NPG Asia Materials*, vol. 10, no. 7, pp. 581–605, 2018.

[114] K. Liu, S. Lee, S. Yang, O. Delaire, and J. Wu, "Recent progresses on physics and applications of vanadium dioxide," *Materials Today*, vol. 21, no. 8, pp. 875–896, 2018.

[115] J. B. Goodenough, "The two components of the crystallographic transition in VO2," *Journal of Solid State Chemistry*, vol. 3, no. 4, pp. 490–500, 1971.

[116] A. Zylbersztejn and N. F. Mott, "Metal-insulator transition in vanadium dioxide," *Physical Review B*, vol. 11, no. 11, pp. 4383–4395, 1975.

[117] A. Cavalleri, T. Dekorsy, H. H. W. Chong, J. C. Kieffer, and R. W. Schoenlein, "Evidence for a structurally-driven insulator-to-metal transition in VO2: A view from the ultrafast timescale," *Physical Review B*, vol. 70, no. 16, p. 161102, 2004.

[118] T. Yao, X. Zhang, Z. Sun, S. Liu, Y. Huang, Y. Xie, C. Wu, X. Yuan, W. Zhang, Z. Wu, G. Pan, F. Hu, L. Wu, Q. Liu, and S. Wei, "Understanding the Nature of the Kinetic Process in a VO2 Metal-Insulator Transition," *Physical Review Letters*, vol. 105, no. 22, p. 226405, 2010.

[119] M. M. Qazilbash, M. Brehm, B.-G. Chae, P.-C. Ho, G. O. Andreev, B.-J. Kim, S. J. Yun, A. V. Balatsky, M. B. Maple, F. Keilmann, H.-T. Kim, and D. N. Basov, "Mott Transition in VO2 Revealed by Infrared Spectroscopy and Nano-Imaging," *Science*, vol. 318, no. 5857, pp. 1750–1753, 2007.

[120] J. Lin, K. Alam, L. Ocola, Z. Zhang, S. Datta, S. Ramanathan, and S. Guha, "Physics and technology of electronic insulator-to-metal transition (E-IMT) for record high on/off ratio and low voltage in device applications," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2017, pp. 23.4.1–23.4.4.

[121] H. Madan, M. Jerry, A. Pogrebnyakov, T. Mayer, and S. Datta, "Quantitative Mapping of Phase Coexistence in Mott-Peierls Insulator during Electronic and Thermally Driven Phase Transition," *ACS Nano*, vol. 9, no. 2, pp. 2009–2017, 2015.

[122] J. Jeong, N. Aetukuri, T. Graf, T. D. Schladt, M. G. Samant, and S. S. P. Parkin, "Suppression of Metal-Insulator Transition in VO2 by Electric Field–Induced Oxygen Vacancy Formation," *Science*, vol. 339, no. 6126, pp. 1402–1405, 2013.

[123] G. Kovacs, D. Buerger, I. Skorupa, H. Reuther, R. Heller, and H. Schmidt, "Effect of the substrate on the insulator-metal transition of vanadium dioxide films," *Journal of Applied Physics*, vol. 109, 2011.

[124] D. Brassard, S. Fourmaux, M. Jean-Jacques, J. C. Kieffer, and M. A. El Khakani, "Grain size effect on the semiconductor-metal phase transition characteristics of magnetron-sputtered VO2 thin films," *Applied Physics Letters*, vol. 87, no. 5, p. 051910, 2005.

[125] K. M. Niang, G. Bai, H. Lu, and J. Robertson, "Microstructure scaling of metal–insulator transition properties of VO2 films," *Applied Physics Letters*, vol. 118, no. 12, p. 121901, 2021.

[126] J. Zhang, Z. Zhao, J. Li, H. Jin, F. Rehman, P. Chen, Y. Jiang, C. Chen, M. Cao, and Y. Zhao, "Evolution of Structural and Electrical Properties of Oxygen-Deficient VO2 under Low Temperature Heating Process," *ACS Applied Materials & Interfaces*, vol. 9, no. 32, pp. 27 135–27 141, 2017.

[127] S. A. Bukhari, S. Kumar, P. Kumar, S. P. Gumfekar, H.-J. Chung, T. Thundat, and A. Goswami, "The effect of oxygen flow rate on metal–insulator transition (MIT) characteristics of vanadium dioxide (VO2) thin films by pulsed laser deposition (PLD)," *Applied Surface Science*, vol. 529, p. 146995, 2020.

[128] L. Fan, X. Wang, F. Wang, Q. Zhang, L. Zhu, Q. Meng, B. Wang, Z. Zhang, and C. Zou, "Revealing the role of oxygen vacancies on the phase transition

of VO 2 film from the optical-constant measurements," *RSC Advances*, vol. 8, pp. 19 151–19 156, 2018.

[129] C. Chen, Y. Zhao, X. Pan, V. Kuryatkov, A. Bernussi, M. Holtz, and Z. Fan, "Influence of defects on structural and electrical properties of VO2 thin films," *Journal of Applied Physics*, vol. 110, no. 2, p. 023707, 2011.

[130] Z. Zhang, F. Zuo, C. Wan, A. Dutta, J. Kim, J. Rensberg, R. Nawrodt, H. H. Park, T. J. Larrabee, X. Guan, Y. Zhou, S. Prokes, C. Ronning, V. M. Shalaev, A. Boltasseva, M. A. Kats, and S. Ramanathan, "Evolution of Metallicity in Vanadium Dioxide by Creation of Oxygen Vacancies," *Physical Review Applied*, vol. 7, no. 3, p. 034008, 2017.

[131] K. Appavoo, D. Y. Lei, Y. Sonnefraud, B. Wang, S. T. Pantelides, S. A. Maier, and R. F. J. Haglund, "Role of Defects in the Phase Transition of VO2 Nanoparticles Probed by Plasmon Resonance Spectroscopy," *Nano Letters*, vol. 12, no. 2, pp. 780–786, 2012.

[132] A. Krammer, A. Magrez, W. A. Vitale, P. Mocny, P. Jeanneret, E. Guibert, H. J. Whitlow, A. M. Ionescu, and A. Schüler, "Elevated transition temperature in Ge doped VO2 thin films," *Journal of Applied Physics*, vol. 122, no. 4, p. 045304, 2017.

[133] Z. Yu, Y. Liu, Z. Zhang, M. Cheng, Z. Zou, Z. Lu, D. Wang, J. Shi, and R. Xiong, "Controllable phase transition temperature by regulating interfacial strain of epitaxial VO2 films," *Ceramics International*, vol. 46, no. 8, Part B, pp. 12 393–12 399, 2020.

[134] M.-W. Kim, W. Jung, Hyun-Cho, T.-S. Bae, S.-J. Chang, J.-H. Jang, W.-K. Hong, and B.-J. Kim, "Substrate-mediated strain effect on the role of thermal heating and electric field on metal-insulator transition in vanadium dioxide nanobeams," *Scientific Reports*, vol. 5, p. 1, 2015.

[135] J. Jian, W. Zhang, C. Jacob, A. Chen, H. Wang, J. Huang, and H. Wang, "Roles of grain boundaries on the semiconductor to metal phase transition of VO2 thin films," *Applied Physics Letters*, vol. 107, no. 10, p. 102105, 2015.

[136] W. Lu, L.-M. Wong, S. Wang, and K. Zeng, "Local phenomena at grain boundaries: An alternative approach to grasp the role of oxygen vacancies in metallization of VO2," *Journal of Materiomics*, vol. 4, no. 4, pp. 360–367, 2018.

[137] L. Wang, Z. Shao, Q. Li, J. Liu, C. Yang, P. Jin, and X. Cao, "Dynamic Modulation of Phase Transition by External Strain Engineering in Quasi-van der Waals Epitaxial VO2 Films on Fluorophlogopite," *Advanced Materials Interfaces*, vol. 9, no. 29, p. 2200864, 2022.

[138] T. Kikuzuki and M. Lippmaa, "Characterizing a strain-driven phase transition in VO2," *Applied Physics Letters*, vol. 96, no. 13, p. 132107, 2010.

[139] R. M. Bowman and J. M. Gregg, "VO2 thin films: growth and the effect of applied strain on their resistance," *Journal of Materials Science: Materials in Electronics*, vol. 9, no. 3, pp. 187–191, 1998.

[140] H.-T. Zhang, L. Guo, G. Stone, L. Zhang, Y.-X. Zheng, E. Freeman, D. W. Keefer, S. Chaudhuri, H. Paik, J. A. Moyer, M. Barth, D. G. Schlom, J. V. Badding, S. Datta, V. Gopalan, and R. Engel-Herbert, "Imprinting of Local Metallic States into VO2 with Ultraviolet Light," *Advanced Functional Materials*, vol. 26, no. 36, pp. 6612–6618, 2016.

[141] H. Navarro, J. del Valle, Y. Kalcheim, N. M. Vargas, C. Adda, M.-H. Lee, P. Lapa, A. Rivera-Calzada, I. A. Zaluzhnyy, E. Qiu, O. Shpyrko, M. Rozenberg, A. Frano, and I. K. Schuller, "A hybrid optoelectronic Mott insulator," *Applied Physics Letters*, vol. 118, no. 14, p. 141901, 2021.

[142] M. Nakano, K. Shibuya, D. Okuyama, T. Hatano, S. Ono, M. Kawasaki, Y. Iwasa, and Y. Tokura, "Collective bulk carrier delocalization driven by electrostatic surface charge accumulation," *Nature*, vol. 487, no. 7408, pp. 459–462, 2012.

[143] Z. Yang, S. Hart, C. Ko, A. Yacoby, and S. Ramanathan, "Studies on electric triggering of the metal-insulator transition in VO2 thin films between 77 K and 300 K," *Journal of Applied Physics*, vol. 110, no. 3, p. 033725, 2011.

[144] Z. Li, Y. Guo, Z. Hu, J. Su, J. Zhao, J. Wu, J. Wu, Y. Zhao, C. Wu, and Y. Xie, "Hydrogen Treatment for Superparamagnetic VO2 Nanowires with Large Room-Temperature Magnetoresistance," *Angewandte Chemie International Edition*, vol. 55, no. 28, pp. 8018–8022, 2016.

[145] W. W. Li, J. J. Zhu, J. R. Liang, Z. G. Hu, J. Liu, H. D. Chen, and J. H. Chu, "External Electric Field Manipulations on Structural Phase Transition of Vanadium Dioxide Nanoparticles and Its Application in Field Effect Transistor," *The Journal of Physical Chemistry C*, vol. 115, no. 47, pp. 23 558–23 563, 2011.

[146] S. H. Lee, M. K. Kim, J. W. Lee, Z. Yang, S. Ramanathan, and S. Tiwari, "Vanadium dioxide (VO2) is also a ferroelectric: Properties from memory structures," in *2011 11th IEEE International Conference on nanotechnology*, 2011, pp. 735–739.

[147] C. Ko and S. Ramanathan, "Effect of ultraviolet irradiation on electrical resistance and phase transition characteristics of thin film vanadium oxide," *Journal of Applied Physics*, vol. 103, pp. 106 104–106 104, 2008.

[148] J. Cao, E. Ertekin, V. Srinivasan, W. Fan, S. Huang, H. Zheng, J. W. L. Yim, D. R. Khanal, D. F. Ogletree, J. C. Grossman, and J. Wu, "Strain engineering and one-dimensional organization of metal–insulator domains in single-crystal vanadium dioxide beams," *Nature nanotechnology*, vol. 4, no. 11, pp. 732–737, 2009.

[149] M. F. Becker, A. B. Buckman, R. M. Walser, T. Lépine, P. Georges, and A. Brun, "Femtosecond laser excitation of the semiconductor-metal phase transition in VO2," *Applied Physics Letters*, vol. 65, no. 12, pp. 1507–1509, 1994.

[150] A. Cavalleri, C. Tóth, C. W. Siders, J. A. Squier, F. Ráksi, P. Forget, and J. C. Kieffer, "Femtosecond Structural Dynamics in VO2 during an Ultrafast Solid-Solid Phase Transition," *Physical Review Letters*, vol. 87, no. 23, p. 237401, 2001.

[151] W. R. Roach and I. Balberg, "Optical induction and detection of fast phase transition in VO2," *Solid State Communications*, vol. 9, no. 9, pp. 551–555, 1971.

[152] Y. H. Matsuda, D. Nakamura, A. Ikeda, S. Takeyama, Y. Suga, H. Nakahara, and Y. Muraoka, "Magnetic-field-induced insulator–metal transition in W-doped VO2 at 500 T," *Nature Communications*, vol. 11, no. 1, p. 3591, 2020.

[153] A. Crunteanu, J. Givernaud, J. Leroy, D. Mardivirin, C. Champeaux, J.-C. Orlianges, A. Catherinot, and P. Blondy, "Voltage- and current-activated metal–insulator transition in VO2-based electrical switches: a lifetime operation analysis," *Science and Technology of Advanced Materials*, vol. 11, no. 6, p. 065002, 2010.

[154] G. Stefanovich, A. Pergament, and D. Stefanovich, "Electrical switching and Mott transition in VO2," *Journal of Physics: Condensed Matter*, vol. 12, no. 41, pp. 8837–8845, 2000.

[155] Y. Zhou, X. Chen, C. Ko, Z. Yang, C. Mouli, and S. Ramanathan, "Voltage-Triggered Ultrafast Phase Transition in Vanadium Dioxide Switches," *IEEE Electron Device Letters*, vol. 34, no. 2, pp. 220–222, 2013.

[156] H.-T. Kim, B.-G. Chae, D.-H. Youn, S.-L. Maeng, G. Kim, K.-Y. Kang, and Y.-S. Lim, "Mechanism and observation of Mott transition in VO2-based two- and three-terminal devices," *New Journal of Physics*, vol. 6, no. 1, p. 52, 2004.

[157] A. Joushaghani, J. Jeong, S. Paradis, D. Alain, J. Stewart Aitchison, and J. K. S. Poon, "Voltage-controlled switching and thermal effects in VO2 nano-gap junctions," *Applied Physics Letters*, vol. 104, no. 22, p. 221904, 2014.

[158] J. Leroy, A. Crunteanu, A. Bessaudou, F. Cosset, C. Champeaux, and J.-C. Orlianges, "High-speed metal-insulator transition in vanadium dioxide films induced by an electrical pulsed voltage over nano-gap electrodes," *Applied Physics Letters*, vol. 100, no. 21, p. 213507, 2012.

[159] E. Corti, "Networks of Coupled VO2 Oscillators for Neuromorphic Computing," PhD dissertation, EPFL, Lausanne, 2021.

[160] G. A. Gibson, "Designing Negative Differential Resistance Devices Based on Self-Heating," *Advanced Functional Materials*, vol. 28, no. 22, p. 1704175, 2018.

[161] A. Zimmers, L. Aigouy, M. Mortier, A. Sharoni, S. Wang, K. G. West, J. G. Ramirez, and I. K. Schuller, "Role of Thermal Heating on the Voltage Induced

Insulator-Metal Transition in VO2," *Physical Review Letters*, vol. 110, no. 5, p. 056601, 2013.

[162] X. Zhong, X. Zhang, A. Gupta, and P. LeClair, "Avalanche breakdown in microscale VO2 structures," *Journal of Applied Physics*, vol. 110, no. 8, p. 084516, 2011.

[163] O. Maher, R. Bernini, N. Harnack, B. Gotsmann, M. Sousa, V. Bragaglia, and S. Karg, "Highly reproducible and CMOS-compatible VO2-based oscillators for brain-inspired computing," *Scientific Reports*, vol. 14, no. 1, p. 11600, 2024.

[164] J. del Valle, N. M. Vargas, R. Rocco, P. Salev, Y. Kalcheim, P. N. Lapa, C. Adda, M.-H. Lee, P. Y. Wang, L. Fratino, M. J. Rozenberg, and I. K. Schuller, "Spatiotemporal characterization of the field-induced insulator-to-metal transition," *Science*, vol. 373, no. 6557, pp. 907–911, 2021.

[165] P. Stoliar, L. Cario, E. Janod, B. Corraze, C. Guillot-Deudon, S. Salmon-Bourmand, V. Guiot, J. Tranchant, and M. Rozenberg, "Universal Electric-Field-Driven Resistive Transition in Narrow-Gap Mott Insulators," *Advanced Materials*, vol. 25, no. 23, pp. 3222–3226, 2013.

[166] S. K. Nandi, S. K. Das, C. Estherby, A. Gentle, and R. G. Elliman, "Understanding modes of negative differential resistance in amorphous and polycrystalline vanadium oxides," *Journal of Applied Physics*, vol. 128, no. 24, p. 244103, 2020.

[167] F. Qaderi, T. Rosca, M. Burla, J. Leuthold, D. Flandre, and A. Ionescu, "Millimeter- and Terahertz-wave stochastic sensors based on reversible insulator-to-metal transition in vanadium dioxide," In Review, preprint, 2021.

[168] D. K. Lee, S. Lee, H. Sim, Y. Park, S.-Y. Choi, and J. Son, "Piezo strain–controlled phase transition in single-crystalline Mott switches for threshold-manipulated leaky integrate-and-fire neurons," *Science Advances*, vol. 10, no. 14, p. eadk8836, 2024.

[169] T. Rosca, F. Qaderi, M. Riccardi, O. J. Martin, and A. M. Ionescu, "An Experimental Study of the Photoresponse of 1T-1R Oscillators Based on Vanadium Dioxide: Towards Spiking Sensing Systems," in *2021 21st International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers)*, 2021, pp. 373–376.

[170] J. M. Wu and L. B. Liou, "Room temperature photo-induced phase transitions of VO2 nanodevices," *Journal of Materials Chemistry*, vol. 21, no. 14, p. 5499, 2011.

[171] J. M. Wu and W. E. Chang, "Ultrahigh Responsivity and External Quantum Efficiency of an Ultraviolet-Light Photodetector Based on a Single VO2 Microwire," *ACS Applied Materials & Interfaces*, vol. 6, no. 16, pp. 14 286–14 292, 2014.

[172] M. J. Tadjer, V. D. Wheeler, B. P. Downey, Z. R. Robinson, D. J. Meyer, C. R. Eddy, and F. J. Kub, "Temperature and electric field induced metal-insulator transition in atomic layer deposited VO2 thin films," *Solid-State Electronics*, vol. 136, pp. 30–35, 2017.

[173] L. Niu, P. Xu, T. Huang, W. Yang, Z. Chen, X. Chen, and N. Dai, "Defect-Driven Light Perception and Memristor Storage with Phase Transition in Vanadium Dioxide," *Advanced Electronic Materials*, p. 2400006.

[174] L. Wang, L. Wang, X.-Y. Ye, X.-H. Xu, L.-Y. Shang, Y.-W. Li, J.-Z. Zhang, L.-Q. Zhu, and Z.-G. Hu, "Thermally modulated photoelectronic synaptic behavior in HfS2/VO2 heterostructure," *Rare Metals*, 2024.

[175] X. Yu, C. Cheng, J. Liang, M. Wang, B. Huang, Z. Wang, and L. Li, "Graphene-Assisting Nonvolatile Vanadium Dioxide Phase Transition for Neuromorphic Machine Vision," *Advanced Functional Materials*, vol. 34, no. 16, p. 2312481, 2024.

[176] F. Guo, Y. Liu, M. Zhang, W. Yu, S. Li, B. Zhang, B. Hu, S. Li, A. Sun, J. Jiang, and L. Hao, "VO2/MoO3 Heterojunctions Artificial Optoelectronic Synapse Devices for Near-Infrared Optical Communication," *Small*, vol. n/a, no. n/a, p. 2310767.

[177] X. Deng, S.-Q. Wang, Y.-X. Liu, N. Zhong, Y.-H. He, H. Peng, P.-H. Xiang, and C.-G. Duan, "A Flexible Mott Synaptic Transistor for Nociceptor Simulation and Neuromorphic Computing," *Advanced Functional Materials*, vol. 31, no. 23, p. 2101099, 2021.

[178] J. Park, C. Oh, and J. Son, "Anisotropic ionic transport-controlled synaptic weight update by protonation in a VO2 transistor," *Journal of Materials Chemistry C*, vol. 9, no. 7, pp. 2521–2529, 2021.

[179] C. Oh, I. Kim, J. Park, Y. Park, M. Choi, and J. Son, "Deep Proton Insertion Assisted by Oxygen Vacancies for Long-Term Memory in VO2 Synaptic Transistor," *Advanced Electronic Materials*, vol. 7, no. 2, p. 2000802, 2021.

[180] S. K. Nath, S. K. Das, S. K. Nandi, C. Xi, C. V. Marquez, A. Rúa, M. Uenuma, Z. Wang, S. Zhang, R.-J. Zhu, J. Eshraghian, X. Sun, T. Lu, Y. Bian, N. Syed, W. Pan, H. Wang, W. Lei, L. Fu, L. Faraone, Y. Liu, and R. G. Elliman, "Optically Tunable Electrical Oscillations in Oxide-Based Memristors for Neuromorphic Computing," *Advanced Materials*, vol. 36, no. 25, p. 2400904, 2024.

[181] Q. Duan, T. Zhang, C. Liu, R. Yuan, G. Li, P. Jun Tiw, K. Yang, C. Ge, Y. Yang, and R. Huang, "Artificial Multisensory Neurons with Fused Haptic and Temperature Perception for Multimodal In-Sensor Computing," *Advanced Intelligent Systems*, vol. 4, no. 8, p. 2200039, 2022.

[182] Y. Wang, C. Su, Y. Zheng, K. Zhou, Z. Wen, Y. Fu, Q. Wang, and D. He, "High consistency VO2 memristor for artificial auditory neuron," *Microelectronic Engineering*, vol. 282, p. 112101, 2023.

[183] C. Zhang, C. Koughia, O. Güneş, J. Luo, N. Hossain, Y. Li, X. Cui, S.-J. Wen, R. Wong, Q. Yang, and S. Kasap, "Synthesis, structure and optical properties of high-quality VO2 thin films grown on silicon, quartz and sapphire substrates by high temperature magnetron sputtering: Properties through the transition temperature," *Journal of Alloys and Compounds*, vol. 848, p. 156323, 2020.

[184] J. Lin, Annadi, S. Sonde, C. Chen, L. Stan, K. Achari, S. Ramanathan, and S. Guha, "Low-voltage artificial neuron using feedback engineered insulator-to-metal-transition devices," in *2016 IEEE International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE, 2016, pp. 34.5.1–34.5.4.

[185] L. V. Yakovkina, S. V. Mutilin, V. Y. Prinz, T. P. Smirnova, V. R. Shayapov, I. V. Korol'kov, E. A. Maksimovsky, and N. D. Volchok, "MOCVD growth and characterization of vanadium dioxide films," *Journal of Materials Science*, vol. 52, no. 7, pp. 4061–4069, 2017.

[186] B. Chae, D.-H. Youn, H.-T. Kim, M. Sunglyul, and K. Kang, "Fabrication and Electrical Properties of Pure VO2 Phase Films," *Journal of the Korean Physical Society*, vol. 44, 2003.

[187] D. H. Jung, H. S. So, K. H. Ko, J. W. Park, H. Lee, T. T. T. Nguyen, and S. Yoon, "Electrical and optical properties of VO2 thin films grown on various sapphire substrates by using RF sputtering deposition," *Journal of the Korean Physical Society*, vol. 69, no. 12, pp. 1787–1797, 2016.

[188] T. Rosca, "Energy Efficient Sensing using Steep Slope Devices," Ph.D. dissertation, EPFL, Lausanne, 2022.

[189] J. del Valle, P. Salev, Y. Kalcheim, and I. K. Schuller, "A caloritronics-based Mott neuristor," *Scientific Reports*, vol. 10, no. 1, p. 4292, 2020.

[190] R. Chodun, B. Wicher, L. Skowrnski, K. Nowakowska-Langier, S. Okrasa, A. Grabowski, R. Minikayev, and K. Zdunek, "Multi-sided metallization of textile fibres by using magnetron system with grounded cathode," *Materials Science-Poland*, vol. 35, no. 3, pp. 639–646, 2017.

[191] R. T. R. Kumar, B. Karunagaran, D. Mangalaraj, S. K. Narayandass, P. Manoravi, and M. Joseph, "Characteristics of amorphous VO2 thin films prepared by pulsed laser deposition," *Journal of Materials Science*, vol. 39, no. 8, pp. 2869–2871, 2004.

[192] I. Safi, "Recent aspects concerning DC reactive magnetron sputtering of thin films: a review," *Surface and Coatings Technology*, vol. 127, no. 2, pp. 203–218, 2000.

[193] Q. Van Overmeere and S. Ramanathan, "Thin film fuel cells with vanadium oxide anodes: Strain and stoichiometry effects," *Electrochimica Acta*, vol. 150, pp. 83–88, 2014.

[194] S. A. Hassanzadeh-Tabrizi, "Precise calculation of crystallite size of nanomaterials: A review," *Journal of Alloys and Compounds*, vol. 968, p. 171914, 2023.

[195] J. E. Greene, "Chapter 12 - Thin Film Nucleation, Growth, and Microstructural Evolution: An Atomic Scale View," in *Handbook of Deposition Technologies for Films and Coatings (Third Edition)*. Boston: William Andrew Publishing, 2010, pp. 554–620.

[196] M. A. Herman, W. Richter, and H. Sitter, *Epitaxy: Physical Principles and Technical Implementation*. Berlin, Heidelberg: Springer, 2004, vol. 62.

[197] Q. Lu, C. Sohn, G. Hu, X. Gao, M. F. Chisholm, I. Kylänpää, J. T. Krogel, P. R. C. Kent, O. Heinonen, P. Ganesh, and H. N. Lee, "Metal–insulator transition tuned by oxygen vacancy migration across TiO2/VO2 interface," *Scientific Reports*, vol. 10, no. 1, p. 18554, 2020.

[198] Y. Liu, J. Liu, Y. Li, D. Wang, L. Ren, and K. Zou, "Effect of annealing temperature on the structure and properties of vanadium oxide films," *Optical Materials Express*, vol. 6, no. 5, p. 1552, 2016.

[199] H.-T. Zhang, L. Zhang, D. Mukherjee, Y.-X. Zheng, R. C. Haislmaier, N. Alem, and R. Engel-Herbert, "Wafer-scale growth of VO2 thin films using a combinatorial approach," *Nature Communications*, vol. 6, no. 1, p. 8475, 2015.

[200] J. Hu, X. Wang, J. Zhang, J. Luo, Z. Zhang, and Z. Shen, "A general mechanism of grain growth -I. Theory," *Journal of Materiomics*, vol. 7, no. 5, pp. 1007–1013, 2021.

[201] M. Kumar, J. P. Singh, K. H. Chae, J. Park, and H. H. Lee, "Annealing effect on phase transition and thermochromic properties of VO2 thin films," *Superlattices and Microstructures*, vol. 137, p. 106335, 2020.

[202] H. Ma, X. Xiao, Y. Wang, Y. Sun, B. Wang, X. Gao, E. Wang, K. Jiang, K. Liu, and X. Zhang, "Wafer-scale freestanding vanadium dioxide film," *Science Advances*, vol. 7, no. 50, p. eabk3438.

[203] E. Corti, B. Gotsmann, K. Moselund, I. Stolichnov, A. Ionescu, and S. Karg, "Resistive Coupled VO2 Oscillators for Image Recognition," in *2018 IEEE International Conference on Rebooting Computing (ICRC)*, 2018, pp. 1–7.

[204] A. Ainabayev, D. Mullarkey, B. Walls, D. Caffrey, K. Zhussupbekov, A. Zhussupbekova, C. Ilhan, A. Kaisha, P. Biswas, A. Tikhonov, O. Murtagh, and I. Shvets, "Epitaxial Grown VO2 with Suppressed Hysteresis and Low Room Temperature Resistivity for High-Performance Thermal Sensor Applications," *ACS Applied Nano Materials*, vol. 6, no. 4, pp. 2917–2927, 2023.

[205] M. J. Miller and J. Wang, "Influence of grain size on transition temperature of thermochromic VO2," *Journal of Applied Physics*, vol. 117, no. 3, p. 034307, 2015.

[206] B. Bhar, A. Khanna, A. Parihar, S. Datta, and A. Raychowdhury, "Stochastic Resonance in Insulator-Metal-Transition Systems," *Scientific Reports*, vol. 10, no. 1, p. 5549, 2020.

[207] M. Jerry, K. Ni, A. Parihar, A. Raychowdhury, and S. Datta, "Stochastic Insulator-to-Metal Phase Transition-Based True Random Number Generator," *IEEE Electron Device Letters*, vol. 39, no. 1, pp. 139–142, 2018.

[208] J. d. Valle, P. Salev, S. Gariglio, Y. Kalcheim, I. K. Schuller, and J.-M. Triscone, "Generation of Tunable Stochastic Sequences Using the Insulator–Metal Transition," *Nano Letters*, vol. 22, no. 3, pp. 1251–1256, 2022.

[209] E. Qiu, P. Salev, F. Torres, H. Navarro, R. C. Dynes, and I. K. Schuller, "Stochastic transition in synchronized spiking nanooscillators," *Proceedings of the National Academy of Sciences*, vol. 120, no. 38, p. e2303765120, 2023.

[210] L. Bao, Z. Wang, B. Wang, K. Liu, G. Bai, Z. Yu, J. Kang, Y. Ling, L. Wu, Q. Chen, K. Niang, Y. Cai, J. Robertson, and R. Huang, "Tunable Stochastic Oscillator Based on Hybrid VO2/TaOx Device for Compressed Sensing," *IEEE Electron Device Letters*, vol. 42, pp. 102–105, 2020.

[211] K. Zhang, B. Wang, F. Wang, Y. Han, X. Jian, H. Zhang, and H. S. P. Wong, "VO2-Based Selection Device for Passive Resistive Random Access Memory Application," *IEEE Electron Device Letters*, vol. 37, no. 8, pp. 978–981, 2016.

[212] G. A. Fonseca Guerra and S. B. Furber, "Using Stochastic Spiking Neural Networks on SpiNNaker to Solve Constraint Satisfaction Problems," *Frontiers in Neuroscience*, vol. 0, 2017.

[213] L. Buesing, J. Bill, B. Nessler, and W. Maass, "Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons," *PLOS Computational Biology*, vol. 7, no. 11, p. e1002211, 2011.

[214] K. Ahmed, A. Shrestha, Q. Qiu, and Q. Wu, "Probabilistic inference using stochastic spiking neural networks on a neurosynaptic processor," in *2016 International Joint Conference on Neural Networks (IJCNN)*, 2016, pp. 4286–4293.

[215] S. Cheng, M.-H. Lee, R. Tran, Y. Shi, X. Li, H. Navarro, C. Adda, Q. Meng, L.-Q. Chen, R. C. Dynes, S. P. Ong, I. K. Schuller, and Y. Zhu, "Inherent stochasticity during insulator–metal transition in VO2," *Proceedings of the National Academy of Sciences*, vol. 118, no. 37, p. e2105895118, 2021.

[216] A. G. Shabalin, J. del Valle, N. Hua, M. J. Cherukara, M. V. Holt, I. K. Schuller, and O. G. Shpyrko, "Nanoscale Imaging and Control of Volatile and Non-Volatile Resistive Switching in VO2," *Small*, vol. 16, no. 50, p. 2005439, 2020.

[217] J. del Valle, P. Salev, F. Tesler, N. M. Vargas, Y. Kalcheim, P. Wang, J. Trastoy, M.-H. Lee, G. Kassabian, J. G. Ramírez, M. J. Rozenberg, and I. K. Schuller, "Subthreshold firing in Mott nanodevices," *Nature*, vol. 569, no. 7756, pp. 388–392, 2019.

[218] J.-G. Ramírez, A. Sharoni, Y. Dubi, M. E. Gómez, and I. K. Schuller, "First-order reversal curve measurements of the metal-insulator transition in VO2: Signatures of persistent metallic domains," *Physical Review B*, vol. 79, no. 23, p. 235110, 2009.

[219] E. Kisiel, P. Salev, I. Poudyal, F. Baptista, F. Rodolakis, Z. Zhang, O. Shpyrko, I. K. Schuller, Z. Islam, and A. Frano, "High-Resolution Full-field Structural Microscopy of the Voltage Induced Filament Formation in Neuromorphic Devices," 2023.

[220] A. G. Shabalin, J. del Valle, A. Charnukha, N. Hua, M. V. Holt, D. N. Basov, I. K. Schuller, and O. G. Shpyrko, "Nanoimaging of Electrical Failure in VO2 Resistive-Switching Nanodevices," *ACS Applied Electronic Materials*, vol. 2, no. 8, pp. 2357–2362, 2020.

[221] J. Jeong, Z. Yong, A. Joushaghani, A. Tsukernik, S. Paradis, D. Alain, and J. K. S. Poon, "Current induced polycrystalline-to-crystalline transformation in vanadium dioxide nanowires," *Scientific Reports*, vol. 6, no. 1, p. 37296, 2016.

[222] N. Ghazikhanian, J. del Valle, P. Salev, R. El Hage, Y. Kalcheim, C. Adda, and I. K. Schuller, "Resistive switching localization by selective focused ion beam irradiation," *Applied Physics Letters*, vol. 123, no. 12, p. 123505, 2023.

[223] P. Salev, E. Kisiel, D. Sasaki, B. Gunn, W. He, M. Feng, J. Li, N. Tamura, I. Poudyal, Z. Islam, Y. Takamura, A. Frano, and I. K. Schuller, "Local strain inhomogeneities during the electrical triggering of a metal-insulator transition revealed by the x-ray microscopy."

[224] Y. Arata, H. Nishinaka, M. Takeda, K. Kanegae, and M. Yoshimoto, "Strain-Induced Modulation of Resistive Switching Temperature in Epitaxial VO2 Thin Films on Flexible Synthetic Mica," *ACS Omega*, vol. 7, no. 45, pp. 41 768–41 774, 2022.

[225] J. Yoon, G. Lee, C. Park, B. S. Mun, and H. Ju, "Investigation of length-dependent characteristics of the voltage-induced metal insulator transition in VO2 film devices," *Applied Physics Letters*, vol. 105, no. 8, p. 083503, 2014.

[226] Y. Kalcheim, A. Camjayi, J. del Valle, P. Salev, M. Rozenberg, and I. K. Schuller, "Non-thermal resistive switching in Mott insulator nanowires," *Nature Communications*, vol. 11, no. 1, p. 2985, 2020.

[227] A. Parihar, M. Jerry, S. Datta, and A. Raychowdhury, "Stochastic IMT (Insulator-Metal-Transition) Neurons: An Interplay of Thermal and Threshold Noise at Bifurcation," *Frontiers in Neuroscience*, vol. 12, p. 210, 2018.

[228] S. K. Das, S. K. Nandi, C. V. Marquez, A. Rúa, M. Uenuma, E. Puyoo, S. K. Nath, D. Albertini, N. Baboux, T. Lu, Y. Liu, T. Haeger, R. Heiderhoff, T. Riedl,

T. Ratcliff, and R. G. Elliman, "Physical Origin of Negative Differential Resistance in V3O5 and Its Application as a Solid-State Oscillator," *Advanced Materials*, vol. 35, no. 8, p. 2208477, 2023.

[229] M. Lange, S. Guénon, Y. Kalcheim, T. Luibrand, N. M. Vargas, D. Schwebius, R. Kleiner, I. K. Schuller, and D. Koelle, "Imaging of Electrothermal Filament Formation in a Mott Insulator," *Physical Review Applied*, vol. 16, no. 5, p. 054027, 2021.

[230] S. Kumar, Z. Wang, N. Davila, N. Kumari, K. J. Norris, X. Huang, J. P. Strachan, D. Vine, A. L. D. Kilcoyne, Y. Nishi, and R. S. Williams, "Physical origins of current and temperature controlled negative differential resistances in NbO2," *Nature Communications*, vol. 8, no. 1, p. 658, 2017.

[231] S. Li, X. Liu, S. K. Nandi, S. K. Nath, and R. G. Elliman, "Origin of Current-Controlled Negative Differential Resistance Modes and the Emergence of Composite Characteristics with High Complexity," *Advanced Functional Materials*, vol. 29, no. 44, p. 1905060, 2019.

[232] M. Jerry, N. Shukla, H. Paik, D. G. Schlom, and S. Datta, "Dynamics of electrically driven sub-nanosecond switching in vanadium dioxide," in *2016 IEEE Silicon Nanoelectronics Workshop (SNW)*, 2016, pp. 26–27.

[233] V. S. Aliev, S. G. Bortnikov, A. G. Karakotsky, and V. I. Litvintsev, "Modeling of VO2 Relaxation Oscillator in the Fast-Slow System Approximation," *IEEE Transactions on Electron Devices*, vol. 69, no. 2, pp. 715–721, 2022.

[234] B. Razavi, *Design of analog CMOS integrated circuits*. New York, NY: McGraw-Hill Education, 2017.

[235] J. Frenkel, "On Pre-Breakdown Phenomena in Insulators and Electronic Semi-Conductors," *Physical Review*, vol. 54, no. 8, pp. 647–648, 1938.

[236] D. Allan, "Statistics of atomic frequency standards," *Proceedings of the IEEE*, vol. 54, no. 2, pp. 221–230, 1966.

[237] J. J. Snyder, "An Ultra-High Resolution Frequency Meter," in *Thirty Fifth Annual Frequency Control Symposium*, 1981, pp. 464–469.

[238] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen, W. L. Smith, R. L. Sydnor, R. F. C. Vessot, and G. M. R. Winkler, "Characterization of Frequency Stability," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, no. 2, pp. 105–120, 1971.

[239] T.-L. Wu, L. Whittaker, S. Banerjee, and G. Sambandamurthy, "Temperature and voltage driven tunable metal-insulator transition in individual WxV1-xO2 nanowires," *Physical Review B*, vol. 83, no. 7, p. 073101, 2011.

[240] I. P. Radu, B. Govoreanu, S. Mertens, X. Shi, M. Cantoro, M. Schaekers, M. Jurczak, S. D. Gendt, A. Stesmans, J. A. Kittl, M. Heyns, and K. Martens, "Switching mechanism in two-terminal vanadium dioxide devices," *nanotechnology*, vol. 26, no. 16, p. 165202, 2015.

[241] S. L. Fang, C. Y. Han, Z. R. Han, B. Ma, Y. L. Cui, W. Liu, S. Q. Fan, X. Li, X. L. Wang, G. H. Zhang, X. D. Huang, and L. Geng, "An Artificial Spiking Afferent Neuron System Achieved by 1M1S for Neuromorphic Computing," *IEEE Transactions on Electron Devices*, vol. 69, no. 5, pp. 2346–2352, 2022.

[242] S. Carapezzi, A. Plews, G. Boschetto, A. Nejim, S. Karg, and A. Todri-Sanial, "How fast can vanadium dioxide neuron-mimicking devices oscillate? Physical mechanisms limiting the frequency of vanadium dioxide oscillators," *Neuromorphic Computing and Engineering*, vol. 3, no. 3, p. 034010, 2023.

[243] D. Bol, "Pushing ultra-low-power digital circuits into the nanometer era," PhD dissertation, UCLouvain, Louvain-la-Neuve, 2008.

[244] C. Frenkel, "Sparsity provides a competitive advantage," *Nature Machine Intelligence*, vol. 3, no. 9, pp. 742–743, 2021.

[245] J. P. Segundo, E. Altshuler, M. Stiber, and A. Garfinkel, "Periodic inhibition of living pacemaker neurons (i): locked, intermittent, messy, and hopping behaviors," *International Journal of Bifurcation and Chaos*, vol. 01, no. 03, pp. 549–581, 1991.

[246] M. A. Fitzurka and D. C. Tam, "A joint interspike interval difference stochastic spike train analysis: detecting local trends in the temporal firing patterns of single neurons," *Biological Cybernetics*, vol. 80, no. 5, pp. 309–326, 1999.

[247] A. Longtin and K. Hinzer, "Encoding with Bursting, Subthreshold Oscillations, and Noise in Mammalian Cold Receptors," *Neural Computation*, vol. 8, no. 2, p. 215, 1996.

[248] M. Jerry, A. Parihar, B. Grisafe, A. Raychowdhury, and S. Datta, "Ultra-low power probabilistic IMT neurons for stochastic sampling machines," in *2017 Symposium on VLSI Technology*, 2017, pp. T186–T187.

[249] H. Liu, X. Zhu, Z. Guo, R. He, X. Li, Q. Sun, X. Ye, C. Sun, Y. Tian, and R.-W. Li, "Artificial Spiking Neuron with Bursting Dynamics for Noise-Resistant Neuromorphic Coding," *ACS Applied Electronic Materials*, vol. 5, no. 6, pp. 3454–3461, 2023.

[250] M. Drouhin, S. Li, M. Grelier, S. Collin, F. Godel, R. G. Elliman, B. Dlubak, J. Trastoy, D. Querlioz, and J. Grollier, "Characterization and modeling of spiking and bursting in experimental NbOx neuron," *Neuromorphic Computing and Engineering*, vol. 2, no. 4, p. 044008, 2022.

[251] J. W. Tukey, *Exploratory data analysis*. Reading, Mass.: Addison-Wesley Pub. Co., 1977.

[252] N. Albanbay, B. Medetov, and M. A. Zaks, "Exponential distribution of lifetimes for transient bursting states in coupled noisy excitable systems," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 31, no. 9, p. 093105, 2021.

[253] E. Merivee, A. Must, K. Nurme, A. Di Giulio, M. Muzzi, I. Williams, and M. Mänd, "Neural Code for Ambient Heat Detection in Elaterid Beetles," *Frontiers in Behavioral Neuroscience*, vol. 14, 2020.

[254] A. Must, E. Merivee, K. Nurme, I. Sibul, M. Muzzi, A. D. Giulio, and I. Williams, "Encoding noxious heat by spike bursts of antennal bimodal hygroreceptor (dry) neurons in the carabid Pterostichus oblongopunctatus," *Cell and Tissue Research*, vol. 368, no. 1, pp. 29–47, 2017.

[255] R. Krahe and F. Gabbiani, "Burst firing in sensory systems," *Nature Reviews Neuroscience*, vol. 5, no. 1, pp. 13–23, 2004.

[256] G. A. McMillan and J. R. Gray, "Burst Firing in a Motion-Sensitive Neural Pathway Correlates with Expansion Properties of Looming Objects that Evoke Avoidance Behaviors," *Frontiers in Integrative Neuroscience*, vol. 9, 2015.

[257] A. Rao, P. Plank, A. Wild, and W. Maass, "A Long Short-Term Memory for AI Applications in Spike-based Neuromorphic Hardware," *Nature Machine Intelligence*, vol. 4, no. 5, pp. 467–479, 2022.

[258] N. Krishnan, S. Boyd, A. Somani, S. Raoux, D. Clark, and D. Dornfeld, "A Hybrid Life Cycle Inventory of Nano-Scale Semiconductor Manufacturing," *Environmental Science & Technology*, vol. 42, no. 8, p. 3069–3075, 2008.

[259] S. B. Boyd, *Life-Cycle Assessment of Semiconductors*. New York, NY: Springer New York, 2012.

[260] M. Garcia Bardon, P. Wuytens, L.-A. Ragnarsson, G. Mirabelli, D. Jang, G. Willems, A. Mallik, A. Spessot, J. Ryckaert, and B. Parvais, "DTCO including Sustainability: Power-Performance-Area-Cost-Environmental score (PPACE) Analysis for Logic Technologies," in *2020 IEEE International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE, 2020, pp. 41.4.1–41.4.4.

[261] T. Pirson, T. P. Delhaye, A. G. Pip, G. L. Brun, J.-P. Raskin, and D. Bol, "The Environmental Footprint of IC Production: Review, Analysis, and Lessons From Historical Trends," *IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING*, vol. 36, no. 1, 2023.

[262] L. Boakes, M. Garcia Bardon, V. Schellekens, I.-Y. Liu, B. Vanhouche, G. Mirabelli, F. Sebaai, L. Van Winckel, E. Gallagher, C. Rolin, and L.-A. Ragnarsson, "Cradle-to-gate Life Cycle Assessment of CMOS Logic Technologies," in *2023 International Electron Devices Meeting (IEDM)*. San Francisco, CA, USA: IEEE, 2023, pp. 1–4.

[263] G. Orchard, E. P. Frady, D. B. D. Rubin, S. Sanborn, S. B. Shrestha, F. T. Sommer, and M. Davies, "Efficient Neuromorphic Signal Processing with Loihi 2," 2021, arXiv:2111.03746 [cs].

[264] "Decadal Plan for Semiconductors - SRC," Semiconductor Research Corporation, Accessed on 2024-07-02. [Online]. Available: https://www.src.org/about/decadal-plan/

[265] M. Buyle, A. Audenaert, P. Billen, K. Boonen, and S. Van Passel, "The Future of Ex-Ante LCA? Lessons Learned and Practical Recommendations," *Sustainability*, vol. 11, no. 19, p. 5456, 2019.

[266] M. Z. Hauschild, *Life Cycle Impact Assessment*. Dordrecht: Springer Netherlands, 2015.

[267] "ISO 14040:2006/Amd 1:2020," ISO standard, Accessed on 2024-07-02. [Online]. Available: https://www.iso.org/standard/76121.html

[268] L. M. Hilty and B. Aebischer, "ICT for sustainability: An emerging research field. In:," in *ICT Innovations for Sustainability,*. Springer, 2015, pp. 3–36.

[269] T. Ekvall, "Attributional and Consequential Life Cycle Assessment," in *Sustainability Assessment at the 21st century*. IntechOpen, 2019.

[270] E. Strubell, A. Ganesh, and A. McCallum, "Energy and Policy Considerations for Deep Learning in NLP," 2019, arXiv:1906.02243 [cs].

[271] K. Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

[272] A. van Wynsberghe, "Sustainable AI: AI for sustainability and the sustainability of AI," *AI and Ethics*, vol. 1, no. 3, pp. 213–218, 2021.

[273] M. Proske, D. Sánchez, C. Clemm, and S.-J. Baur, "Life Cycle Assessment of the Fairphone 3," Fraunhofer IZM, Tech. Rep., 2020.

[274] M. S. Branham, "Semiconductors and sustainability : energy and materials use in integrated circuit manufacturing," Ph.D. dissertation, Massachusetts Institute of Technology, 2008.

[275] T. Pirson, T. P. Delhaye, A. G. Pip, G. L. Brun, J.-P. Raskin, and D. Bol, "The Environmental Footprint of IC Production: Review, Analysis, and Lessons From Historical Trends," *IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING*, vol. 36, no. 1, 2023.

[276] S. Cucurachi, C. van der Giesen, and J. Guinée, "Ex-ante LCA of Emerging Technologies," *Procedia CIRP*, vol. 69, pp. 463–468, 2018.

[277] P. Nuss and M. J. Eckelman, "Life Cycle Assessment of Metals: A Scientific Synthesis," *PLOS ONE*, vol. 9, no. 7, p. e101298, 2014.

[278] "IEA World Energy Statistics and Balances," International Energy Agency (IEA), Accessed on 2024-07-02. [Online]. Available: https://www.oecd-ilibrary.org/energy/data/iea-world-energy-statistics-and-balances_enestats-data-en

[279] "Greenhouse gas emission intensity of electricity generation in Europe," European Environment Agency, Accessed on 2024-07-02. [Online]. Available: https://www.eea.europa.eu/en/analysis/indicators/greenhouse-gas-emission-intensity-of-1

[280] S. Kim and M. Overcash, "Energy in chemical manufacturing processes: gate-to-gate information for life cycle assessment," *Journal of Chemical Technology & Biotechnology*, vol. 78, no. 9, pp. 995–1005, 2003.

[281] A. A. Yaroshevsky, "Abundances of chemical elements in the Earth's crust," *Geochemistry International*, vol. 44, no. 1, pp. 48–55, 2006.

[282] S. Kadivar, H. Akbari, and E. Vahidi, "Assessing the environmental impact of gold production from double refractory ore in a large-scale facility," *Science of The Total Environment*, vol. 905, p. 167841, 2023.

[283] E. D. Williams, R. U. Ayres, and M. Heller, "The 1.7 Kilogram Microchip: Energy and Material Use in the Production of Semiconductor Devices," *Environmental Science and Technology*, vol. 36, no. 24, pp. 5504–5510, 2002.

[284] G. Indiveri, E. Chicca, and R. Douglas, "A VLSI Array of Low-Power Spiking Neurons and Bistable Synapses With Spike-Timing Dependent Plasticity," *IEEE Transactions on Neural Networks*, vol. 17, no. 1, pp. 211–221, 2006.

[285] J. M. Cruz-Albrecht, M. W. Yung, and N. Srinivasa, "Energy-efficient neuron, synapse and STDP integrated circuits," *IEEE transactions on biomedical circuits and systems*, vol. 6, no. 3, pp. 246–256, 2012.

[286] I. Sourikopoulos, S. Hedayat, C. Loyez, F. Danneville, V. Hoel, E. Mercier, and A. Cappy, "A 4-fJ/Spike Artificial Neuron in 65 nm CMOS Technology," *Frontiers in Neuroscience*, vol. 11, p. 123, 2017.

[287] K. Nalliboyina and S. Ramachandran, "An energy-efficient hybrid CMOS spiking neuron circuit design with a memristive based novel T-type artificial synapse," *AEU - International Journal of Electronics and Communications*, vol. 173, p. 154982, 2024.

[288] S. Zhu, H. Hu, H. Yang, Y. Qu, and Y. Li, "Mini-Review of Best Practices for Greenhouse Gas Reduction in Singapore's Semiconductor Industry," *Processes*, vol. 11, no. 7, p. 2120, 2023.

[289] B. Son, Y. Suh, S. Kim, H. Jung, J.-S. Kim, C. Shin, K. Park, K. Lee, J. Park, J. Woo, Y. Roh, H. Lee, Y. Wang, I. Ovsiannikov, and H. Ryu, "4.1 A 640×480 dynamic vision sensor with a 9μm pixel and 300Meps address-event representation," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, 2017, pp. 66–67.

[290] C. Li, L. Longinotti, F. Corradi, and T. Delbruck, "A 132 by 104 10um-Pixel 250uW 1kefps Dynamic Vision Sensor with Pixel-Parallel Noise and Spatial Redundancy Suppression," in *2019 Symposium on VLSI Circuits*, 2019, pp. C216–C217.

[291] C.-S. Poon and K. Zhou, "Neuromorphic Silicon Neurons and Large-Scale Neural Networks: Challenges and Opportunities," *Frontiers in Neuroscience*, vol. 5, 2011.

[292] Z. Dong, X. Ji, G. Zhou, M. Gao, and D. Qi, "Multimodal Neuromorphic Sensory-Processing System With Memristor Circuits for Smart Home Applications," *IEEE Transactions on Industry Applications*, vol. 59, no. 1, pp. 47–58, 2023.

[293] P. Stoliar, M. Rozenberg, E. Janod, B. Corraze, J. Tranchant, and L. Cario, "Nonthermal and purely electronic resistive switching in a Mott memory," *Physical Review B*, vol. 90, no. 4, p. 045146, 2014.

[294] S. Carapezzi, G. Boschetto, C. Delacour, E. Corti, A. Plews, A. Nejim, S. Karg, and A. Todri-Sanial, "Advanced Design Methods From Materials and Devices to Circuits for Brain-Inspired Oscillatory Neural Networks for Edge Computing," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 11, no. 4, pp. 586–596, 2021.

[295] M. Quet, "La critique des technologies émergentes face à la communication promettante. Contestations autour des nanotechnologies," *Réseaux*, vol. 173-174, no. 3-4, pp. 271–302, 2012.

[296] M. Borup, N. Brown, K. Konrad, and H. Van Lente, "The sociology of expectations in science and technology," *Technology Analysis & Strategic Management*, vol. 18, no. 3-4, pp. 285–298, 2006.

[297] V. C. Müller, "Ethics of Artificial Intelligence and Robotics," in *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, 2023.

[298] L. Drew, "The ethics of brain–computer interfaces," *Nature*, vol. 571, no. 7766, pp. S19–S21, Jul. 2019.

[299] E. J. Davidoff, "Agency and Accountability: Ethical Considerations for Brain-Computer Interfaces," *The Rutgers journal of bioethics*, vol. 11, pp. 9–20, 2020.

[300] D. V. Christensen, R. Dittmann, B. Linares-Barranco, A. Sebastian, M. L. Gallo, A. Redaelli, S. Slesazeck, T. Mikolajick, S. Spiga, S. Menzel, I. Valov, G. Milano, C. Ricciardi, S.-J. Liang, F. Miao, M. Lanza, T. J. Quill, S. T. Keene, A. Salleo, J. Grollier, D. Marković, A. Mizrahi, P. Yao, J. J. Yang, G. Indiveri, J. P. Strachan, S. Datta, E. Vianello, A. Valentian, J. Feldmann, X. Li, W. H. P. Pernice, H. Bhaskaran, S. Furber, E. Neftci, F. Scherr, W. Maass, S. Ramaswamy, J. Tapson, P. Panda, Y. Kim, G. Tanaka, S. Thorpe, C. Bartolozzi, T. A. Cleland, C. Posch, S. Liu, G. Panuccio, M. Mahmud, A. N. Mazumder, M. Hosseini, T. Mohsenin, E. Donati, S. Tolu, R. Galeazzi, M. E. Christensen, S. Holm, D. Ielmini, and N. Pryds, "2022 roadmap on neuromorphic computing and engineering," *Neuromorphic Computing and Engineering*, vol. 2, no. 2, p. 022501, 2022.

[301] S. Cheng, M.-H. Lee, X. Li, L. Fratino, F. Tesler, M.-G. Han, J. del Valle, R. C. Dynes, M. J. Rozenberg, I. K. Schuller, and Y. Zhu, "*Operando* characterization of conductive filaments during resistive switching in Mott VO_2," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, p. e2013676118, 2021.

[302] "Allan variance," 2013, lecture from the engineering school of the Iowa State University, Accessed on 2024-10-02. [Online]. Available: https://home.engineering.iastate.edu/~shermanp/AERE432/lectures/Rate%20Gyros/Allan%20variance.pdf

Appendices

A Appendix A

The presence of an initial *electro-forming cycle* has been reported in a number of VO_2 two-terminal devices [120, 159, 166, 220, 221]. This phenomenon takes place during the first electrically-induced transition measured in the device, and can be triggered both using current- and voltage-driven measurements. Typically, upon the first cycle, the device transitions at a relatively high bias point. This initial transition results in some form of damage to the crystalline structure of the VO_2 film inside the device channel. In some polycrystalline films, the VO_2 can locally appear to have "melted" [159, 220], with grain boundaries no longer visible. This phenomenon is more prominent in films with relatively high OFF/ON resistivity ratio [120], as the film could locally reach significantly high temperature upon the formation of the first filament, due to large local current densities. The nature of these damages (transition from polycrystalline to crystalline [221], change in stoichiometry [301] or VO_2 phase/metallicity [220]) seem to be very dependent on the specific device and film properties, as well as the electro-forming conditions [166].

After this first cycle, the resistance of the device in its insulating state is usually lowered. When testing a large number of devices, a certain proportion of them do not show reliable electrically-triggered transition properties anymore. However, some of them still present repeatable resistive switching, usually at a lower bias point (I_{IMT}, V_{IMT}) than upon the first cycle. The specific design parameters impacting the *yield* of functional devices among those requiring electro-forming are not yet well identified.

In our work, only the large-grains devices showed such electro-forming behavior, possibly due to their larger OFF/ON resistivity ratio. This phenomenon is illustrated for one specific device in Fig. A.1. These damages are more pronounced in devices of larger length L, as shown on Fig. A.2 and A.3. They also seem to be more pronounced when the electro-forming is triggered through a voltage-driven measurement (Fig. A.2), as opposed to a current-driven measurement (Fig. A.3). Indeed, in current-driven measurements, the current flowing through the device is limited by the source as the filament gradually forms. In voltage-driven measurements, the filament formation leads to a fast increase in total current: even when setting a current compliance on the source, this phenomenon could be too fast for the equipment detection dynamics, leading to a current spike through the device.

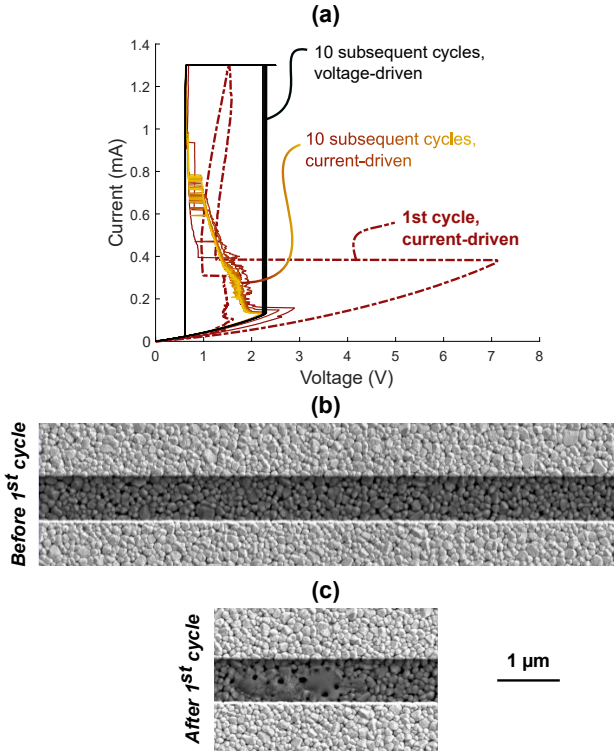


Fig. A.1 Electroforming behavior illustrated in one of the large-grains devices (see Chapter 2 for specific process parameters). Devices dimensions are L = 1.2 µm, W = 200 µm. **(a)** I-V characteristics of the devices before and after electro-forming. Upon the first current-driven cycle, the device transitions at a relatively high bias point. After the first cycle, the 10 consecutive current-driven measurements show that the device transitions at a lower bias point, and has a lower resistance in the insulating state. A pronounced NDR behavior is observed. Some form of training is required for the device to show repeatable I-V characteristics. Resistive-switching properties are preserved and demonstrated upon 10 consecutive voltage-driven measurements. **(b)** SEM image of the device in its pristine state, prior to any electrical measurement. **(c)** SEM image of the device after the first cycle, at the specific location in the channel where damage has occurred, showing some form of melting or coalescence of the grains.

On the other hand, our small-grains devices did not show any electro-forming behavior. For each measured device, the first I-V curve was similar to the subsequent ones, with no significant variations in the device electrical parameters (other than the cycle-to-cycle variations discussed in detail in Chapter 2). We attribute this to the lower OFF/ON resistivity ratio of the small-grains film (below 10^2 *vs* more than 10^3 for the large-grains film), resulting in a lower local current density inside the metallic filament.

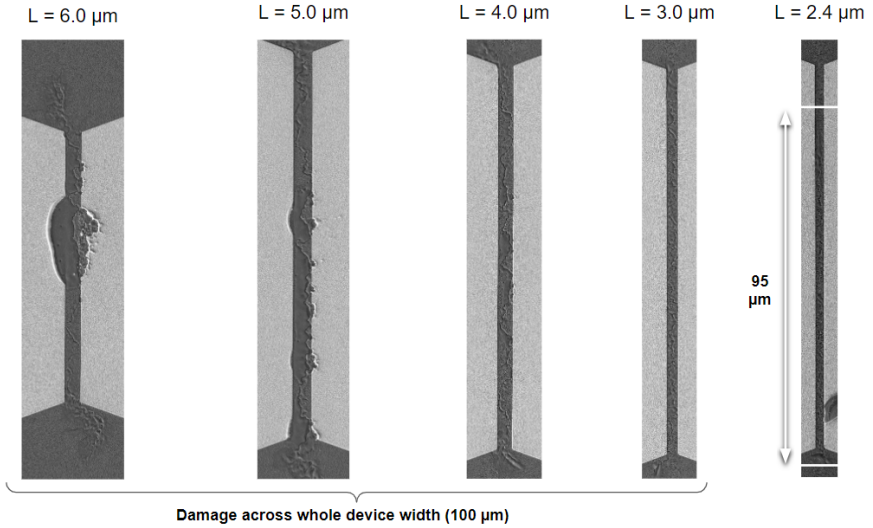


Fig. A.2 SEM imaging of large-grains devices of various lengths (L = 6 µm to 2.4 µm) and identical 100 µm width, after the first electro-forming cycle performed through a **voltage-driven** measurement. Width of the damages are indicated.

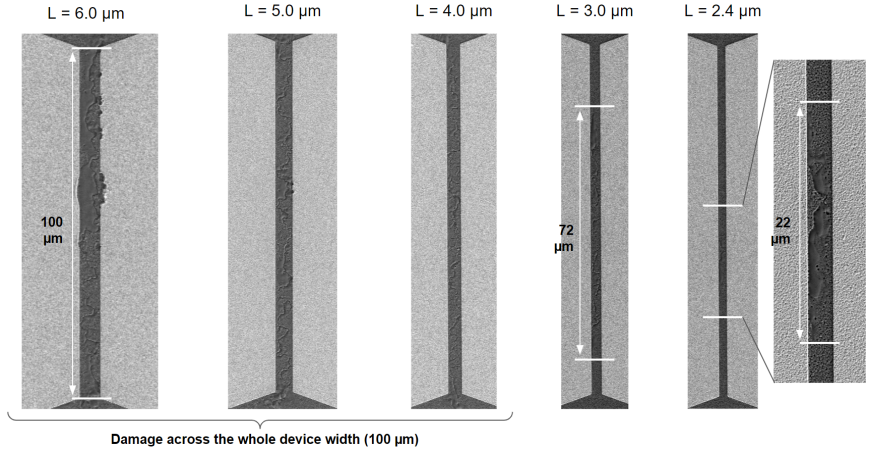


Fig. A.3 SEM imaging of large-grains devices of various lengths (L = 6 µm to 2.4 µm) and identical 100 µm width, after the first electro-forming cycle performed through a **current-driven** measurement. Width of the damages are indicated, showing that they are more pronounced in devices of larger length.

The fact that the large-grains devices show an electro-forming behavior that is absent in small-grains devices could also explain the observed difference in stochasticity investigated in Chapter 2. Indeed, the damages to the crystalline structure could impact the filament formation. To understand how, high-resolution microscopy imaging of the large-grains devices should be performed to assess the location of the filament over several cycles. When attempting to perform such measurements with our optical setup, we could not distinguish the filament as clearly as for the small-grains devices, possibly because it has a much smaller width, or because the optical contrast is not as pronounced.

In all cases, our hypothesis regarding the correlation between the number of grains along the channel length and the stochastic behavior is backed up by the stochasticity dependency to device length.

B Appendix B

This appendix details the computation of the confidence interval around the Allan deviation [302]. The lower and higher bounds involve the use of the chi-squared distribution χ^2. To estimate a 90 % confidence interval, the lower bound uses the 95 % range of the distribution $\chi^2(0.95)$, and the higher bound uses the 5 % range $\chi^2(0.05)$:

$$\sqrt{\frac{df \cdot \sigma_{y,estimated}(t_{obs})^2}{\chi^2(0.95)}} < \sigma_{y,true}(t_{obs}) < \sqrt{\frac{df \cdot \sigma_{y,estimated}(t_{obs})^2}{\chi^2(0.05)}} \quad , \tag{B.2}$$

with $\sigma_{y,estimated}(t_{obs})$ the estimator for the Allan deviation detailed in Eq. 3.14 and 3.15 of Chapter 3, $\sigma_{y,true}(t_{obs})$ the true Allan deviation, and df the degrees of freedom for the Allan deviation estimator. For a noise of the white frequency modulation type, these degrees of freedom are given by the following expression:

$$df \cong \left(\frac{3(N-1)}{2m} - \frac{2(N-2)}{N}\right)\frac{4m^2}{4m^2+5} \quad , \tag{B.3}$$

with m the number of oscillator periods in the considered interval such that $t_{obs} = mT_{spike}$, and $N - 2m$ the total number of overlapped intervals considered in the Allan deviation estimator.