Towards Oxide Electronics: a Roadmap


Published in:
Applied Surface Science

Link to article, DOI:
10.1016/j.apsusc.2019.03.312

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Towards Oxide Electronics: a Roadmap


Corresponding author.
E-mail address: fabio.miletto@spin.cnr.it (F. Miletto Granozio).

https://doi.org/10.1016/j.apsusc.2019.03.312

0169-4332/ © 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).
Abstract
At the end of a rush lasting over half a century, in which CMOS technology has been experiencing a constant and breathtaking increase of device speed and density, Moore's law is approaching the insurmountable barrier given by the ultimate atomic nature of matter. A major challenge for 21st century scientists is finding novel strategies, concepts and materials for replacing silicon-based CMOS semiconductor technologies and guaranteeing a continued and steady technological progress in next decades. Among the materials classes candidate to contribute to this momentous challenge, oxide films and heterostructures are a particularly appealing hunting ground. The vastity, intended in pure chemical terms, of this class of compounds, the complexity of their correlated behaviour, and the wealth of functional properties they display, has already made these systems the subject of choice, worldwide, of a strongly networked, dynamic and interdisciplinary research community.

Oxide science and technology has been the target of a wide four-year project, named Towards Oxide-Based Electronics (TO-BE), that has been recently running in Europe and has involved as participants several hundred scientists from 29 EU countries. In this review and perspective paper, published as a final deliverable of the TO-BE Action, the opportunities of oxides as future electronic materials for Information and Communication Technologies ICT and Energy are discussed. The paper is organized as a set of contributions, all selected and ordered as individual building blocks of a wider general scheme. After a brief preface by the editors and an introductory contribution, two sections follow. The first is mainly devoted to providing a perspective on the latest theoretical and experimental methods that are employed to investigate oxides and to produce oxide-based films, heterostructures and devices. In the second, all contributions are dedicated to different specific fields of applications of oxide thin films and heterostructures, in sectors as data storage and computing, optics and plasmonics, magnonics, energy conversion and harvesting, and power electronics.
Preface

Mariona Coll, Josep Fontcuberta, Nini Pryds, Fabio Miletto Granzio

In early 2007, motivated by the discovery of high mobility and quantum Hall effect in a transition metal oxide [1] A.P. Ramirez, enthusiastically argued that "the era of oxide electronics" had been finally entered [2]. Trying to summarize in a preface, more than ten years later, what makes oxide-based electronics potentially so special is still an arduous task. In the realm of oxides diversity and complexity play a key role. Oxide physics is not easily reduced to a few words.

The unparalleled wealth of oxide's functional properties is first rooted in the extreme diversity characterizing, in purely chemical terms, this materials platform. The introductory paper 'A unique exploration' sharply clarifies that, at the closing of a technological era prevalently dominated by Si, oxides open the whole periodic table as the playground of tomorrow's materials scientists. Still, this spectacular chemical diversity only partially contributes to making oxides arguably the richest class of electronic materials. The sensitivity of oxides electronic systems to their structural background is in fact such, that competing ground states with highly different properties can be also obtained under minimal (percental) chemical variations. Even at fixed composition, phase transitions can be induced under the effect of different structural knobs modifying the systems boundary conditions: e.g. thickness, strain, grain boundaries and interfaces.

The main reason that oxide materials are 'complex' (using this word in its literal physical sense [3]) can be largely traced back to the narrowness of their bandwidths. The chemical origin of this lies, in turn, in the ionic character of oxygen bonds, if compared with the highly covalent IV and III-V semiconductors. Bandwidth reduction becomes extreme in the case of transition metal (TM) oxides, dominated by the scarcely overlapping and partially occupied metal d orbitals, with m = 3, 4 or 5 and n = 1,.., 10. For such materials the bandwidth W approaches the eV or even sub-eV energy scale. In this regime, a number of intrinsic degrees of freedom, e.g. electrostatic on-site repulsion, spin-orbit coupling, magnetic interactions and electron-lattice interactions, having comparable associated energies, come strongly into play in determining the system's electronic ground state. It's no surprise, therefore, that based on this intrinsic competition of charge, spin orbital and lattice degrees of freedom, TM oxides present the richest variety of emergent states and that new paradigms are needed to tackle their theoretical understanding, as explained in the second introductory contribution ('What theoretical approaches can provide: a perspective on oxides electronics'), to both interpret the experimental data and foresee the properties of not-yet-synthesized materials.

The level of complexity rises even further when interfaces between oxides are considered. Adjacent materials affect each other through, e.g., charge transfer related to band mismatch, built-in fields, dielectric/ferroelectric surface charges, strain, structural distortions and magnetic (exchange, superexchange, Dzyaloshinskii-Moriya) interactions. The discovery of the existence of a high mobility and superconducting electron gas at interfaces between large band gap insulating oxides [4] is a clear example of the new "opportunities that interfaces bring to Oxide Electronics" [5]. In a new twist, emerging topological properties arising at interfaces and surfaces of oxide thin films are now driving a renewed attention [6]. In 2016, M. Lorentz and M.S. Ramachandra Rao brilliantly brought together views from experts in the field and published an Oxide electronic materials and oxide interfaces road-map that constitutes a snapshot of the exciting oxide-based science [7].

The maturity of the field, both in terms of materials, concepts and methods, now demands a renewed attention to address a few crucial questions about the potential technological and social benefits of a research that has involved enormous amount of human, financial and technical resources in last decades. The purpose of the current paper "Towards Oxide Electronics: a Roadmap", published as Final Action Dissemination of the TO-BE COST Action MP1308 ("Towards oxide-based electronics"), is to shift the focus, with respect to the previously mentioned effort [7], addressing more directly some crucial questions:

- What is the future role of oxide thin film in modern technologies?
- How far are oxides from taking this role?
- Which are the competing technologies?
- Which are the hurdles presently preventing the diffusion of marketable applications?
- Which are the chances these hurdles can be overcome?
- Which are the advancement in synthesis and characterization methods that can help us facing them?

In making our selection of topics and in defining the index of the present paper, some filters were obviously applied. On the one hand, we tended to privilege the technologies and the ideas that already gained some recognition among the principal figures (companies and academic institutions) of the ongoing technological rush, with respect to others that remained so far mostly confined to the 'oxide' community. On the other hand, we focused explicitly on thin-film-based applications, and when possible on epitaxial films, deliberately neglecting vast sectors of oxide technologies based e.g. on bulk devices and on nanocrystalline samples/surfaces. Due to the latter choice, not only many well established technologies resorting to oxide components were ignored, but even some novel and very active fields of oxide research, as gas sensing and catalysis, were not considered.

Most of the applications presented in the following are expected to impact the fields of ICT and Energy. In times in which mankind is facing the extreme ambition of artificially emulating the computing power of a human brain, exceeding the exaflop range and corresponding, with the present energy-hungry electronic technologies, to a power consumption in the several GW range, the two fields are related as never before.

Selected topics of this Roadmap include: (a) Insights from theory and modelling, (b) advanced growth, nanofabrication and characterization techniques and (c) applications on: data storage and computing, optics and plasmonics, magnonics, energy conversion and harvesting, and power electronics.

Our hope is that this Roadmap will show to be valuable both as a collective self-analysis of the oxide electronics community, providing an updated picture of the state-of-the-art in the field, and as a dissemination tool, helpful to whoever is willing to spread knowledge about oxide science and to attract towards this exciting field of research new young scientists and further public and private resources.

Acknowledgments

This paper is based upon work from COST Action MP1308 "Towards oxide-based electronics" (TO-BE), supported by COST (European Cooperation in Science and Technology) - www.cost.eu. Josep Fontcuberta and Mariona Coll express their acknowledgement for the financial support from the Spanish Government, through the "Severo Ochoa" Programme for Centres of Excellence in R&D (SEV-2015-0496) and the MAT2017-85232-R, and MAT2017-83169-R projects and from Generalitat de Catalunya (2017 SGR 1377) during the preparation of this document. M.C. acknowledges the Ramon y Cajal Ryc2013-12448 contract. Nini Pryds is gratefully acknowledge the Danish Council for Independent Research Technology and Production Sciences for the DFF-Research Project 2 grants No. 6111-00145B (NICE) and the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 801267 (BioWings)
A unique exploration
J. Mannhart *, H. Boschker
Max Planck Institute for Solid State Research, Heisenbergstr. 1, 70569 Stuttgart, Germany
* Corresponding author.
E-mail: office-mannhart@fkf.mpg.de (J. Mannhart).

Abstract
The Roadmap of Oxide Technologies for Electronic Applications provides a view into the future of oxide electronics. In this introduction to the roadmap, we scrutinize possible merits of oxide electronics, oxide electronics research, and the objectives of the roadmap.

Keywords: Oxides; Oxide electronics; Periodic table

Why is research valuable to our society? Which projects should be pursued? And, from the perspective of us scientists, what research is worth our own personal time and effort? Apparently, there are two obvious, diverging answers:

First, there is fundamental research to understand how the universe works. Such explorations are an essential trait of humans, which is why numerous space missions, and research institutions such as CERN and LIGO have received and continue to enjoy funding. It is less widely known, however, that profound questions about the basic properties of nature are also being answered by solid-state research. For example, current research seeks to resolve the wonders of macroscopic quantum states that raise quantum physics to the human scale, to reveal the existence and properties of novel quasiparticles, or to illuminate the astounding capabilities of emerging physical behaviors. It is worth noting that the study of electron systems in oxides is one of the methods with which these intriguing questions are being explored. For instance, consider the studies of the d-wave symmetry of the cuprate superconductors' macroscopic order parameter (comparable to a gigantic molecular state) and the symmetry's influence on the current-carrying properties of high-Tc cables [8]. Or take the following, less applied examples: the search for a possible electric dipole moment of the electron using multiferroics to investigate possible charge-parity (CP) violations [9,10], the search for magnetic monopoles in spin ice [11,12], or the exploration of artificial atoms using complex oxides [13].

Second, research has sometimes led to applications that have improved our lives and may even be essential for our future, and this clearly provides abundant motivation to pursue such work. In fact, the most important, revolutionary applications developed to date have typically arisen from curiosity-driven investigations. Often, the subsequent applications had not even been imagined when the original research was performed, as exemplified by spectacular results from solid-state science. For example, Kamerlingh Onnes' drive to cool matter to 4.2 K laid the base for magnetic resonance imaging (MRI) with its enormous value in medicine, or Michael Faraday's studies of electrical conductivity in Ag₂S [14], which led to the discovery of semiconductors, or—to mention but one such pioneer—Ferdinand Braun's investigations of rectifying metal contacts to PbS or FeS crystals, which induced the invention of the diode, the solar cell, the transistor, the integrated circuit and the Internet. It will be exciting to see the unexpected applications to which oxide electronics will lead in the future.

In addition to these breakthrough discoveries, numerous technical advances have occurred in a more gradual and—some say—more predictable manner. Our roadmap will provide an overview of these developments in the field of oxide electronics, ranging from advances in film growth and the characterization of heterostructures at the atomic level to ferroelectrics, oxide photovoltaics, solid-state fuel cells, neuromorphic computing, oxide microelectromechanical systems (MEMS), and transparent conductors.

Note that, whereas this roadmap illuminates the progress of oxide technologies and electronics from the perspective of applications, the advances achieved by improving fabrication technologies, analytical tools and our understanding of oxide devices will feed back into fundamental science. In addition, addressing and questioning the limits of technology, which this roadmap intends to trigger, will continue to inspire scientists to raise fundamental questions about our physical world.

Another consideration we deemed relevant in putting this road-map into perspective was expressed by Rolf Landauer in a notable paper [15]. Based on his long-time experience at IBM Research, he vividly stresses in this sobering publication that only a small fraction of the innovative devices envisaged, invented, or developed by scientists ever become successful products. Although this is obviously true, Landauer's warning in no way calls for us not to invent novel devices, at least if common-sense is used in considering their possible value. As we know, some inventions will be quite successful, and often it is simply impossible to foresee which of the many innovations will become winners. Indeed, oxides have already made a huge impact on electronics. Consider for example (Ba,Sr)TiO₃-based capacitors, PZT-based transducers, Hf-based oxinitrides for high-k gate insulators, and indium tin oxide as transparent conductors for displays, smart windows and photovoltaics.

These considerations underscore that research in an applied and scientifically exciting field such as oxide electronics is relevant for both, applications as well as fundamental science. But there is even more to oxide electronics. Oxide electronics is a trailblazer in a unique adventure: the pioneering endeavor of opening the entire periodic table to applications in electronic devices. To appreciate the relevance of this effort, we should recall the astounding fact that only a few but extensively exploited chemical elements have driven the phenomenal success of complementary metal-oxide-semiconductor devices based on silicon (Si-CMOS), the workhorse of standard semiconductor technology: silicon and germanium are used as semiconductors, oxygen is added to make SiO₂, copper and aluminum serve as interconnects, hafnium and nitrogen for gate dielectrics, and gold for the contacts. In addition to these pervasive Si-CMOS elements, a much larger number of elements and compounds are relevant constituents of industrial semiconductor devices. To name but a few examples besides Si and Ge, compounds such as GaAs, InSb, GaP, CdTe and ZnO are obviously important semiconductors. Furthermore, numerous elements such as B, As, P and Ga are dopants of choice. And let's not forget materials such as Se, Cu₂O, and CuS, which played important roles in semiconducting devices of the past.

It is remarkable that virtually all of these materials comprise semiconductor electron systems only. In essence, mean-field systems define the arena of materials for semiconducting electronics, because semiconducting devices rely almost exclusively on mean-field phenomena. Band diagrams are an essential tool for semiconductor engineers. Recall Herbert Kroemer's notable statement, "If, in discussing a semiconductor problem, you cannot draw an Energy Band Diagram, this shows that you don't know what you are talking about" [16]. Mean-field behavior provides the basis for the rules of band-diagram design. One of the few exceptions to the mean-field behavior of standard semiconductors is the quantum-Hall effect where, by applying magnetic fields, we can induce correlations in mean-field systems [13].

Although a fairly large and growing number of elements besides silicon and numerous compounds thereof are used in semiconductor technology today, they are but a subset of the periodic table, and the number of compounds used in semiconductor applications pales in comparison to the full range of electronic materials at our disposal. In a vigorous development that was much initiated by the discovery of high-Tc superconductivity in the copper oxides, the scientific community is now making rapid progress in learning how to grow films and heterostructures using almost any element of the periodic table that is not a noble gas, super-toxic, or radioactive (see Fig. 1).

This endeavor is not restricted to pure elements. Countless
While the exploration of the oxides is of special interest for devices and is at the core of our oxide electronics roadmap, oxide electronics research has to be understood as part of a more general trend toward using complex compounds. The advances, possibilities, and challenges discussed in this roadmap of designing, growing and using oxides are valid in principle also for nitrides, sulfides and many other compounds, and to some extent even for organic systems. The ongoing adventure is to explore and leverage virtually all elements of the periodic table and compounds thereof for practical applications as well as for fundamental science.

As more elements become available for use in devices, the number of possible compounds grows astronomically [19]. The fabrication of highly complex compounds is of course a generic goal of chemistry and materials science. It is already possible today to grow many of such materials as epitaxial films and heterostructures, and to pattern them into nanodevices. In fact, with our capability to grow heterostructures with atomic precision, the phase space of electronic systems is becoming virtually unlimited.

In this development, not only the number of available compounds is growing, but, more importantly, also their range of functional properties. The mean-field compounds discussed above are based almost exclusively on s and p-electron systems. Thanks to transition-metal oxides and elements, electronic correlations are becoming increasingly relevant by providing completely new arenas for device applications, such as phase-transition transistors, negative capacitors and photovoltaic cells based on correlation effects, all of which are complementing and adding to mean-field Si-CMOS. Nevertheless, despite these bright prospects for device applications, we should be mindful of the leap from device demonstrations to successful products, as described by Rolf Landauer.

There is yet another issue to ponder: Exploring the periodic table to discover and synthesize materials for electronic devices will be a one-off research adventure because there is of course only one periodic table. In their ground states, chemical elements just possess a limited set of s, p, d and f orbitals and linear combinations thereof, with no additional orbitals to come. The effects we will find, the discoveries we hope to make, and the issues we will uncover regarding their applications will be what they are, as there will never be a new set of chemical elements. The periodic table is the one stomping ground we have.

The scientific community is developing increasingly refined tools to design, grow and characterize these materials with atomic precision. The accuracy and supercell size of DFT calculations are being extended to inform the design of heterostructures and devices. Handling smaller supercells, DMFT calculations allow correlation effects to be assessed. Advanced film growth techniques based, for example, on molecular beam epitaxy (MBE), pulsed laser deposition (PLD), and atomic layer deposition (ALD) are now available, and analytical tools such as scanning transmission electron microscopy (STEM) and electron energy loss spectroscopy (EELS), resonant inelastic X-ray scattering (RIXS) and angle-resolved photoelectron spectroscopy (ARPES), transport measurements and tunneling spectroscopy are being used with great success for characterization. With recent advances in stoichiometry control, substrate termination, and reflection high-energy electron diffraction (RHEED), it is now possible to grow heterostructures of complex compounds of surprising quality. Although defect concentrations and mobilities of semiconducting heterostructures are distinctly superior to those of complex oxides, according to their STEM images and EELS analyses, in terms of quality, oxide heterostructures already seem at least comparable to today’s most advanced III-V heterostructures, an achievement that was unthinkable twenty years ago (see, e.g., [18]).

While the exploration of the oxides is of special interest for devices and is at the core of our oxide electronics roadmap, oxide electronics research has to be understood as part of a more general trend toward using complex compounds. The advances, possibilities, and challenges discussed in this roadmap of designing, growing and using oxides are valid in principle also for nitrides, sulfides and many other compounds, and to some extent even for organic systems. The ongoing adventure is to explore and leverage virtually all elements of the periodic table and compounds thereof for practical applications as well as for fundamental science.

As more elements become available for use in devices, the number of possible compounds grows astronomically [19]. The fabrication of highly complex compounds is of course a generic goal of chemistry and materials science. It is already possible today to grow many of such materials as epitaxial films and heterostructures, and to pattern them into nanodevices. In fact, with our capability to grow heterostructures with atomic precision, the phase space of electronic systems is becoming virtually unlimited.

In this development, not only the number of available compounds is growing, but, more importantly, also their range of functional properties. The mean-field compounds discussed above are based almost exclusively on s and p-electron systems. Thanks to transition-metal oxides and elements, electronic correlations are becoming increasingly relevant by providing completely new arenas for device applications, such as phase-transition transistors, negative capacitors and photovoltaic cells based on correlation effects, all of which are complementing and adding to mean-field Si-CMOS. Nevertheless, despite these bright prospects for device applications, we should be mindful of the leap from device demonstrations to successful products, as described by Rolf Landauer.

There is yet another issue to ponder: Exploring the periodic table to discover and synthesize materials for electronic devices will be a one-off research adventure because there is of course only one periodic table. In their ground states, chemical elements just possess a limited set of s, p, d and f orbitals and linear combinations thereof, with no additional orbitals to come. The effects we will find, the discoveries we hope to make, and the issues we will uncover regarding their applications will be what they are, as there will never be a new set of chemical elements. The periodic table is the one stomping ground we have.

The scientific community is developing increasingly refined tools to design, grow and characterize these materials with atomic precision. The accuracy and supercell size of DFT calculations are being extended to inform the design of heterostructures and devices. Handling smaller supercells, DMFT calculations allow correlation effects to be assessed. Advanced film growth techniques based, for example, on molecular beam epitaxy (MBE), pulsed laser deposition (PLD), and atomic layer deposition (ALD) are now available, and analytical tools such as scanning transmission electron microscopy (STEM) and electron energy loss spectroscopy (EELS), resonant inelastic X-ray scattering (RIXS) and angle-resolved photoelectron spectroscopy (ARPES), transport measurements and tunneling spectroscopy are being used with great success for characterization. With recent advances in stoichiometry control, substrate termination, and reflection high-energy electron diffraction (RHEED), it is now possible to grow heterostructures of complex compounds of surprising quality. Although defect concentrations and mobilities of semiconducting heterostructures are distinctly superior to those of complex oxides, according to their STEM images and EELS analyses, in terms of quality, oxide heterostructures already seem at least comparable to today’s most advanced III-V heterostructures, an achievement that was unthinkable twenty years ago (see, e.g., [18]).
effort, and time. Indeed, it is imperative that we make good, sensible use of the elements available to us.

Acknowledgements

The authors gratefully acknowledge helpful discussions with Dieter Fischer.

What theoretical approaches can provide. A perspective on oxide electronics

Mario Cuoco\textsuperscript{a,b,}\textsuperscript{*}, Jeroen van den Brin\textsuperscript{c}
\textsuperscript{a}CNR-SPIN, IT-84084 Fisciano SA, Italy
\textsuperscript{b}Dipartimento di Fisica “E. R. Caianiello”, Università di Salerno, IT-84084 Fisciano SA, Italy
\textsuperscript{c}Institute for Theoretical Solid State Physics, IFW-Dresden, Helmholtzstr. 20, D-01069 Dresden, Germany
\textsuperscript{*} Corresponding author.
E-mail: mario.cuoco@spin.cnr.it (M. Cuoco).

Abstract

Theoretical approaches and materials design strategies are discussed that can contribute in tackling present experimental and technological challenges in oxide materials with the aim of providing a perspective on the ways-to-go towards new concepts, effects and materials with an impact particularly in the area of oxide spintronics, spinorbitronics, 2D electronics, and topotronics.

Keywords: Theoretical modelling; Materials design; Oxides; Spintronics; Spinorbitronics; 2D electronics; Topotronics

1. Introduction

Developing and enhancing applications in the area of oxide electronics relies on detailed knowledge of the behavior of lattice, electronic, magnetic and orbital degrees of freedom in these materials, both in bulk and at interfaces. From the general experimental perspective two different roles are played by theoretical approaches.

The first takes experimental observations on a given material system as a starting point. Theoretical modelling and computational methodologies are then applied to explain experimental outcomes as well as provide full access to the underlying microscopic environment. Such an approach is probably the most common one in materials science. The other trajectory is inverted and employs theoretical computation and modelling to predict from an abstract starting point which materials might exhibit interesting new effects or fit better for some desired physical property, functionality and potential ensuing technology. Both directions can greatly support experimental research efforts, although with different perspectives and impact, as well as increase the success rate in the discovery of novel functional materials and novel functionalities of materials.

Since the relevant elementary length-scale in oxides is the atomic one, quantum mechanical modelling that scale constitutes a fundamental starting point. In this respect basically two approaches are pursued as illustrated in Fig. 2. The \textit{ab initio}, first principles one, aims to tackle the full many body Schrödinger equation and approximations thereof, for instance based on density functional theory [20] or via quantum chemistry approaches [21]. Pure first principles approaches have limited validity for materials in which electron-electron interactions play an important role such as a number of strongly correlated transition metal oxides, but they can be augmented by mean field, or dynamical mean field treatments of electronic correlation effects which introduces additional parameterizations which takes the methodology out of the strict \textit{ab initio} realm [22].

The complementary atomistic approaches use fully parameterized microscopic models, in particular many-body model Hamiltonians, to describe selected aspects of the lattice, electronic, magnetic and orbital degrees of freedom and their interactions [23]. Due to the restrictions of the Hilbert space, larger systems can be handled - some highly idealized strongly interacting models can even be solved exactly in the thermodynamic limit. Constructing model Hamiltonians makes the machinery of many-body physics available to materials physics.

The relevance of such models for a specific material obviously depends on the approximations made while deriving the model and the validity of the choice of the parameterization. \textit{Ab initio} methods can contribute to such parameterizations for instance via a tight-binding model of the most relevant bands in the electronic structure that enters an atomistic model Hamiltonian. Vice versa investigating different parameter regimes of model Hamiltonians can provide qualitatively new effects and properties that subsequently can be investigated quantitatively in real materials with first principles methods. In the context of the employed approximations for the solution of the model Hamiltonians, it is worth pointing out that for a single orbital many-electron problem including only the local Coulomb interaction, as described by the Hubbard model, the acquired knowledge of the ground state and excited states is much satisfactory to the extent that benchmarks of a wide range of numerical algorithms have been obtained. The application of many methods (e.g. auxiliary-field quantum Monte Carlo, bare and bold-line diagrammatic Monte Carlo, method of dual fermions, density matrix theories, dynamical cluster approximation, etc.), allowed the identification of limitations and errors in the whole phase diagram [24]. The performed analysis indeed allowed to validate methods and test also new approaches. When moving to a multi-orbital low-energy description which is closer to the description of realistic materials, one does not have the same level of knowledge and benchmarking of the above mentioned different methodologies for the single-orbital Hubbard model. The DMFT (Dynamical Mean Field Theory) is one of the most common and powerful approach to investigate realistic Hamiltonians which are constructed through density-functional theory (DFT) in DFT + DMFT methods. While extensions of DMFT point to cluster approaches in real or momentum space, which become critically relevant especially in low dimensions, the computational time can increase very quickly with the number of orbitals and sites in the unit cell.

![Fig. 2. Schematic representation of the relations between different theoretical approaches and strategies in relation to experimental research efforts and applications for oxide electronics.](Image)
Moreover, the key bottleneck for the computational success of such extensions is represented by the lack of fast and efficient quantum impurity solvers. A comprehensive insight into the ground state and excited states properties of multi-orbitals many-body problem still sets a challenge to date.

As schematically described in Fig. 2, a number of macroscopic physical properties of oxides can be directly determined by the ground state atomistic electronic structure - e.g. stability, elasticity and dynamics of the lattice, magnetic ordering, band offsets, electronic instabilities such as charge ordering. However, an understanding of, for instance, the transport of electrons, collective modes (e.g. magnons) and heat requires knowledge of defects and related scattering amplitudes. A theoretical understanding of those necessitates modelling of the electronic structure at mesoscopic length-scales in order to make contact with experiment. Such approach also holds for larger, emergent patterns and defects in long-range ordering such as magnetic domain walls, skyrmonic spin textures and lattice twin boundaries.

A number of parameters that appear in mesoscopic models can often be related to first principles electronic structure calculations.

2. Theoretical approaches and functional materials design

Materials identification strategies. For the inductive ab initio approach, the growth of the computational power is paving the way towards a substantial transformation of how materials science can be handled in the future. Main stream ideas in the last decade suggest that instead of continuing to synthesize new materials based on an experimental trial-and-error explorative synthesis approach, the computer modelling in synergy with machine-learning can provide interesting alternatives by generating candidate materials with specified functionalities.

From a general point of view, the development of new compounds relies on factual synthesis/growth of a material system which not only requires knowledge about stable structures and their properties, but also about the synthetic path that is required to produce such a system. An absence of a practically achievable synthetic path poses an obvious restriction. Nevertheless, using a systematic search without relying only on a trial-and-error laboratory process allows for a number of new material identification strategies: simulation-based predictions of functional properties of known structures and the determination of the structure-property relationship from experimental data, data mining on functional properties of known structures and the determination of the desired global low-energy structure which sets the main computational bottleneck. If local structural stability and thermodynamic stability for a given composition are achieved, one still needs to determine whether it tunes the low-energy Hamiltonian along the desired direction concerning the target phenomenology-gies related with formation/evolution of electronic transport and spin-charge-orbital ordered patterns. These final steps presently set the forefront challenges with lots of room and needs for new conceptual and computational approaches.

Case studies attempting high-throughput calculation for correlated materials are not much common. For instance, a representative and paradigmatic application used as a descriptor the correlation between the superconducting critical temperature ($T_c$) and the charge-transfer energy between the ligand and the transition metal element using DMFT combined with LDA, aiming to find new superconductors. To limit the huge amount of possible combinations, the applied strategy was to move in a restricted family of high temperature superconductors selecting both CuO as building blocks and singling out the Hg in the chemical formula, being the Hg the element for which the highest critical temperature is obtained in the cuprates family. The study identified layered oxysulfides [29,30] as a potential new family of high temperature superconductors.

Such example underlies the main limitations which mine the predictive power of a high-throughput calculation for correlated materials. Basically, often the phenomenology of the targeted interacting phases is not fully settled, thus indicators cannot be simply identified, and, moreover, small variations in the parameters space can lead to significantly different quantum states, thus quickly making the combinatorial search fast growing. Hence, on the basis of the status quo, only local exploration in the parameters space and elemental phenomenology seem to be currently achievable.

Moving further along this direction, one clearly foresees the fundamental difficulties that electronic correlations cause for a fast and reliable fully data-driven search are evident when considering the question what theory can deliver for the class of oxide materials based
on transition metal elements, including their combination (e.g. heterostructures or superlattices), and regarding applications for electronics and information technology based on the manipulation of spin-orbital-charge degrees of freedom as classical (i.e. bit) or quantum (i.e. qubit) information carrier.

In this context, an alternative starting point to face these issues is the identification of specific needs for theory given the experimental status quo and likely future directions. As such one has to establish the key demands for a materials search and for core quantities to be computed. For instance, in the area of electronics there is a strong motivation to go beyond the silicon-based CMOS technology, which is coming to the end of its technological roadmap [31]. The practical challenge in this context is to overcome the use of the electron’s charge as information carrier and element for Boolean logic operations. The societal and technological impulse for this arises from the recent dramatic developments in information technology creating a demand for even larger capacity data storage, faster data processing and lower energy consumption. Current proposals include the concept of employing the electron’s spin and orbital degrees of freedom (e.g. in spintronics and spin-orbitronics) as a fundamental element for developing faster progress in storage, transport and manipulation of information. The infant technologies of spintronics, spin-orbitronics, 2D electronics and topotronics are therefore considered in more detail in the following within the context of oxide materials.

Spintronics. When dealing with spintronics one is facing a large variety of effects, phenomenology and materials geared at controlling and manipulating the electron spin on a nanoscopic length-scale. The overall control and manipulation of the electron spin in spintronics devices includes many different aspects, namely spin injection-detection, spin-accumulation, spin-transfer, spin-damping, spin-resonance, spin-interference, and pure quantum spin-design. These elements generally set the basis for the development of the next generations of spintronic devices [32]. An immediate observation is that the foundational spintronic elements are interrelated to many physical aspects of the building block materials and a systematic search is not simply feasible for all the involved mechanisms and functionalities. Selecting some of the main spintronic requirements can make the task more feasible.

Naturally, ferromagnetic and, more recently, antiferromagnetic materials play a key role in spintronic devices, having the magnetization manipulated with currents or external fields. For most of the functionalities of the spintronic devices it is relevant to employ ferromagnetic materials with (i) low magnetic moment, (ii) high Curie temperature for room temperature implementations, (iii) high-spin polarization (half-metallicity), and (iv) low Gilbert damping.

What theory can do to support the identification of optimal materials or to fully comprehend the key physical properties of the targeted magnetic materials? Considering the points (i)-(iii), approaches based on density functional theory can provide accurate estimate for the magnetic ground state energies, magnetic moments and the overall spin-polarization. Then, taking into account the magnetic ground state configurations one can map (for instance by spin-polarized relativistic KKR calculations [33,34]) the system into an effective classical Heisenberg Hamiltonian and obtain the magnetic temperature (e.g. via Monte Carlo simulations). A successful demonstration of such approach has been recently reported for screening antiferromagnetic Heusler compounds with a Néel temperature above room temperature [35] and similar applications can be found in the realm of oxides as well.

For instance, concerning the growing field of antiferromagnetic materials for spintronics, antiferromagnet with an asymmetric semiconductor behavior have been theoretically proposed in the form of double-perovskite (e.g. $A_2B_CMO_6$ with $A$ = Ca, Sr, Ba; $M$ = Ru, Os). Such hybrid oxide solutions are representative of a more general scheme for designing the properties of oxide materials by combining the different atomic properties of 3d and 4d/5d transition elements. These double-perovskites have the potential to behave like magnetic semiconductors exhibiting both room-temperature magnetic ordering and large spin polarization at the transition metal sites. Remarkably, since the magnetic moments are carried by non-equivalent transition-metal elements, they can lead to reciprocal cancellations through antiferromagnetic coupling. Here, the design roles underlie the fact that strong antiferromagnetic super-exchange interaction can typically lead to materials with high magnetic ordering temperature while the presence of inequivalent transition metals in such hybrids can lead to tunable spin-polarization of electronic states close to the Fermi level.

More difficulties emerge when facing the problem of the spin-dynamics and to predict the spintronic characteristic length or time scales which enter in the transfer of spin-polarization. Furthermore, a spintronic device is basically made of two- or three-terminals. Hence, one has to deal with junctions and interfaces involving both magnetic and non-magnetic materials. The knowledge and prediction of the interfacial properties are then highly desirable for the experimentalists. However, from a theoretical point of view, already the analysis of the electronic structure requires the use of large unit cells that strongly bounds the possibility to compute realistic structures involving different magnetic layers. In this respect, the further determination of the electronic transport properties are computationally demanding and often can be considered only within suitable approximation and effective modelling that include the scattering due to the electron-electron interaction, electron-phonon, or with impurities and other types of defects. In other words, the computational complexity of the quantum problem becomes quickly very high and not tractable.

Another critical point is that in transition metal oxides an electronic reconstruction commonly takes place at the interface, that is, a formation of interface states which differ significantly from those in the adjacent bulk occur, thus adding inhomogeneities and extra quantum degrees that are hard to be handled especially if Coulomb interaction or electron-phonon coupling gets more and more relevant. In practice, for oxide heterostructures the predictive power for spintronic material applications based on electronic interface design is limited and an unbound computational effort is required to get accurate results concerning the electronic spectra and the emergent long-range orderings at the interface.

Another critical point is that in transition metal oxides an electronic reconstruction commonly takes place at the interface, that is, a formation of interface states which differ significantly from those in the adjacent bulk occur, thus adding inhomogeneities and extra quantum degrees that are hard to be handled especially if Coulomb interaction or electron-phonon coupling gets more and more relevant. In practice, for oxide heterostructures the predictive power for spintronic material applications based on electronic interface design is limited and an unbound computational effort is required to get accurate results concerning the electronic spectra and the emergent long-range orderings at the interface.

Spin-orbitronics. Conventional spintronic devices make use of the exchange interaction between conduction electron spins and local spins in magnetic materials to create spin-polarized currents or to manipulate nanomagnets by spin-transfer from spin-polarized currents. A novel direction of spintronics exploits the spin-orbit coupling (SOC) generally in nonmagnetic materials instead of the magnetic exchange interaction to generate, control and detect spin-polarized currents. This approach opens the way to the spin-orbitronics where the SOC is the main physical ingredient to design devices that have the basic property to function with nonmagnetic materials and without magnetic fields. Thus, the first crucial need-issue in this framework is to understand the nature, strength and tunability of the SOC. There are three fundamental SOCs: (i) the atomic interaction, (ii) the Rashba-type spin-orbit coupling due to inversion symmetry breaking at the surface or interface in heterostructures, and (iii) the Dresselhaus like interaction arising from the inversion asymmetry in the bulk host material. It is worth noting that typically in oxide materials it is the conspiracy of the atomic spin-orbit interaction with the inversion symmetry breaking sources that generates both Rashba and Dresselhaus like emergent terms within the
multi-orbital electronic configurations close to the Fermi level. The role of these interactions are currently transforming the physical scenario especially in the direction of having novel functionalities and high-performances in the field of information technology. For instance, heavy metals with large atomic spin-orbit coupling can allow a significant charge-to-spin current conversion by exploiting spin Hall and Rashba-Edelstein effects. Remarkably, the emergent field of the spin-orbitronics is strongly boosted by the potential of employing the non-trivial topology of the band structure of the topological insulators or semimetals, or the spin-textures in magnetic materials. Progress in this direction is at present motivated by the discovery of many topological materials with unconventional spin-orbit driven magneto-transport properties as well as the achievement and control of topological magnetic textures (i.e. skyrmions) in multilayers made of heavy metals and magnetic layers.

Since spin-orbitronics relies on the control of pure spin currents through spin-orbit effects, the spin-orbit coupling due to the inversion symmetry breaking at the interface of oxides, also due to its gate tunability, is of prime importance. In conventional Rashba spin-orbit coupled two-dimensional electron gases (2DEGs), the flow of a charge current results in the creation of a non-zero spin accumulation due to uncompensated spin-textured Fermi surfaces.

Recently, a very large, gate-tunable inverse Rashba effect - i.e. a spin-to-charge conversion - was observed in NiFe/AlO/SrTiO heterostructures, with an efficiency more than an order of magnitude larger than at the Ag/Bi(111) interface or in spin Hall materials [36].

Such experimental findings suggest that oxide interfaces have a potential for spintronics and spin-orbitronics, both for the generation or detection of spin currents through direct or inverse Rashba effects, and for their electrical modulation. From a theoretical perspective, it would be desirable to quantitatively predict the spin-Hall and inverse spin-Hall transport properties and the resulting spin-to-charge or charge-to-spin conversion efficiency which is typically parameterized through the inverse Rashba length. In a Rashba spin-orbit coupled environment this length corresponds to the distance the electrons travel in the 2DEG between collisions while keeping spin-momentum locking.

Here, the mechanisms and the phenomenology are not yet completely settled. Furthermore, although a simplified picture would point to a single-orbital description with Rashba spin-orbit coupling, the experimental findings seem to clearly indicate a need for a multi-orbital modelling and the inclusion of charge reconstruction of spin-orbit effects on both charge and spin transport. In this framework, both ab initio and atomistic modelling are needed to get deeper insight into the problem and, if successful, can contribute to figure out the impact of 2DEG and oxide interfaces at large for spin-orbitronics purposes.

2D oxide electronics. Engineering a two-dimensionally confined electron liquid is the basis of 2D electronics and there are different approaches to realize such in oxides. One direction lies on the formation of a high-mobility 2DEG at TiO2-terminated interfaces in the polar/non-polar LaAlO3/SrTiO3 (LAO/STO) heterostructures. The discovery of conduction and a high-mobility 2D electron gas [4] at the interface between these bulk insulators has been, without doubt, a milestone for supporting the engineering of nanoscale devices that could outperform with respect to other solid-state platforms. The progress in creating conducting transition-metal-oxide interfaces has opened a new field of materials and physics research especially due to the continuously increasing level of experimental control. The prototypical LAO/STO interface put into the limelight the potential and the limitations of this class of materials for both fundamental and application-oriented purposes.

One fascinating aspect of 2DEG is that many mechanisms cooperate/compete in the formation of the interface electron gas. Indeed, on a general ground, there are at least five main mechanisms that can be considered for the formation of a 2DEG at the interface of an oxide heterostructure. (i) By band bending a suitable modulation of the doping and a 2DEG can be achieved in a hybrid system made of a wide band-gap and a narrow band-gap insulator. There, n-type doping in the wide band-gap subsystem can introduce additional electrons that get transferred from the conduction band of the wide band-gap to that one of the interfaced narrow-band-gap. For instance, a typical example of 2DEG based on this mechanism is the AlGaAs/GaAs hybrid system. (ii) A polar discontinuity at the interface between a polar and non-polar layer causes the electronic reconstruction with an ensuing charge transfer that creates the 2DEG at the interface (e.g. in LAO/STO or similar oxide interfaces). (iii) A 2DEG is due to a discontinuity of the strain potential which can yield an internal electrostatic field and polarization charges bound at the interface (an example is provided by ZnMgO/ZnO heterostructure). (iv) Oxygen vacancies and cation intermixing lead to extra electron/hole dopants at the interface and are commonly present even for extremely controlled epitaxial growth. They are responsible for the creation of effective electron or hole doping that then leads to the 2DEG. Finally, (v) strain induced polarization and piezoelectric polarization effects in non-polar/non-polar interfaces can drive the formation of bound interface charges. Such effects can be at work in non-polar/non-polar oxide heterostructures of the type CaZnO/SrTiO. Although a large number of theoretical investigation of the above mechanisms have been carried out, a comprehensive understanding of their interplay is not yet completely settled and again points to a successful integration of ab initio.

Novel technological directions that bridge the 2D oxide electronics with the spintronics and spin-orbitronics areas require the design of new interfaces which display a 2DEG with not only high conductivity but also high spin-polarization. There are two main ways to generate spin-polarized 2DEG: (i) interfacing Mott insulator to band-insulator as for instance for the GeTe interface interfacing SrTiO [37] and (ii) sandwiching band insulators with few layers of a magnetic material, as for the case of the SrTiO/ErTiO/LaAlO heterostructure [38].

These two pathways towards spin-polarized 2DEG are fundamentally different. First of all, by employing magnetism out of a Mott scenario, one deals with a Mott insulating phase in the heterostructure with strong spin-orbital entanglement in the low energy and a high level of complexity when considering the behavior of intrinsic or extrinsic electron-hole doping effects at the interface. Indeed, the Mott state is generally accompanied by spin-orbital orders whose doping evolution can lead to exotic phase of matter and bad metallic configurations. Moreover, the Mott transition is generally of first-order type and thus naturally leads to phase separation scenario in a window of doping around the critical Coulomb interaction. Hence, spin-orbital reconstruction, intrinsic inhomogeneities and electronic phase separation have to be addressed at the band-Mott insulating interface. On the other hand, for interfaces between band insulators inhomogeneities are instead tied to the presence of defects or impurities rather than to electron-correlation effects, thus providing a very different microscopic environment for setting the physical nature of the 2DEG. Also for this case, due to the complexity behind the inhomogeneous electronic quantum problem [3], the most suited theoretical approaches are those based on atomistic modelling combined with disordered statistical schemes of computation.

Moving along akin directions for designing spin-polarized 2DEG at the oxide interfaces, a geometric perspective points to the possibility of designing heterostructures with other oxide crystal structures apart from the perovskite one. For instance, an example of stabilizing a spin-polarized 2DEG is given by the case of interfacing two insulating spinel oxides CoFe2O4 and MgA12O4 [39]. Due to the cation distributions in these two oxides and their polar character, a charge discontinuity can occur at the interface, inducing the appearance of a polar catastrophe scenario similar to that proposed of LAO/STO interface, with an insulating-to-metal transition appearing at a critical thickness of CoFe2O4. Due to the magnetic exchange and the energy mismatch at the interface the transfer of charge is expected only for minority spin electrons, and due to the high magnetic temperature of CoFe2O4, the 2DEG may remain spin-polarized up to room temperature.
Another path for developing 2D oxide electronics is to go down to one-atom-thick layer structures, inspired by graphene, a one-atom-thick carbon sheet with a honeycomb structure that can be obtained by exfoliation of bulk graphite or through a layer-by-layer deposition. As for graphene a single-layer of transition metal oxide may maintain its structure because of strong covalent bonding within a layer, whereas weak van der Waals forces acting between charge-neutral layers can allow isolation. Such approach has been successful for 2D hexagonal boron nitride (h-BN) as well as for 2D metal oxides [40,41], including perovskite-based oxides, that, so far, reported unique physical properties.

In contrast to the above cases, non-layered materials with wurtzite and cubic crystal structures were theoretically predicted to form 2D layers, which widens the range of potential 2D materials to be fabricated. Recently, graphene-like 2D zinc oxide clusters and 2D iron clusters with a square lattice have been observed. These oxide materials can lead to novel functionalities in 2D electronics especially for developing new sensors [42]. On the theoretical side, affordable key issues concern with the stability of the single sheets as well as the prediction of which structures are favored and what type of crystal symmetries can be realized.

For the above systems it would be highly desirable to have a full microscopic control on the electronic mobility as well as on its spin dependence. This is a formidable task if compared to conventional semiconductors, because, apart from requiring a knowledge of the main scattering processes occurring in the material, it involves detailed knowledge of the charge and orbital reconstruction occurring at the interface as well as of other possible broken symmetries. Whether the layers at the interface are conductive or insulating strongly depends on growth conditions and the various proposed mechanisms for conduction (i.e. oxygen vacancies, polar catastrophe, cation intermixing) are very difficult to fine tune due to their intrinsic nanoscale nature.

Addressing the resulting complex problem theoretically requires ab initio atomistic modelling in an intrinsically inhomogeneous microscopic quantum environment. In this respect, although this field reached a high degree of maturity both on the side of the fundamental physics and device engineering a complete control of all the functional properties is still far away thus providing an obvious future research direction for these hybrid systems. Material-specific theory can be fruitful in this context by strengthening the synergy between model Hamiltonian and ab initio approaches especially when dealing with the engineering of correlated electron physics at the nanoscale.

Many-body models become particularly relevant when exploring parameter regimes such as Mott insulating phases, Kondo lattices, unconventional superconductivity and topological non-trivial states of matter. Still, as mentioned above, the emerging high-throughput computational materials design methodology can enter here to accelerate the process of discovery for new materials. In particular, interesting attempts to systematically search novel perovskite-based 2D electron gasses have led to the identification of a group of combinatorial descriptors including the polar character, lattice mismatch, band gap, and band alignment between the polar perovskite oxides and the STO substrate. Then, on the basis a perovskite-oxide-oriented materials repository using the automatic framework AFLow that is designed for high-throughput materials discovery, a large number of candidate perovskite oxides have been singled out for further exploring the possibility of producing 2D electron gas at the interface [43].

**Topotronics.** A topological phase is generally marked by quantized macroscopic observables which are substantially unaffected by changes in the local environment as for the well-known example of the quantum Hall effect. Topotronics aims at developing concepts and devices which implement topological materials and their quantized observables to store or transport information. In this respect, for instance, topological insulators [44,45] have been viewed as a special class of spintronics materials, with their surface states used for pure spin generation and transportation due to their topological protection to backscattering.

Still in order to achieving topological behavior not all the perturbations are generally allowed, thus symmetry protection and topological phases have to be addressed together. Strong topological phases commonly arise from the protection of internal symmetries, such as time-reversal, particle-hole, or chirality, while translation or point group symmetries of the lattice can give rise to so called weak and crystalline topological states. Weak or strong character indicates if the protecting symmetry can be achieved in a realistic configuration. For instance, while time reversal symmetry can be controlled by avoiding magnetic impurities, for crystalline topological phases the translation symmetry can be broken by impurities in a crystal. However, even when disorder breaks the lattice symmetry, a topo-logical phase can still persist if the system remains symmetric on average. General theories of topological classification that take into account the symmetries of weakly interacting electron systems have been successfully developed in the last decades. The ground state can be then classified [46] according to certain topological number depending on its dimensionality and symmetries. Such approaches combined with ab initio and microscopic modelling can provide a clear guide to single out a large variety of topological materials.

While the number of topological phases proposed in theory is still growing, experimental confirmations are mainly limited to the systems of groups IV-VI elements. In oxides various topological states matter have been predicted as well [47–51]. However, indications of their existence are so far still lacking and one of the fundamental question to face is why topological phases have not been detected in other abundant materials such as oxides. There are different design paths which have been indicated so far. One common direction focuses on the exploitation of large amplitude of the atomic spin-orbit coupling thus the target is centered on oxide-based on 4d and 5d elements. Another one points to the geometry of the superlattices [52], as a key building block, by growing oxide heterostructure in a way that the connectivity of the d-orbitals can mimic an effective honeycomb lattice (e.g. in the case of thin films growing along the [111] direction). Such geometric approach for the superlattice fabrication is motivated by the search of semimetallic phases from which one can drive topological insulating configurations by applying suitable perturbations with specific symmetries.

This line of search is significantly growing in the last years and can set the stage for relevant future technological directions which can employ topologically protected gapless electronic phases. Indeed, after the great impact of topological insulators, there has been a significant expansion towards metals and semimetals as well quantum materials combining topological and conventional forms of order. In topological semimetals conduction and valence bands cross at certain points or lines in the Brillouin zone and the crossing are protected by a symmetry of the system and the ensuing topological invariants.

In oxides, a promising direction is the search for new variants of semimetals that combine Dirac fermions physics, magnetism and superconductivity due to the common presence of these ordered phases in most of the transition oxide materials. For instance, while most of the currently known semimetals are non-magnetic, antiferromagnetic ones can be obtained where both time and inversion are broken while their combination is kept or due to the combination of internal symmetries with non-symmorphic transformations. Although at the beginning, such issues can be well addressed by combining ab initio and many-body modelling including DFT + schemes and other correlated schemes of computation.

Along this path, it is worth to mention that the combination of Dirac semimetals and superconductivity is one possible way to design topological superconductivity exhibiting edge Majorana bound states. In oxides, there is a large amount of potential materials and hetero-structures in which Majorana modes or topological superconductivity might be achieved and that thereby can be relevant for topological quantum computation. One of the main obstacle is the control of the interface properties as already mentioned above. Currently, the use of 2DEGs at polar-non polar interface seems to represent an interesting
quantum phase transition. There both the gate tunability of the Rashba spin-orbit coupling and the coexistence of superconductivity and magnetism provide the basic physical ingredients to obtain a topological superconducting phase. At the present stage, theoretical predictions indicate that topological superconductivity can be obtained in 2DEGs but there are no observations that can confirm its existence. While this field is definitely at its infancy, it can provide significantly potential of application for oxide materials. From a theoretical point of view, apart from the potential use of combined atomistic and ab initio approaches that can also integrate the superconductivity, there are still open issues and limited means to fully access the nature of the electron pairing and the resulting topological behavior.

There is also quite some room for new concepts and effects which are still largely unexplored due to the additional presence of Mott insulating phases and the interplay with electron-electron correlations, as for instance the case of topological phases of matter where electron correlations yield fractionalization inside the topological insulator [53]. In this framework, referring to the scheme of Fig. 2, most of the advancements are driven by many-body microscopic and mesoscale modelling while full ab initio schemes are still beyond the current computational and conceptual state-of-the-art. In other words, while examples of interacting driven topologically non-trivial phases have been proposed including a set of possible Hamiltonians for their description, identifying which specific materials or structures that can exhibit such phases is still a major challenge.

The capacity to simulate, understand and predict the properties of non-collinear magnetic phases (e.g. skyrmionic or spiral phases), rely on a combination of various approaches including ab initio, model Hamiltonian and mesoscale modelling. This is mainly because full ab initio approaches become quite limited once dealing with large unit cells and inhomogeneities in the magnetic textures. Following the scheme of Fig. 2, skyrmionic phases of matter are a paradigmatic case for which one needs to fully employ the flow from ab initio to microscopic and mesoscale modelling in order to search for materials and screen the most prominent ones for applications. From a technological perspective, taking into account the so far synthesized materials, the main obstacles for the realization for highly energy-efficient device applications is that skyrmions have mainly been observed only for few specific materials at low temperatures for which one typically has to employ external magnetic fields. Furthermore, fast and energetically efficient current driven motion of single skyrmions is still to be demonstrated experimentally.

3. Concluding remarks

The complexity of the oxides clearly sets a need to strengthen the interrelation between the ab initio, atomistic modelling, and mesoscopic modelling blocks. While each processing part is quite mature with well-defined internal structures and successfully generating outcomes, a robust cross-length scale modelling approach rooted in first principles atomic calculations is far from so - achieving it can serve as a dot at the horizon for the present day computational and theoretical materials science research effort.

Perspectives for applications of ultimate (atomic) control of oxide films using PLD

Gertjan Koster *, Mark Huijben, Guus Rijnders

MESA+ Institute for Nanotechnology, University of Twente, 7500 AE Enschede, Netherlands

* Corresponding author.

E-mail: g.koster@utwente.nl (G. Koster).

Abstract

Whether oxide thin films prepared by atomically controlled pulsed laser deposition (PLD) are ever going to be used in real applications depends on several factors. Most importantly, one needs to establish a technology platform that allows these reconfigurable complex oxide materials in thin film form to be integrated with current industrial processes and standards, such as large area and the use of technical substrates (e.g. silicon). Three key ingredients are currently in need of attention in order to access the full potential of these reconfigurable materials: atomic control at industrial dimensions, detailed understanding of atomistic surface processes under industrial process conditions, and demonstrations of function-ality on technical substrates.

Keywords: Complex oxides; Thin film growth; Pulsed laser deposition; Industrial scale; Silicon

1. State of the art

Starting in 1988, pulsed laser deposition (PLD) has proven to be a valuable thin film deposition technique to create heterostructures of complex oxide materials (for example, superconducting YBa2Cu3O7 [54] and piezo/ferro-electric Pb(Zr,Ti)O3 [55]) with control on the atomic level. The highest level of control is achieved when one shows the ability to transfer material stoichiometrically from a mul-ticomponent target to a growing film, which has been demonstrated with PLD under ideal circumstances.

Controlling stoichiometry in a reproducible manner is very important for complex oxide films, such as Pb(Zr,Ti)O3 (PZT), because their physical properties strongly depend on the precise control of the chemical composition. The subsequent development of an all-oxide piezoMEMS technology was achieved by including conductive-oxide SrRuO3 electrodes which are well-known for their chemical stability [56]. A dramatic improvement of the fatigue resistance of PZT thin film devices was found, which is very important for the performance of

Fig. 3. (La,Sr) MnO2/BaTiO3/Co-Au ferroelectric tunnel junction-Dependence of the junction resistance measured at Vread = 100 mV after the application of 20 ns voltage pulses (Vwrite) of different amplitudes. The different curves correspond to different consecutive measurements, with varying maximum (positive or negative) Vwrite. Taken with permission from [58].
piezoMEMS with PZT films as the active layers [57]. Another example is use of the ferroelectric compound BaTiO3 in so-called ferroelectric tunnel junctions. In this device, which has been demonstrated for usage as a memristor, the tunnel resistance can be manipulated by tuning the ferroelectric polarization in the thin BaTiO3 layer, see Fig. 3. Such a device would only work properly if the interfaces in the heterostructure are of sufficient quality in addition to the requirement of a low leakage ferroelectric layer, which is not trivial at the reported thicknesses.

Worth mentioning is the LaAlO3-SrTiO3 heterostructure, where at the interface between these insulating materials a two-dimensional electron liquid (2-DEL) forms [30], exhibiting interesting phenomena, such as superconductivity and magnetism. While this example is probably less important as potential functional material, it was demonstrated that both small variations in the composition of the LaAlO3 layer eventually determine the conductive and/or magnetic state measured at the hetero-interface, as well as the need to control the precise terminating layer at the hetero interface [59]. Many more examples are available in literature and some of these are discussed in other sections of this road-map.

Besides having to aim for perfect stoichiometry as well as crystallinity, epitaxially grown materials have been shown to offer the possibility to modify or reconfigure the properties of materials to match that of an application.

Two complex oxide systems beautifully exemplify this future potential. Recently, it was demonstrated that the magnetocrystalline anisotropic properties of SrRuO3 in a heterostructure with Ca0.5Sr0.5-TiO3 (0-4 monolayers thick) grown on a GdScO3 substrate can be modified [60]. The Ru—O—Ti bond angle of the SrRuO3/Ca0.5Sr0.5TiO3 interface was engineered by layer-by-layer control of the Ca0.5Sr0.5TiO3 layer thickness, as can be seen in Fig. 4 where the change in lattice parameters show the symmetry change as a function of the thickness by which the magnetic anisotropy in the entire SrRuO3 layer was tuned. In another system a similar effect was seen, epitaxial La0.67Sr0.33MnO3 films can be reconfigured to change the magnetocrystalline anisotropy by carefully selecting specific ultrathin buffer layers of SrTiO3 [61], which reduce the propagation of the octahedral tilt system of the substrate into the thin film on top.

Finally, in another system, it has been shown that piezo-electric-like properties can be achieved in epitaxial heterostructures of the non-ferroelectric and non-piezoelectric SrTiO3 material [62], where reconfiguration of properties occurs due to the application of a strain gradient in the film by means of the bending of a cantilever, see for a schematic in Fig. 5.

Here, we would like to emphasize again that in the examples above, PLD was used to fabricate one or more of the functional layers. All of the aforementioned examples have used single crystalline, small scale substrates under ideal laboratory conditions. An important step towards applications would require the development of a technology combining complex oxides with silicon on larger scales. Extensive literature exists on this issue and several reliable solutions are used or are being developed [63,64].

![Fig. 4. Lattice constants of the SrRuO3 layer as a function of the Ca0.5Sr0.5TiO3 layer thickness, revealing that the monoclinic SrRuO3 layer changes into the tetragonal form when the Ca0.5Sr0.5TiO3 layer thickness is greater than three monolayers. The lattice constants were obtained from the (620)o,(260)o, and (442)o SrRuO3 reflection positions. Taken with permission from [60].](image)

![Fig. 5. Schematic comparing flexoelectric actuation and piezoelectric bimorph actuation in nanoscale actuators. In a piezoelectric bimorph actuator, a homogeneous mechanical strain is generated on application of an electrical voltage to the piezoelectric layer. The mechanical clamping induced by the non-piezoelectric layer creates a strain gradient across the structure, converting the piezoelectric strain into a flexural motion. On the other hand, any dielectric sandwiched between the electrodes can, in principle, act as a flexoelectric actuator. In this case, the bending moment arises from a symmetry-breaking strain gradient generated at the unit cell level. Taken with permission from [62].](image)
2. Future prospects/current and future challenges

Reconfigurable functionalities in epitaxial thin films of complex oxide materials, as was exemplified above, result from the controlled crystallinity and defect structure in combination with selected templates and buffer layers. A crucial next step will be the establishment of an industrial platform for novel electronic devices based on reconfigurable thin film oxide materials grown by PLD, which delivers performance and functionality that could enhance state-of-the-art CMOS.

The pulsed laser deposition technique used broadly in research laboratories worldwide is conceived as relatively simple, mainly because of the fact that the heating source for evaporation (ablation), i.e., a powerful laser, typically of the excimer type, is located outside the process chamber. It derives its popularity thanks to its versatility. However, the actual processes and mechanisms that take place are highly non-trivial and dynamic: when the laser beam hits the ceramic target material ablation takes place. The superheated plasma will subsequently expand, cool and interact with the background gas by collisions and ultimately the (reacted) species will arrive at the substrate surface, where surface processes related to local inhomogeneities and chemistry and composition of the thin films, which is essential in obtaining the desired control of layer thickness at the atomic-layer level. In recent years, much has been learned about the role of arriving (reacted) species in the deposition process and the observed type of crystal growth on small scale wafers [65–67] as well as the intricate nature and composition of arriving species by means of laser induced fluorescence detection [68,69], see for example Fig. 6 for the case of SrTiO\textsubscript{3}. Note that in the case of small wafers, only a small portion of the particles in the plasma plume is collected, usually those originating from the center part of the plume.

Techniques based on scanning PLD should be the next step in the technological application of reconfigurable oxide materials, for more industrial CMOS compatible production processes (and standards) on technical substrates (e.g., silicon) of sufficient wafer sizes. The key difference with the examples from the laboratory for small scale is that now particles originating from the entire plasma plume are collected onto the large wafers. Unfortunately, fundamental knowledge is lacking on the relation between a scanning plasma and the achieved structure and composition of the thin films, which is essential in obtaining the observed reconfigurable functional properties in a reliable and reproducible manner. This topic is in need of direct attention in order to make the important step towards integration of complex oxide materials made by PLD with industry standard technical wafers. However, as mentioned earlier, until now such control was only possible on laboratory scale on single crystalline wafers.

The success of being able to use and reconfigure complex oxide thin films critically depends on being able to atomically control the pulsed laser deposition technique, e.g., by virtue of \textit{in situ} diagnostic tools (e.g. RHEED and LIF) at the industrially imposed boundary conditions. In terms of deposition tools, important progress has recently been made in terms of plasma scanning techniques on large wafers (e.g. Solmates B.V.) and compatibility with silicon. More activity is expected to incorporate diagnostic tools into the large area deposition tools. One may expect a more detailed understanding of the role of the plasma species and their distribution, the implications of scanning on surface kinetics and stoichiometry and with the obtained knowledge more successes in demonstration of functionalities on industrially relevant platforms.

3. Concluding remarks

Understanding the thin film properties of complex oxides on single crystal (research) substrates is a very active field of research and offers real potential. However, three key technological ingredients are missing in order to access the full application potential of these materials: 1. Full control of epitaxial thin film growth at sizes suitable for industrial applications, in particular in terms of a scanning plasma plume during PLD. 2. In case of scanning PLD, the detailed understanding of atomistic surface processes related to local inhomogeneities and chemistry versus global stoichiometry 3. First demonstration of reconfigured complex oxide functionalities on technical substrates such as silicon wafers, comparable to what was already achieved on small single-crystalline oxides. Progress in these three areas plus the exploration of new materials systems should go hand-in-hand.

Acknowledgements

The authors acknowledge financial support from the Netherlands Organization for Scientific Research (NWO).

Oxide MBE and the path to creating and comprehending artificial quantum materials

Darrell G. Schlom*a,b,*, Kyle M. Shenb,c

*Department of Material Science and Engineering, Cornell University, Ithaca, NY 14853, USA
*bKavli Institute at Cornell for Nanoscale Science, Ithaca, NY 14853, USA
*cLaboratory of Atomic and Solid State Physics, Department of Physics, Cornell University, Ithaca, NY 14853, USA
* Corresponding author.
E-mail: schlom@cornell.edu (D.G. Schlom).

Abstract

Molecular-beam epitaxy (MBE) is renowned for preparing semiconductor heterostructures with high purity, high mobility, and exquisite control of layer thickness at the atomic-layer level. In recent decades it has become the definitive method for the preparation of oxide heterostructures as well, particularly when it is desired to explore their intrinsic properties or use the unparalleled synthesis precision of
MBE to break symmetries to expose hidden ground states and emergent phenomena. The demonstrated combination of oxide MBE with angle-resolved photoemission spectroscopy (ARPES) enables a powerful path forward to create, study, and provide relevant feedback to theory that will lead to an improved understanding of artificial quantum materials.

**Keywords**: Oxide molecular-beam epitaxy; MBE + ARPES; Emergent phenomena; Quantum materials

1. **State of the art**

When it comes to growing oxide films with high purity, high mobility, superb perfection, and exquisite control of layer thickness at the atomic-layer level, nothing comes close to oxide MBE [70]. Transport properties provide a sensitive test of the quality of electronic materials. Recent examples of the unparalleled transport properties of oxide films grown by MBE include the reigning mobility records for ZnO [71], SrTiO$_3$ [72], and BaSnO$_3$ [73], the observation of the quantum Hall effect in SrTiO$_3$ [74] and the fractional quantum Hall in ZnO [75], as well as films of SrVO$_3$ [76], CaRuO$_3$ [77], and SrRuO$_3$ [77] with the highest residual resistivity ratio, $\rho_{300K}/\rho_{4K}$. In addition, MBE has achieved the highest superconducting transition temperature in thin films of Sr$_2$RuO$_4$, a (likely odd-parity topological) superconductor that is extremely sensitive to disorder [78].

When growing epitaxial thin films, it is possible to exploit the presence of the substrate to gain access to phases that are meta-stable by changing the energetics of the thermodynamic system through a process known as epitaxial stabilization [79]. The thermodynamic system includes both the substrate and the film and energies associated with the interface between them. Judicious choice of the substrate can be used to break symmetries and even break Pauling’s first rule of crystal structures [80]. An example of this is shown in Fig. 7 for the case of a commensurate SrTiO$_3$ epitaxial film deposited (schematically) on substrates of different lattice spacings.

In bulk, SrTiO$_3$ satisfies Pauling’s first rule [80], meaning if the ions are considered as hard spheres that the titanium ion is sufficiently large that when surrounded by six oxygen ions that radius ratios are such that the oxygen ions do not overlap. This results in the titanium ion being located in the center of the resulting oxygen coordination polyhedron, which in this case is an octahedron. The energetic landscape that the titanium ion feels is shown schematically in Fig. 1(d). The titanium ion lies at the center of symmetry of the SrTiO$_3$ structure and this I symmetry precludes SrTiO$_3$ from being ferroelectric.

Now consider what happens when SrTiO$_3$ is grown commensurate on an isostructural substrate with a larger in-plane lattice constant—the case shown schematically at the right of Fig. 7(a). The octahedral cage surrounding the titanium ion is stretched in-plane, making it so that the titanium ion has room to “rattle” inside of this cage, altering the energetic landscape of the titanium ion to be that shown in Fig. 7(e). Such rattling, which violates Pauling’s first rule [80], also destroys the center of symmetry of the structure. An electric field can now be used to reposition the titanium ion between the two symmetry equivalent stable positions in this new ground state of SrTiO$_3$; this previously hidden ground state, accessed by breaking the symmetry of the SrTiO$_3$ through the introduction of an appropriate substrate, is ferroelectric. Similarly, if an isostructural substrate with a smaller in-plane spacing is used (the case shown schematically at the left of Fig. 7(a)), the cage around the titanium is compressed in-plane, but elongated out-of-plane, allowing the titanium ion to rattle in the out-of-plane direction. The energetic landscape for this latter case is shown in Fig. 7(c). Though simplistic, the cartoons shown in Fig. 7(a) and (c)-(e) correspond to the strain phase diagram calculated by thermodynamic analysis (Fig. 7(b) [82]) and have been verified by experiment [82].

The use of “active” substrates, which through strain [83] or other means break symmetries in overlying films and unleash hidden ground states, provides a huge opportunity to create and manipulate the

---

**Fig. 7.** (a) Schematic of the influence of the lattice spacing of an underlying substrate on the position of the titanium ion (shown in green) within SrTiO$_3$. The strontium ions are shown in red and the oxygen ions are located at the vertices of the coordination octahedron surrounding the titanium ion. (b) Calculated strain phase diagram of SrTiO$_3$ (after Ref. [83]). The energetic landscape felt by the titanium ion as a function of in-plane biaxial strain is schematically shown in (c)-(e), where movement of the titanium ion off center gives rise to a spontaneous polarization.
properties of artificial materials. This is because these systems can have a plethora of competing ground states close by in energy to each other. An example is shown in Fig. 8 for the case of LuFe$_2$O$_4$ [84]. According to first-principles calculations, hidden ground states with properties of far greater interest than the ground state lie just a few meV per formula unit higher in energy. Such a low-lying hidden multiferroic ground state was achieved experimentally in (LuFeO$_3$)$_{1-x}$(LuFe$_2$O$_4$)$_x$ superlattices by using the puckered motif of an underlying LuFeO$_3$ layer to break the horizontal mirror plane of the LuFe$_2$O$_4$ layer in these superlattices [84]. Another way to break symmetry is through atomically precise layering; indeed this has been extensively used in MBE-grown oxide superlattices to break inversion symmetry in perovskite [85,86], Ruddlesden-Popper [87], and brownmillerite phases [88,89]. Regardless of the means used to break symmetry and unleash hidden ground states, the result is an "artificial" material. We refer to these as "artificial" materials because they are all metastable in bulk form and only exist in artificial heterostructures.

2. Future prospects

A tremendous opportunity for materials discovery involves identifying and accessing the hidden ground states of artificial quantum materials—materials whose electronic structure and properties are not well described by density functional theory (quantum materials) and that only exist in artificial heterostructures. This materials design space is gigantic and as our interest is to identify and make artificial quantum materials having superior properties, a means to navigate this immense space is crucial. "Materials-by-design" has been shown to be an effective approach to discover materials with superior properties in cases where density functional theory provides reliable guidance. But what about those materials for which it does not, i.e., quantum materials?

In many ways this is the most exciting frontier as the properties involved—superconductivity, ferromagnetism, metal-insulator transitions, magnetoelectric multiferroics—have the potential to be extremely useful. Despite the enormous excitement and interest in quantum materials, the fundamental study and understanding of these systems has been hampered by a critical constraint: the state-of-the-art techniques developed for the evaluation of the electronic structure of bulk single crystals (such as high-resolution angle-resolved photoemission (ARPES) [90] and scanning tunneling microscopy (STM) [91]), have until recently been unable to access artificial quantum materials, since they exist only as ultrathin epitaxial films.

The direct, ultra-high vacuum connection between MBE and ARPES systems [92–97] has changed this situation. Now it is possible to measure the electronic structure of quantum materials and then subject them to systematic changes in strain [95], doping [94,95], cation ordering [92], or dimensional confinement [94] and see how the electronic structure evolves. This is an ideal situation for theorists working on quantum materials as they can see the effects of adjusting specific control parameters on electronic structure and with that input refine their theoretical methods to better describe these materials. Further, this approach frees ARPES of its traditional limitation of only being applicable to materials that (1) can be prepared as single crystals and (2) cleave. MBE + ARPES allows the electronic structure of materials that do not cleave [95,96] as well as artificial heterostructures [92] to be readily imaged. It enables a vast variety of artificial quantum materials to be made, systematically tweaked, and, most importantly, better understood.

The unusual ground states of quantum materials often emerge near phase transitions, due to competing interactions. One possible route to artificial quantum materials, schematically illustrated in Fig. 9, is to start with a phase believed to be near an instability, e.g., the metal-to-insulator instability depicted in Fig. 9, and tweak the phase by various control parameters. These control parameters include carrier density and electron overlap (as depicted in Fig. 9), which can be varied by doping and strain, but there are many additional control parameters
including dimensionality, octahedral rotations, ionic order or disorder, etc. The ability of MBE to tailor doping, layering, dimensionality, etc. in combination with active substrates enables the electronic structure of artificial quantum materials to be tuned as schematically shown by the white arrows in Fig. 9.

An example of this approach is the recent MBE + ARPES study of the spin-triplet superconductor Sr$_2$RuO$_4$ and its isoelectronic counterpart Ba$_2$RuO$_4$. Strain has been inferred to cause major changes in the electronic structure of Sr$_2$RuO$_4$ from electrical transport studies of uniaxially strained Sr$_2$RuO$_4$ single crystals in combination with theory [98], but in Fig. 10 the effect of biaxial strain on the electronic structure of Sr$_2$RuO$_4$ and Ba$_2$RuO$_4$ is directly imaged by ARPES. As can been seen, biaxial tensile strain of the RuO$_2$ plane in Sr$_2$RuO$_4$ and Ba$_2$RuO$_4$ drives the Fermi surface sheet through the van Hove singularity as it undergoes a Lifshitz transition and the Fermi surface sheet changes from an electron pocket into a hole pocket. Percent level strains can be applied to oxide thin films—far beyond where such materials would crack in bulk [83].

Seeing the effect of biaxial strain on electronic structure motivated a new strategy, one in which theoretical developments in the weak-coupling renormalization-group approach are combined with experimental developments in lattice-strain-driven Fermi surface engineering [99]. The result is that theory has predicted spin triplet superconductivity with an enhanced transition temperature to be driven predominantly by two-dimensional bands near the van Hove singularity [99]. This demonstrates the ability of MBE in combination with active substrates to alter the band structure of promising odd parity topological superconductors and ARPES to reveal the effect of biaxial strain on electronic structure and pinpoint where the $T_c$ should be highest. Figure from Ref. [97].

**Fig. 10.** (a)-(d) Fermi surface maps and (e)-(h) spectral weight along the $(0,k_y)$ direction [thick red line in (a)] for selected strain states of Sr$_2$RuO$_4$ and Ba$_2$RuO$_4$ thin films. The data in (e) were measured at an elevated temperature ($T = 100$ K) to thermally populate the states above the Fermi level ($E_F$); the rest of the data were taken at 15 K. To show the dispersion near the van Hove singularity above $E_F$, the spectral weight was divided by the Fermi function in (e) and (f). The substrate number line shows the room temperature lattice constants and strain values relative to bulk Sr$_2$RuO$_4$. This demonstrates the ability of MBE in combination with active substrates to alter the band structure of promising odd parity topological superconductors and ARPES to reveal the effect of biaxial strain on electronic structure and pinpoint where the $T_c$ should be highest. Figure from Ref. [97].

A more general materials-by-design loop for artificial quantum materials is shown in Fig. 11. Targeted materials are synthesized by MBE with feedback that is crucial to optimizing their structural perfection using a combination of X-ray diffraction, electrical transport measurements, and transmission electron microscopy. Analytic electron microscopy is a vital part of knowing what structure is actually grown, including the oxidation states of the ions, the structure and faceting present at interfaces, interdiffusion, and other key aspects of the microstructure. Sometimes growth conditions yielding optimized X-ray diffraction patterns do not correspond to targeted structures and real-space structural information with chemical specificity at the atomic resolution can be vital to understanding underlying growth mechanisms to achieve desired artificial quantum materials [100,101]. Next the electronic structure of the artificial quantum material is determined in situ using high-resolution ARPES without exposure to air, which would irrevocably contaminate the surface. This experimentally determined band structure is then compared to that expected from theory. Theory is

**Fig. 11.** An integrated materials-by-design discovery loop for artificial quantum materials.
then used to guide the perturbation of control parameters-in a way to make the electronic structure the one that is desired. The revised targeted artificial quantum material is then grown by MBE, measured by ARPES, and compared to the revised theory until the desired electronic structure is achieved.

3. Concluding remarks

A crucial difference between today’s materials-by-design approach and what we suggest here is the class of materials involved. Contemporary materials-by-design involves “conventional” materials, where first principles density functional calculations employing the local density approximation work reasonably well because electron exchange and correlation effects are minimal. Quantum materials, on the other hand, have historically been driven by experimental discovery with little or no theoretical guidance. The MBE + ARPES approach that we advocate here takes materials-by-design into the exciting realm of artificial quantum materials. Though the methodology described is quite advanced, MBE + ARPES as well as integrated design loops benefiting from MBE + ARPES are becoming increasingly accessible; for example, PARADIM users can now select among 62 elements of the periodic table for materials they wish to create and study by MBE + ARPES [102].

Acknowledgements

We acknowledge fruitful interactions with our many collaborators in the realm of artificial quantum materials, especially the groups of J.C. Séamus Davis, Craig J. Fennie, Eun-Ah Kim, Lena F. Kourkoutis, and David A. Muller. This work was supported by the National Science Foundation (Platform for the Accelerated Realization, Analysis, and Discovery of Interface Materials (PARADIM)) under Cooperative Agreement No. DMIR39918.

Nanoscale patterning of complex-oxide materials
Alexei Kalaboukhov a,*, Hans Boschker b

*aDepartment of Microtechnology and Nanoscience - MC2, Chalmers University of Technology, Göteborg, Sweden
bMax Planck Institute for Solid State Research, Heisenbergstrasse 1, 70569 Stuttgart, Germany

* Corresponding author.
E-mail: alexei.kalaboukhov@chalmers.se (A. Kalaboukhov).

Abstract

Nano-scale patterning is a pre-requisite for the realization of functional integrated devices and the investigation of size and shape-induced phase transitions in functional oxides. Complex crosstalkigraphic and chemical properties of functional oxide materials hamper the utilization of the standard silicon-based technology prompting the development of dedicated surface patterning techniques. Here, we review major achievements in the field of nano-patterning of various functional oxide materials and provide outlook for new physics of nanoscale oxide devices.

Keywords: Nanoscience; Lithography; Complex oxides

1. Nanoscale complex-oxide materials

In this contribution, we focus on the nanoscale patterning of complex-oxide thin films and heterostructures. In contrast to semiconductor technology, the driving force for smaller oxide structures and devices is only partially the technological advantage in integration density that miniaturization brings. Even more important is the new physics that is expected to emerge when nanoscale oxide structures become smaller and smaller [13,103]. Most intrinsic length scales governing the physics of the complex oxides are on the order of 10-100 nm, so sub-100 nm patterning is required. A revealing example of the physics at play is the metal-insulator transition in the manganites. Electronic phase separation occurs with a typical length scale of 70-200 nm, resulting in broad transitions in macroscopic devices. Nanowire devices with a width of 50 nm, however, show intrinsic first-order transitions with very sharp, single-step changes in the magnetoconductance [104]. Such improvement in performance in nanoscale devices is promising for future applications. In the following we discuss the main challenges in patterning complex oxides at sub-100 nm length scales.

2. Patterning strategies

Whereas atomic-scale control over layer thicknesses has been achieved already a long time ago in epitaxial processes, lateral patterning is much more challenging. A large variety of techniques have nowadays been used to pattern materials into sub-100 nm scale structures. Several examples are shown in Fig. 12. The standard patterning method for the creation of nanostructures is by exposing and developing either UV light- or electron-beam-sensitive resists. This method has been well developed and nowadays structures smaller than 10 nm are being realized [109,110]. Electron beam lithography has been applied to oxide nanopatterning in both top-down and bottom-up approaches. At the current state of the art a large number of materials systems can be patterned at sub-100 nm resolution. The main challenges in patterning oxides are in transferring the patterns to the oxides.

In top down processes, etching of the material is required. This can be done using wet chemical etching, ion milling, and reactive ion etching. Wet chemical etching has the advantage that it is material specific and the disadvantage that the etching is not directional. The latter point precludes the creation of small structures with high aspect ratios. Most of complex oxides are also very inert to standard wet etching. Ion milling and reactive ion etching are directional and thereby well suited for nanostructure patterning. A common problem with ion milling is that it damages the material and post-annealing steps to repair the damage can be necessary. This can be circumvented by using a protection layer. For example, a 50 nm thick gold layer allowed the fabrication of YBa2Cu3O7 nanowires with line-widths of 50 nm that showed very high critical current densities approaching the theoretical de-pairing limit [111]. A further problem is that the electron beam resists do not withstand a lot of ion irradiation, thereby limiting the thickness of the patterned material. This problem can be circumvented by using bi-layer hard amorphous carbon/resist masks. This way ion beam etching of thick (up to 200 nm) oxide films was achieved [112]. For certain oxides, low intensity ion irradiation can be directly used to pattern structures without etching the material [113,114].

In lift-off processes, the patterned resist is used as a shadow mask in a subsequent deposition step and is later removed. Most oxides have to be grown at temperatures and pressure where the resists are not stable. Hence, hard masks of sacrificial or inert materials are required that can be grown at low temperatures [115]. A promising strategy is to use a sacrificial bilayer that is composed of an easily soluble layer and a thermodynamically stable rigid layer [116]. Since two lift-off steps have to be performed, the maximum height of the structures is limited.

Nano-shadow masks (nano-stencils) allow the fabrication of large-area arrays of nanoscale structures. For oxides, the use of nano-stencils is complicated due to high deposition temperatures and high surface mobilities that limit the resolution. Pulsed laser deposition of SrTiO3 and BaTiO3 nano-arrays with dimensions down to 50-70 nm through SiN4 nanostencils has been demonstrated [117]. Moreover, ultrathin anodic alumina membranes were used to pattern extended arrays of ferroelectric Pb(Zr0.20Ti0.80)O3 for individually addressable epitaxial ferroelectric nanocapacitor memories with near Tb inch2 density [118].
For oxides that can be grown from liquid precursors, an alternative strategy exists. The precursors can be patterned with soft lithography techniques such as microcontact printing and micromoulding in capillaries. This has resulted in structures smaller than 100 nm in for example WO₃ and PbZr₁₋ₓTiₓO₃ [119].

Another promising direction in the nano-patterning of oxides is direct writing using a conducting-tip atomic-force microscope (CAFM). Nano-structures with lateral dimensions as small as only 2 nm have been patterned by CAFM in the 2-dimensional electron gas at the LaAlO₃/SrTiO₃ interface [120]. The writing mechanism is based on surface protonation at the presence of H₂O which dissociates into OH⁻ and H⁺ by a biased conducting AFM tip [108]. In another application, CAFM lithography was used to control the electrical resistance of SrTiO₃ surface and thin film multilayer structures through selective manipulation of oxygen vacancies in natural or artificially created crystal dislocations [121]. This paves the way for the realization of ultra-high density redox-based resistive random access memories [122]. However, the CAFM writing is inherently limited by low speed and by slow relaxation processes that may occur after writing.

To overcome the problem of induced damage during etching processes, various non-invasive bottom-up patterning techniques have been developed that rely on physical and chemical molecular self-assembly. Self-assembly of nano-composites is usually achieved by decomposition (pulsed-laser, sputtering, or CVD) of two or several intermixed materials on a substrate. A correct selection of thermodynamic conditions and growth kinetic parameters allows the formation of vertically aligned nano-structures (VAN) forming well-defined heterointerfaces [123]. The phases that are formed are not necessarily the same as those in the target material. For example, growth of BiFeO₃:CoFe₂O₄ (BFO:CFO) VAN films has been achieved by controlling the surface adatoms diffusion length [124]. Both vertically straight and gradient heterointerfaces can be realized by tuning the deposition frequency. In a different approach, 3-dimensional SrTiO₃ nanostructures have been fabricated with atomic precision from a metastable amorphous oxide film using a scanning transmission electron microscope [105,125].

The utmost importance in the nano-patterning of functional oxide materials is the ability to control and exploit point defects on an atomic level. In this respect, bottom-up nano-patterning using self-assembly accompanied by first-principle calculations of the stability of the nanostructures holds great promises for further miniaturization of oxide devices. For instance, controllable manipulation of oxygen vacancies in SrCuO₂ results in the formation of ordered, nanostruct-tured, rhombohedral SrCrO₂.₈ phases [126]. Charge accumulation in domain walls in ferroelectrics is another recent example that opens prospects for novel nano-electronic device fabrication [127].

Next to the patterning of the complex-oxide materials, devices also require patterning of additional layers used in the device fabrication, such as gate insulators, gate metals and contacting layers. Using a mixture of both bottom-up and top-down methods in electron beam lithography, significant progress has recently been made with the realization of device structures smaller than 100 nm [107,128]. We expect this progress to continue and that contacting and gating single complex-oxide nanoparticles will become possible. The key challenge is maintaining an acceptable contact resistance when the contact area is scaled down. To achieve this, chemical reactions between the contacting metals and the complex oxides will have to be controlled.

3. Concluding remarks

Complex oxide technology is in its infant phase, mostly driven by scientific curiosity rather than by circuit level integration and production cost minimization. From the device point of view, a significant progress is achieved in the fields of ferroelectric and resistive random access memories. The development will continue on a path towards transferring oxide technology to processes compatible with silicon semiconductor technology.
We expect main progress in the patterning resolution to continue in the next years and especially to be used for an ever larger variety of oxide materials. Furthermore, we envision more complex heterostructure integration including amongst others many active layers in the devices and integrated bottom, top and side gates. In terms of scientific merit, scaling down oxide devices allows not only miniaturization but also the exploration of new functionalities coming from quantum confinement and other finite size effects. We expect this to be a rewarding field of physics in the next decades with many exciting discoveries.

Epitaxial oxide films on semiconductor substrates
F. Sánchez, M. Spreitzer
aInstitut de Ciència de Materials de Barcelona (ICMAB-CSIC), Campus UAB, Bellaterra 08193, Barcelona, Spain
bAdvanced Materials Department, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
* Corresponding author.
E-mail: fsanchez@icmab.es (F. Sánchez).

Abstract
Crystalline oxides have a much wider range of functional properties than semiconductors, offering opportunities to improve existing microelectronic devices or to develop new ones. Microelectronics is based on semiconductors, generally using Si(0 0 1) wafers, but their incompatibility with most oxides makes the epitaxial growth of functional oxides on semiconducting wafers difficult. The few oxides that can be grown epitaxially are used as buffer layers and are overgrown with various functional oxides. However, the complex processing conditions required for crystalline growth makes oxide-semiconductor electronics difficult to implement. Ferroelectric oxides are already used in poly-crystalline film in commercial memories and can soon have a great impact with improved optical devices and new ferroelectric memories. Here we present the current state of the integration of epitaxial oxide films with silicon and other semiconductors, and we discuss some of the current and future challenges.

Keywords: Silicon; Oxide thin films; Ferroelectric oxides; Epitaxy; Oxides on Si

1. State of the art

The progress in epitaxial growth of oxides on semiconductors has started with monolithic integration of SrTiO$_3$ (STO) on silicon using molecular beam epitaxy (MBE) [129] and has triggered the development of beyond the complementary metal-oxide-semiconductor (CMOS) logic, as well as some emerging memory devices. Furthermore, due to a large number of exciting electrical, magnetic, and optical properties of various oxides and heterostructures, their integration with semiconductors has stimulated the development of devices with new functionalities, with, for instance, sensing characteristics, incorporated on a single chip together with logic capabilities. In light of this, growth of oxides on Si has been studied by many research groups in the past, but the progress has always been hindered by the problem of the materials’ incompatibility, making epitaxial integration difficult [130].

Most oxides react with silicon, and only a few of them can be integrated epitaxially without a thick interfacial layer. Epitaxial growth on silicon permits their use as a virtual substrate for the overgrowth with functional oxides and heterostructures. STO and yttria-stabilized zirconia (YSZ) are generally used as buffer layers. In the first case, epitaxy of STO on Si(0 0 1) is achieved with MBE using a complex process that was established two decades ago [129]. It starts with surface deoxidation, which results in a formation of 1×2 reconstructed surface. As-prepared surface is passivated with 1/2 monolayer of metallic Sr and overgrown with STO using the so-called “kinetically controlled sequential deposition” method, which results in an atomically sharp interface and thus enable electronic coupling and band gap engineering between the constituents. Since pulsed laser deposition (PLD) is generally used to grow complex functional oxides, their integration with STO-buffered Si(0 0 1) typically involves a combination of MBE and PLD in ex-situ processes. In the case of YSZ buffer layers, PLD was used from the first reports of epitaxial growth on Si(0 0 1) in 1990 [131]. The deposition process is simple, and epitaxy does not require removing the native amorphous SiO$_x$. Moreover, the films are stable at temperatures above 800 °C, although the thickness of a regrown SiO$_x$ amorphous interfacial layer is increased. The few other oxides, like MgO, that can be grown on Si(0 0 1) with cube-on-cube epitaxy are rarely used [132]. TiN, as well as SrO, have also been prepared as buffer layers using PLD [133,134]. On the other hand, (1 1 1)-oriented functional oxide films and interfaces are receiving increasing interest. For its epitaxial integration with silicon, (1 1 1)-oriented epitaxial buffer layers are needed to avoid the formation of crystal domains. Rare earth (RE) oxides like RE$_2$O$_3$ (RE = Sc, Y, Gd, etc.) grow epitaxially on Si(1 1 1), in some cases having an atomically flat interface [135]. The bixbite structure of RE$_2$O$_3$ does not match well with perovskite oxides, but it permits the epitaxial growth of spinel CoFe$_2$O$_4$, in spite of the large lattice mismatch [136]. An alternative to the bixbite buffer layer is the spinel γ-Al$_2$O$_3$, which has already been used to grow epitaxially spinel ferrimagnetic [137] and perovskite ferroelectric oxides [138].

**Fig. 13.** Slot-waveguide structures with the BaTiO$_3$/SrTiO$_3$ layer stack integrated epitaxially on silicon-on-insulator substrates. A low-temperature anneal sufficiently removed hydrogen and enables low propagation losses of 6 dB/cm. Figure adapted with permission from reference [145].
The epitaxy of a few functional oxides, like Fe$_2$O$_3$, LaNiO$_3$ or half-doped manganites, on bare Si wafers has been reported [139]. However, the films generally present degraded structural and functional properties, and thick interfacial layers. In contrast, the ferromagnetic semiconductor EuO can be directly integrated with Si(0 0 1). EuO, which is ferromagnetic below around 70 K and has a spin polarization of almost 100% and a conductivity that can be adjusted by doping, makes possible an efficient spin injection with silicon. Recently, the requirement of a SrO buffer for epitaxy has been overcome by using a sub-monolayer Eu template or other passivation processes [140].

A large variety of functional oxides, including ferroelectric oxides, have been integrated epitaxially with buffered Si wafers [63,133,141]. The most important progress has been achieved with piezoelectric and ferroelectric oxides. The largest transverse piezoelectric coefficient (27 C/m$^2$) was measured for epitaxial PbMg$_{1/3}$Nb$_{2/3}$O$_3$PbTiO$_3$ (PMN-PT) films on SrRuO$_3$/STO/Si(0 0 1) [142]. In addition to the piezoelectricity exhibited by ferroelectric oxides, the existence of two permanent polarization states in ferroelectric materials makes them useful for non-volatile memories. The simplest ferroelectric memory is the ferroelectric RAM (F-RAM), which is on the market for about 25 years. Its speed is high and its power consumption low, but the scalability is limited and the readout is destructive. Polycrystalline PZT is used in commercial F-RAM memories, despite the Pb's toxicity. The epitaxial growth of PZT in noncommercial devices is achieved on CeO$_2$/YSZ [143] or STO [144] buffered Si(0 0 1), generally with SrRuO$_3$ bottom electrodes.

BaTiO$_3$ (BTO) is a Pb-free ferroelectric with a spontaneous polarization of about 26 uC/cm$^2$ and very large electro-optical coefficients, which make it of interest for memory and optical devices, respectively. In the case of optical devices the polar c-axis has to be in the plane. BTO waveguides with large effective electro-optic coefficients and low losses were fabricated on MgO(0 0 1) substrates more than 20 years ago, and very recently, effective Pockels coefficients of more than 150 pm/V and low losses of around 6 dB/cm were achieved in epitaxial films deposited by MBE on STO/Si(0 0 1) (Fig. 13) [145]. On the other hand, in memory devices the ferroelectric polar axis has to be along the out-of-plane direction, but in tetragonal oxides on Si(0 0 1) it tends to be in the plane due to the thermal expansion mismatch between the materials. The epitaxial growth of c-oriented BTO on Si(0 0 1) by MBE is elusive, particularly in films thicker than a few tens of nm. MBE-deposited BTO films on Si(0 0 1) are generally characterized by piezoresponse force microscopy, with no reports of polarization loops. In contrast, a large polarization of above 10 uC/cm$^2$ has been reported in PLD-grown epitaxial films, hundreds of nanometers thick and thus likely to be free of epitaxial strain [146].

The recent discovery of ferroelectricity in doped HfO$_2$ is a breakthrough because of the material's full compatibility with CMOS fabrication processes. Doped HfO$_2$ is usually deposited at a low temperature, followed by a crystallization step at high temperature [147]. The films are polycrystalline, with a mixture of phases, including a metastable orthorhombic phase that is ferroelectric. The stability of the orthorhombic phase depends on the crystal size, and the films are generally only ferroelectric for a range of thicknesses. However, depending on the composition, ferroelectricity has also been confirmed in films either hundreds of nanometers thick or only a few nanometers thick, and thus potentially suitable for tunnel junctions. Although much less investigated, ferroelectricity in epitaxial, doped-HfO$_2$ films on YSZ single crystals has been reported [148].

Epitaxial complex oxides have also been integrated with other semiconducting substrates, such as GaN and InP, and particularly GaAs and Ge. In the case of Ge(0 0 1), the native oxide can be removed more easily than that of Si(0 0 1), and STO, BTO and other oxides can be grown epitaxially with atomically sharp interfaces (Fig. 14) [149]. GaAs is more unstable under oxidizing conditions, although several ferroelectric oxides have been deposited epitaxially on STO/GaAs(0 0 1). Furthermore, for epitaxial STO on an np-GaAs(0 0 1) photocathode efficient solar hydrogen production was demonstrated with an incident photon-to-current efficiency of more than 50%, which is based on the optimal electronic band alignment of an atomically sharp oxide-semiconductor interface [150]. A summary of epitaxial oxides on these semiconductors has been published recently [63].

The large-area growth of epitaxial oxides on silicon has mainly been demonstrated using MBE. However, due to the slow growth rate of MBE, development of alternative, industrially acceptable technologies with a high potential for epitaxial integration of oxides with large-area Si wafers is progressing intensively. Commercial PLD, sputtering, atomic layer deposition (ALD) and other chemical techniques are already available for large area deposition. As an example of the progress, the high-quality growth of PZT films on Si wafers with diameters of 200 mm was demonstrated and thus proved promising for the implementation of oxide electronics [151].

2. Current and future challenges

STO and YSZ are the commonly used buffer layers to grow functional
oxides on Si(0 0 1). However, they have limitations in comparison with oxide single-crystalline substrates. The STO/Si(0 0 1) interface can be unstable under the temperature and oxygen-pressure conditions needed for complex oxide epitaxy. At the interface a silicate layer often forms, causing depletion of the STO buffer layer. The use of thick STO buffer layers is limited by the low growth rate of MBE. The epitaxial integration of STO by PLD replicating the MBE process [152,153] could make the growth of functional oxides on STO/Si(0 0 1) in a single process easier. With respect to YSZ, the interplanar distance a = 5.14 Å/√2 = 3.63 Å is significantly smaller than the lattice parameter of most functional perovskite oxides. A CeO2 layer (a = 5.41 Å) is generally inserted to accommodate progressively the lattice mismatch, but the 3D growth of CeO2 challenges the deposition of flat functional oxides on CeO2/YSZ/Si(0 0 1). It would be important to achieve the 2D growth of CeO2 on YSZ or to replace the CeO2 with an alternative oxide layer. In spite of these limitations, a large number of functional oxides have been integrated epitaxially with Si(0 0 1). In contrast, the epitaxial growth on Si(1 1 1) is less investigated, regardless of the increasing interest in (1 1 1)-oriented oxide films and interfaces.

The functional properties of some complex oxide thin films on buffered Si(0 0 1) are close to those of films on single-crystalline oxide substrates, paving the way towards their wider use in commercial devices. The progress in the integration of ferroelectrics has been particularly relevant. For example, the large electro-optical coefficients and low losses of monolithically integrated BTO on Si(0 0 1) represent a breakthrough for BTO-based modulators and other optical devices [145]. BTO is also relevant for memory devices, generally requiring a c-orientation and high polarization. These properties, elusive for MBE-grown films, can be achieved by PLD either on YSZ [146] or STO [154] buffer layers, and c-orientation has also been accomplished with ALD [155]. The defects in BTO films grown using the high-energy plasma of PLD under a moderately low oxygen pressure (in the range of 0.01 mbar) might be related to the unit-cell expansion, mainly along the out-of-plane direction, thus favoring the c-orientation on Si(0 0 1) substrates. There is a lack of information about the microstructure of these films and its influence on ferroelectric switching kinetics, fatigue, or retention. Control over these defects can be used to engineer the lattice strain and ferroelectric properties of BTO and other oxides.

Doped HfO2 is the most promising candidate for ferroelectric memory devices. The ferroelectric phase is usually stabilized in a polycrystalline film rather than an epitaxial one. Epitaxial HfO2 is relevant because of the effect of the crystal orientation and the interfaces on the ferroelectricity of the thin films. Moreover, the homogeneity of an epitaxial layer can be critical in the case of ultrathin films. Thus, for a better understanding of the material, to improve properties like endurance, and for the fabrication of devices, more progress in epitaxial, ferroelectric, doped HfO2 is crucial.

Ferroelectric field-effect transistors (FET) were elusive due to the very low retention time caused by the insulating buffer layer between the Si and the ferroelectric film. But the interest in ferroelectric FETs has been renewed with the demonstration of a retention time of years and the improved endurance of perovskite and HfO2-based memories [156]. The used ferroelectric films are poly-crystalline or textured. Epitaxial ferroelectric films with out-of-plane polarization and a high coercive field on an ultrathin, high-k buffer layer are needed to enhance the properties. Moreover, the possibility of negative capacitance in ferroelectric FETs is of particular interest to reduce the sub-threshold slope of standard FETs and then allow lower voltage operation. Another emerging ferroelectric device is the ferroelectric tunnel junction (TJ), a nonvolatile memory with a non-destructive readout that can overcome some of the limitations of F-RAM memories. The complex device response, with resistive switching mechanisms added to tunnel transport, can have a strong influence on the switching times. Very recently, ferroelectric TJs with a high tunneling electroresistance ratio of above 109 at room temperature were fabricated on Si(0 0 1) [157]. The large tunneling electroresistance at room temperature is promising, although improved endurance and information about the switching times are needed. On the other hand, the fabrication of ferroelectric TJs using CMOS-compatible HfO2-based ferroelectrics will be a major step. For this purpose, epitaxial HfO2-based layers could be much more convenient than polycrystalline materials.

3. Concluding remarks

A number of oxides have been integrated with Si(0 0 1) epitaxially and can exhibit functional properties close to those of films on perovskite substrates. In many cases epitaxy is possible at a moderately low temperature and, in addition to PLD and MBE, ALD and other chemical techniques can be used in some cases. Ferroelectrics are very relevant for improved or new commercial devices, with doped HfO2 for memories and epitaxial BaTiO3 for optical devices being particularly promising. The spectacular progress in the epitaxial integration of functional oxides has been focused on (0 0 1) epitaxy, whereas the much less explored (1 1 1) oxides will probably be of increasing interest in the future.

Acknowledgements

Financial support from the Spanish Ministry of Economy and Competitiveness (SEV-2015-0496, MAT2014-56063-C2-1-R, and MAT2017-85232-R), from Generalitat de Catalunya (2014 SGR 734), and from Slovenian Research Agency (P2-0091 and J2-6759) is acknowledged.

Recent achievements and challenges in atomic layer deposition of complex oxides for heterostructures

Judith L. MacManus-Driscoll*, Mari Napari

Department of Materials Science and Metallurgy, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, UK

* Corresponding author.

E-mail: jld35@cam.ac.uk (J.L. MacManus-Driscoll).

Abstract

Compared to more traditional thin film deposition techniques, atomic layer deposition (ALD) has many advantages, including low deposition temperature and conformal growth of pinhole-free films. While progress has been made both in the development of the deposition processes as well as in the investigation of new functional materials, there are still some challenges for ALD to overcome. This paper gives a brief review of the current status of ALD processing of oxide electronic materials, with the emphasis on the epitaxial growth, together with discussion of the current challenges and future directions.

Keywords: Atomic layer deposition; Oxide electronics; Complex oxides; Epitaxy; Perovskites

1. State of the art

The essence of ALD is the surface-limited chemisorption of highly reactive precursors. Different gaseous precursors are exposed to a substrate surface in a sequential manner under low vacuum conditions (10−2–10−1 mbar). In contrast to conventional chemical vapour deposition, in ALD inert gas purging steps between precursor pulses are applied to prevent the chemical reactions in the gas phase. The precursors react with the surface species in a self-limiting way, which leads to formation of a film with thickness of a fraction of one atomic layer after each ALD cycle, the growth per cycle depending on the steric hindrance of the precursor molecules and the availability of the reactive surface sites [158]. Besides the traditional thermal ALD processing, which
directly exploits the reactivity of pre-cursors, approaches utilizing plasmas (plasma-enhanced ALD, PEALD) and energetic photons (UV-/ photoassisted ALD) have been developed [159,160]. Another technique that has gained interest recently, mainly due to its capability for fast large-area and roll-to-roll processing even under atmospheric pressure, is spatial ALD, where the precursor pulses and purging are separated in space rather than in time. This approach allows up to two orders of magnitude higher deposition rates (~1nm/s) compared to the traditional ALD, which makes spatial ALD attractive to industries where high throughput is needed [161].

ALD enables the growth of high quality films at low temperatures, typically below 400 °C, and film properties, such as crys-tallinity, can be controlled by the precursor chemistry and growth temperature, as well as the substrate material [162]. The self-limiting nature of ALD allows nearly atomically precise control of both the film thickness and composition, which makes it unique among the thin film deposition techniques. It also enables conformal coatings on complex surfaces, including high aspect ratio trenches and three-dimensional nanostructures [163], and deposition of multilayers with nanometer thicknesses [164]. In addition, thin films of compounds with alkali-metals, including Li [165], Na and K [166,167], and even Rb [168], can be grown by ALD. These chemistries are typically not stable at elevated temperatures and are thus inaccessible to most of the physical deposition methods.

2. Future prospects challenges

In the past two decades the developments in ALD have been strongly driven by the semiconductor industry, where it is used to deposit e. g. high-k gate dielectrics for transistors and diffusion barriers for metal interconnects, and in the production of memory devices [163]. As the technology node continues to decrease in size, ALD can be expected to strengthen its position in the future. In addition to semiconductor technologies, ALD has emerged in several areas including energy conversion and storage (photovoltaics [166], Li-ion batteries [167], etc.), catalysis [168], and sensors [169]. The low deposition temperature allows ALD to coat temperature-sensitive surfaces, such as polymers, fibers, and organic materials, which expands its potential towards biomedical applications and flexible electronics [170,171].

The vast majority of the current applications utilizing ALD rely on binary compounds, of which oxides are the most typical. In general, the ALD of a binary oxide comprises of two reactants i. e. the metal precursor and the oxidiser, making their process development fairly straightforward, and, consequently, binary oxides have already been deposited for more than the half of the elements in the periodic table [162]. However, many applications could exploit more complex oxide compounds, and hence the ALD growth of materials with ternary, quaternary, and even quinary chemistries have been established [165]. As an example of the complex oxides, the compounds that have gained a lot of interest in recent years are the functional perovskites, that exhibit a range of electric and magnetic properties, including ferroelectricity and -magnetism, and superconductivity [172].

The increase in compound complexity also makes the ALD process development more challenging. To obtain the desired stoichiometry several process parameters must be precisely controlled. These include the matching of the different precursors and ALD temperature windows of individual oxide processes, i. e. the temperature ranges where the growth is self-limiting, as well as the design of appropriate cycling sequences of each precursor in an ALD super-cycle [163]. Recent advancements in the ALD of multi-component oxides has benefited from the development of new volatile and reactive metal precursors, such as different amideates, alkoxide, and cyclopentadienyl, to accompany the traditional precursor chemistries like β-diketonates. However, for many complex materials, the offset of temperature compatibility between the precursors still remains as a challenge, and processing temperature windows are often narrow [173]. Another issue is the control of oxidation state. Oxide films of e.g. transition metals in a reduced state could be utilized in several applications, but in majority of the oxide ALD processes the strongly oxidising conditions needed for complete surface reactions also result in fully oxidised films. Because of the aforementioned reasons some technologically important materials, such as SrRuO3 and FeTiO3 still lack viable ALD processes [172].

Together with the complex oxide chemistry, in many applications epitaxial growth, and/or high quality heterostructures and multilayers are required. These properties have traditionally been connected to the films deposited by physical, high temperature techniques, such as pulsed laser deposition (PLD) and molecular beam epitaxy (MBE). Despite the original name of the technique, ALE, standing for atomic layer epitaxy, the epitaxial growth with ALD is typically far from straightforward. As a low-temperature deposition method compared to physical routes such as PLD, which typically take place at temperatures around 700 °C, in ALD the growth is more limited by the mobility of the species on the surface. The reduced kinetics hinder the epitaxial growth and can lower the crystalline quality of the films. Additionally, as a method based on surface chemistry, in ALD the substrate material plays a critical role in the formation of the crystal structure, and hence a single-crystal substrate or surface template is needed to realise the epitaxial growth, where the lattice-match between the substrate and film materials governs the film phase and orientation [162,173,174]. However, despite these challenges, the rapid progress made in ALD has questioned the assumption of chemical deposition routes being unsuitable for producing epitaxial films.

The epitaxial ALD growth has been demonstrated for a variety of binary oxides including TiO2 [175], NiO [176], CeO2 [177], and HfO2 [178] as well as more complex chemistries such as tin doped indium oxide (ITO) [179], Co₃FeO₄ [180], Mg₀.₇Ca₁.₃O [181], and La₀.₅Ca₀.₅MnO₃ [182] to name a few. Owing to the low deposition temperatures, the as-deposited oxide epitaxy typically benefits from energy enhancement either by highly reactive oxidizers, such as ozone or plasma-activated oxygen radicals [159], but, depending on the material, epitaxial growth can also be possible when weaker oxidants like water, are used.

When post-deposition annealing is added to the film processing the selection of ALD-grown epitaxial oxides and heterostructures expands immensely [173]. For example, many epitaxial perovskites can be fabricated by post-deposition annealing of ALD grown films. Fig. 15 shows an example of the ALD process of BiFeO₃ from (1-methoxy-2-methyl-2-propanoyl)methyl bismuth (Bi(mmp)₃) and bis(1-clophenoxydiethyl) iron (Fe(Cp)₂) precursors, and ozone, as well as the epitaxial structure of the film after subsequent annealing at 700 °C [183]. In addition to BiFeO₃, which is a promising oxide material for electronic applications because of its room-temperature ferroelectric and magnetic properties, ALD processes have also been demonstrated for almost 30 functional perovskites or perovskite-like structures [172] including epitaxial Bi₃Ti₂O₇ [184] and LiNbO₃ [185], as well as for high-k materials such as SrTiO₃ [186] and LaAlO₃ [187]. The ability of ALD to deposit conformal ultra-thin epitaxial layers also opens a possibility of tailoring structure interfaces with new functional properties [172].

Despite the tremendous progress in the development of ALD processes for high quality oxide materials, there are still some challenges to overcome. For example, while several materials, including perovskites, have been deposited on semiconductor Ge and GaAs substrates, there has been no success in epitaxial deposition of perovskite-structures directly on silicon as of today [173]. Further investigations into new processes and precursors are also needed in order to increase the range of complex oxide materials and number of applications enabled by them. This includes both the compatibility of precursors in the combined ALD process conditions as well as to the improvement of the as-deposited film quality in applications where high-temperature post-deposition annealing cannot be applied.
3. Concluding remarks

While the heterojunctions and epitaxial growth by ALD have not yet been exploited at the application level, there has been a great effort in the development of new approaches and techniques for atomic layer growth as well as in the achievement of new materials and processes. In terms of new oxide materials, the synthesis of functional complex oxides has advanced greatly, and in many cases ALD has been demonstrated to be competitive to traditional physical deposition routes. Despite the remaining challenges, the recent progress paves the way for a plethora of new emerging applications for ALD-grown electronic oxide materials.

Acknowledgements

The authors acknowledge funding from the EPSRC grant EP/P027032/1. Prof. O. Nilsen is acknowledged for valuable discussions on the current status of ALD perovskites.

Structure solving and refining, and strain gradients mapping in epitaxial thin films by X-ray diffraction techniques

Carlos Frontera *
Institut de Ciència de Materials de Barcelona (ICMAB/CSIC), Campus of UAB, 08193 Cerdanyola del Vallès, Spain
* Corresponding author.
E-mail: frontera@icmab.es.

Abstract

One of the main characteristics of transition metal oxides is the dependence of their properties on the precise composition (including oxygen content) and on the structural details (bond distances and angles). Diffraction techniques easily allow obtaining these two characteristics in bulk oxides, but the small amount of material present in thin epitaxial films makes them elusive. Two facts can help to compensate this handicap. The first is that epitaxial films behave in most cases like single crystals and concentrate the diffracted intensity in single spots that can be easily detected even for very thin films. The second is the use of synchrotron X-ray sources. Also of great relevance for different materials, especially for ferroelectric ones, is the strain state and its variation inside the film. In this short contribution, we briefly overview some of the X-ray diffraction-based techniques used so far to obtain the details of the structure of thin films, as well as the analysis of their strain gradients, and give our opinion on how can/must they evolve in the near future to spread their usage and enlarge their capabilities.

Keywords: Thin films; X-ray diffraction; Structure refinement; Strain gradient

X-ray diffraction has allowed understanding the strong dependence of the properties of many oxide materials on their structure and/or on their structural details (bond distances, angles, precise stoichiometry, etc.). Single crystal and powder diffraction have been widely used in bulk studies and have become an essential technique for the understanding and tuning of materials properties. Such a crucial contribution in bulk samples has not been fully transmitted to epitaxial thin films research where, in general, the precise determination of the atomic positions in the structure has been very scarce. In this short contribution we give a brief (and personal) overview of the methods used so far for this purpose, without entering into mathematical details, and how we believe that they must evolve to make them more popular in oxide based thin films. We also make a short overview of the methods used to study the strain distributions inside thin films. This point is of special relevance for ferroelectric oxide materials.

1. State of the art

The structure determination/refinement in epitaxial thin films is quite unusual. In general, structural studies are limited to determine the thickness, the strain state, and cell parameters of the film, in the best cases. Only a very tiny fraction of experimental works on epitaxial thin film are accompanied by a determination of the position of the atoms in the film. Historically, works detailing the structure are based on synchrotron radiation, and a very fruitful technique used is the coherent Bragg rod analysis (COBRA) method [188]. This method was initially developed for interfaces/surfaces (pure 2D structures) [189] but afterwards expanded and successfully used for the 3D structures of very thin films (4-5 nm, e.g. [190–192]).

COBRA method requires the use of synchrotron radiation. The reason for this is twofold. First, it is based on the measure of the so called crystal truncation rods [188] (emerging between Bragg peaks from the substrate, in the direction perpendicular to the film). These rods appear due to the lack of translational periodicity in the direction perpendicular to the film, are very weak, and require the use of a very powerful source to measure them. Second, the method requires a highly coherent beam, only provided by synchrotron sources. COBRA is
extremely powerful for studying very thin layers of high quality. It renders a detailed reconstruction of the electronic density, allows a very deep comprehension of the anomalies taking place at the interfaces, and provides information about strain and possible composition inhomogeneity. As a counterpart, COBRA method assumptions make it only valid for very thin films growing with the in-plane-periodicity imposed by the substrate.

Structure of thicker films has been successfully solved/refined using laboratory based equipment [193,194]. The method developed in these works is based on the use of extended 2D detectors allowing the integration of the whole intensity diffracted by the film at each Bragg position. After the appropriate corrections [193] one can obtain the structure factors and use conventional software for X-ray analysis like XLENS, EXPO, FullProf or SHELLX for structure resolution [193] and/or refined [194]. This method has been successfully used for determining bond distances and angles in different thin films from 10 nm to above [194]. Unfortunately, it has a very important limitation: reflections coming from the film must be separated from those coming from the substrate (out-of-plane cell parameters of film and substrate must be significantly different). Unpublished efforts to overcome this difficulty have been unfruitful.

Synchrotron radiation has been used for studying the structural details in the case of films (20 nm) with high overlapping between substrate and film diffraction peaks [195]. This last work is based on the fact that total diffraeted intensity is the result of the interference between the substrate and film diffracted beams. From this coherent addition, not only the modulus of the structure factor can be found but also its phase, provided that the substrate structure is known. In contrast with COBRA method, in Refs. [193–195] structure is described as usual by means of a crystallographic cell, a space group, and the positions of the atoms in the asymmetric unit.

As mentioned, strain analysis has been much more extended than structural analysis. Nonetheless, analyses are usually limited to the average strain state, as the position of diffraction peaks (symmetric and non-symmetric peaks) is enough for this purpose [196]. Besides this average strain, it is of interest to obtain information from possible lattice inhomogeneity inside the sample. In bulk samples, cell inhomogeneity is known as microstrain and, besides the microstrain analysis that can be done using different programs for Rietveld refinement, two methods are used to obtain them: the Williamson-Hall plots, and the Warren-Averbach method. In all cases, it is assumed that cell parameters inside the system follow a Gaussian distribution and only the variance of this distribution is obtained (average microstrain). Additionally, not only the modulus of the structure factor can be found but also its phase, provided that the substrate structure is known. In contrast with COBRA method, in Ref. [195] structure is described as usual by means of a crystallographic cell, a space group, and the positions of the atoms in the asymmetric unit.

More detailed information on the inhomogeneous distribution of strain in the film can be reached by assuming an analytic dependence of cell parameter on depth. This allows calculating the expected profile for diffraction lines and to refine the parameters describing the function dependence by comparing with experimental diffraction profiles obtained using laboratory sources [197]. More detailed information on the inhomogeneous distribution of strain in the film can be reached by assuming an analytic dependence of cell parameter on depth. This allows calculating the expected profile for diffraction lines and to refine the parameters describing the function dependence by comparing with experimental diffraction profiles obtained using laboratory sources [197]. The need of modeling has been circumvented by taking advantage of the high coherence of synchrotron radiation [198].

Another technique used to obtain the strain gradient or, more precisely, the variation of cell parameters on sample depth is X-ray diffraction using grazing incidence geometry [196,199]. Unfortunately, it is difficult to gain detailed information in the range of tenths of nanometers or below.

The study of strain, and even composition, distribution has reached an impressive level in nano-objects (nanoparticles [200], nanoslands [201,202] or even nanorods [203,204]). Some of these studies [200,202–204] are based on the reconstruction of three dimensional electron density of the nano-objects using a technique called coherent X-ray diffraction imaging. The reconstruction is done thanks to over-sampled 3D diffraction patterns measured using 2D detectors and highly coherent synchrotron radiation. The oversampling allows recovering the phase lost in the intensity measurement and, through an inverse Fourier transform, the real-space electron density. As in the case of COBRA, this method provides space-resolved information on strain and composition.

In fact, the use of 2D detectors in synchrotron sources has open the door to new perspectives in thin film structural characterization, as it allows to reconstruct the 3D intensity maps around diffraction peaks. Different works have recently evidenced the potential of enlarging reciprocal space maps from 2D to 3D in describing complex thin films. Domain formation in BiFeO3 [205] or in PMNPT [206] are clear examples of the capabilities of the technique. The 3D intensity distribution of the diffraction peaks retain very detailed information about the imperfections of film lattice [207,208].

2. Future prospects/current and future challenges

At present, the main handicap of the different existing methods to determine/refine structure of thin films is the lack of easy software tools for their use. This hinders any possible extended application of any of them. In general, their accessibility is restricted to the groups/researchers that have actively participated in their development. Thus, probably the main future challenge is the development of software, as user-friendly as possible, that could make these techniques "easy" to use for a wider amount of research groups.

Moreover, the access to synchrotron sources is much more difficult and delayed in time than access to laboratory equipment. Thus, a second challenge is the development of techniques based on laboratory sources. As explained in Section 1, in the case that diffraction peaks coming from substrate and film are separated enough, detailed structural studies can be done with Cu-radiation and the use of 2D detectors [193,194]. In general, this type of studies can be done in a reasonable amount of time: usually a weekend is enough to collect enough data for a confident refinement of the structure.

In the case of overlapping between substrate and film peaks, nowadays solvable using synchrotron radiation, two ways can be explored in the next future for the use of laboratory sources. The first one is to reproduce the methodology developed in Ref. [195] using high resolution Cu diffractometers. Line scans similar to those used in Ref. [195] are readily available using present laboratory based X-ray equipment. Thus, a similar analysis of this type of data could be performed on a sufficiently large set of collected line scans. The difficulties of this approach are twofold: (i) the time needed to collect such large amount of line scans and (ii) the effect of the non-coherence of laboratory X-rays. The second way is based on the use of 2D detectors. As mentioned, previous efforts to deconvolute signals from substrate and film have been unsuccessful, but these efforts have not included the known information from substrate structure nor treated the interference between both signals.

Concerning the characterization of lattice imperfections and their distribution across the film, a promising technique for a future development is the detailed analysis of 3D reciprocal space maps. The complete reconstruction of the electron density achieved for nanoparticles and nonorods [200,202–204] seems unrealistic as a thin film has two very large dimensions (and only one nanometer-sized dimension). Besides, and as an advantage, films have a very well know shape (this knowledge is solely limited by it roughness), and the only unknown contributions to lattice imperfections (i.e. those features breaking the translational symmetry of the cell) are strain distribution and composition inhomogeneity. The strain field produces a displacement of the unit cells that according to a kinematical description of diffraction in the Fraunhofer (far field) approximation renders, under certain assumptions (including composition homogeneity), that the intensity around a diffraction peak (at reciprocal lattice point h) is given by \( I_h(Q) = \frac{|\mathbf{F}(h)|^2}{v} |A_h(Q)|^2 \) where \( A_h(Q) = \int s(r) \exp(-ih\cdot r) dr \) (1).
In this last expression, \( s(r) \) represents the shape function of the illuminated crystal \( \left[ s(r) = 1 \text{ inside}, \ 0 \text{ outside} \right] \), and \( u(r) \) the strain field, i.e. the displacement of the cell located at \( r \) with respect to the ideal position expected for a perfect crystal lattice. Thus, at every diffraction peak, one could find, by inverting the diffraction intensity, the projection of the strain filed \( u(r) \) over the direction of the reciprocal lattice point \( h \) of the corresponding peak.

At present, 3D reciprocal space maps are mainly collected at synchrotron sources. In principle, they could also be collected using laboratory sources, but obtaining detailed information on strain distribution requires the use of high resolution diffractometers that are rarely equipped with 2D detectors. In any case, counting time for every diffraction peak would be large, but on the other hand, a limited number of diffraction peaks must provide enough information for strain field reconstruction.

3. Concluding remarks

The detailed structural characterization of thin films, understood as the precise description of the atomic positions in the system, is nowadays possible in most of the cases. Nonetheless, there are two main handicaps for this detailed description to be frequently made: the need, in many cases, of synchrotron radiation and the lack of easy-to-use software for data analysis. We suggest in previous section different ways that can help to circumvent the first difficulty and to obtain atomic positions from laboratory source data. It is worth mentioning that in the case of more complex heterostructures like multilayers (usually of higher technological interest) the situation is worse as the complexity increases and signal from different layers mix. Some of the techniques applied for single thin films could be extended to simple multilayers [195] but using synchrotron sources.

In comparison, the description of strain gradients (or strain field) has not been so deeply investigated. The advances done in other nanostructures (nano-objects), in which strain fields and nano-object shape have been fully extracted from X-ray data, must be extended to thin films under some assumptions/simplifications. In the future, this can render a way for detailed mapping the strain field in thin films.

Acknowledgements

Author thanks fruitful discussions with Dr. J. Santiso. Financial support from the Spanish Ministry of Economy and Competitiveness through the ‘Severo Ochoa’ Programme for Centres of Excellence in R&D (SEV-2015-0496), and projects MAT2015-71664-R and MAT2015-67593-P is acknowledged.

Characterization of point defects in functional oxide thin films

D.J. Keeble*  
Carnegie Laboratory of Physics, SUPA, School of Science and Engineering, University of Dundee, Dundee DD1 4HN, UK  
*Corresponding author.  
E-mail: d.j.keeble@dundee.ac.uk

Abstract

The current status of experimental point defect characterization of oxide thin films is briefly reviewed. The primary focus is on complex oxides, in particular perovskite oxide thin films. The evidence from different point defect sensitive characterization methods, such as electron paramagnetic resonance, positron annihilation spectroscopy, and optical spectroscopy is evaluated. Current and future challenges for point defect characterization of perovskite oxide films are detailed and suggestions for further work to aid the development of electronic grade complex oxide thin films are given.

Keywords: Point defects; Oxide materials; Defect characterization

1. State of the art

Oxide epitaxial thin films and heterostructures provide the prospect of enabling innovation in a wide range of technology challenge areas. The materials of interest span from established binary oxides, for example ZnO, to the diverse families of complex oxides, dominated by the ABO3 perovskite oxide materials. The sustained acceleration in functional oxide thin film research has been enabled by advances in thin film deposition methods, in particular pulsed laser deposition (PLD) and novel molecular beam epitaxy (MBE) approaches [70,210].

Achieving the desired optimized functional property, or properties, can require that the concentration of point defects, such as vacancies, interstitials, antisites, substitutional or interstitial impurity atoms, be sufficiently suppressed [70,210–212]. By contrast, the functional property can be due to a specific point defect, and some oxide devices provide some clear examples [213–217]. Experimental characterization aims to detect, and identify the atomic scale structure, of technologically relevant point defects in thin film oxides. For mature semiconductor technologies, where approximately sixty years of industrial effort has resulted in ultra-pure starting materials, the concentration of point defects can routinely be reduced below \( 10^{10} \text{ cm}^{-3} \), which for silicon corresponds to defect concentrations of less than 0.2 ppb (parts per billion) [218]. The situation for oxide materials is starkly different, achieving defect concentrations lower than hundreds of ppm can already be challenging. A defect concentration of 1000 ppm corresponds to 0.1%, advanced characterization methods such as wavelength dispersive X-ray spectroscopy may achieve this sensitivity, and secondary ion mass spectrometry exceeds it. Many conventional characterization methods, including diffraction techniques, typically have percent level sensitivities. Sensitivity is a necessary condition for candidate defect spectroscopies, but the nature of the information obtained is also of central importance. Some methods provide a single parameter, e.g. an energy level position, while others provide local atomic environment information enabling defect identification. The detection of an energy level, or an optical absorption or emission band, for example, can provide a valuable ‘fingerprint’ but further studies are required to unambiguously establish the precise nature of the defect.

Binary oxide devices provided direct examples for the role of point defects [216]. The defect physics of oxides such as ZnO [219], and MgO, is well established. They are amenable to study by the full range of defect spectroscopies, including electron paramagnetic resonance (EPR), local vibrational mode spectroscopy, positron annihilation spectroscopy (PAS), etc., and will not be discussed further. Here the focus will be complex oxides, specifically the perovskite oxides. Point defects are of central importance in ABO3 materials. Traditional electroceramic applications engineer these, guided by defect chemistry, typically through the incorporation of either donor or acceptor impurities [220]. From the close packed nature of the structure it is assumed that Schottky disorder is dominant [221]. Antisite defects may also exist under certain growth conditions. A Ti3+ center observed by EPR in SrTiO3 was assigned to Ti4+ defects [222,223], and more recent density functional theory (DFT) calculations have also considered the possible stability of cation antisite defects [224]. Nevertheless, the most probable point defects are vacancy defects and substitutional impurity ions. To obtain epitaxial perovskite oxide thin films with the highest quality electronic properties vacancy defect concentrations should be minimized, typically A-site or B-site cation vacancies act as acceptor defects, while oxygen vacancies are donors [225].

The development of aberration correction transmission electron microscopy (TEM) methods has enabled the presence of specific vacancy defects in ABO3 materials to be inferred directly from the atom column intensities [226–229]. Oxygen vacancies, \( V_O \), [226,228],
isolated Sr vacancies, $V_{\text{O}}$, [227], and $V_{\text{O}}$ clusters [229], have been observed in SrTiO$_3$. Local vacancy concentrations of several percent are, however, required. In the case of $V_{\text{O}}$, the relaxation of the near neighbor Ti atoms was also observed [227].

Cation ion vacancies are expected to be normally negatively charged, this makes them amenable to study using positron spectroscopy methods [230]. Positron annihilation lifetime spectroscopy (PALS) has observed cation vacancy defects in bulk PbTiO$_3$, Pb(Zr,Ti)O$_3$, and SrTiO$_3$. The B-site vacancy lifetimes have been found to be in the region of 180-190 ps, and the A-site lifetimes in the range 280–290 ps, in good agreement with the values calculated by DFT [231,232]. Variable energy (VE) positron beam measurements enable positron lifetime spectra to be measured as a function of depth, typically down to ~1 μm. VE-PALS measurements on PLD grown SrTiO$_3$ thin films where the stoichiometry was varied systematically clearly demonstrated positron trapped to $V_{\text{O}}$ with a lifetime of 181(3) ps and $V_{\text{O}}$ with a lifetime of 284(4) ps [230,233,234]. PAS methods can normally detect vacancy concentration as low as ~0.1 ppm. If the concentration of positron trapping vacancy defects is sufficiently high that the probability of trapping approaches unity, saturation trapping is observed and only positrons annihilating from vacancy traps are detected. The concentration at which this occurs depends on the defect specific positron trapping coefficient, but a value of the order of 100 ppm is typically inferred [230,233].

A powder EPR study of BaTiO$_3$ has reported a resonance at $g = 1.997$ attributed to the Ba-vacancy [235], but a detailed single crystal study and assignment was not performed. Extensive EPR studies of perovskite oxides have been made, but an unambiguous identification of a paramagnetic state of cation vacancy has yet to be reported. A broad optical absorption peak centered at approximately 1.3 eV has been observed in SrTiO$_3$, and has been tentatively assigned to $V_{\text{O}}$$^->$VT defect complexes on the basis of LDA + U DFT calculated in-gap energy level positions [236]. More recently DFT calculations have suggested that a luminescence peak observed around 2.0 eV in SrTiO$_3$ can be assigned to $V_{\text{O}}$, and that an emission at 1.2 eV is due to $V_{\text{O}}$ [225]. Measurements of optical conductivity over a wide energy range up to 35 eV have inferred that $V_{\text{O}}$ are responsible for large spectral weight changes from valence band states to unoccupied Ti-3d states in the 15-35 eV range, as well as formation of mid-gap states [237].

Despite the ubiquity of $V_{\text{O}}$ defects in perovskite oxide materials, and the extensive literature focused on their study, unambiguous spectroscopic identification of isolated $V_{\text{O}}$ defects remains challenging. EPR can provide detailed defect identification, however, there must be an accessible paramagnetic state and, ideally suitable $I \neq 0$ nuclei in the local environment of the unpaired spin density. Careful EPR studies in SrTiO$_3$ and BaTiO$_3$ have observed candidate centers involving spin density localized on Ti$^{3+}$, but uncertainty in the interpretation remains [223,238,239]. More recently a weak broad EPR signal has been reported in LaAlO$_3$, the signal intensity limited the characterization of the center [240,241], however, a detailed DFT study of $V_{\text{O}}$ defects in LaAlO$_3$ included calculation of EPR parameters which were reported to be consistent with the experimental values [242].

The room temperature optical emission in the blue from Ar$^+$$^+$ irradiated SrTiO$_3$ is well known, and is attributed to the $V_{\text{O}}$ [243,244]. A number of optical studies of SrTiO$_3$ with varying oxygen deficiency have been performed and have provided evidence for a number of bands inferred to be defect related [245–248]. Recent DFT calculations propose a detailed explanation involving the double donor nature of the $V_{\text{O}}$, where the first electron transitions to delocalized conduction band state while the second localizes as a polaron close to the vacancy [225].

The expected optical transitions for the vacancy defects in SrTiO$_3$ were also calculated. In addition to emission and absorption spectroscopy, optical conductivity measurements have also been reported and observe higher energy transitions that are sensitive to oxygen deficiency in SrTiO$_3$ [237,248].

In addition to the monovacancy defects a range of vacancy defect complexes are also possible. The presence of small vacancy cluster defects has been observed in VE-PALS studies of thin film SrTiO$_3$ grown by both PLD and MBE [211,234,249]. Thin films from three different sources exhibited a lifetime component in the approximate range 410–450 ps. DFT calculations provide evidence that this is a vacancy cluster that is larger than a Schottky defect, possibly containing on the order of ten vacancies [234]. Vacancy cluster defects have also been observed on the surface of SrTiO$_3$ by scanning tunneling microscopy and identified as Ti$_x$O$_3$ defects [250].

2. Current and future challenges

The experimental detection and detailed structural and electronic characterization of isolated oxygen monovacancy defects remains a primary challenge. Plausible theoretical models for the electronic behavior and local structure of $V_{\text{O}}$ have emerged from increasingly accurate DFT studies, but these require further testing against experiment to firmly establish their validity, in particular for model materials such as SrTiO$_3$ and LaAlO$_3$ [225,242].

Oxygen vacancy concentrations may span many orders of magnitude, reaching values above 1 % in some circumstances. At these high levels conventional defect spectroscopies often suffer from broadening or saturation problems, however, direct observation by aberration corrected electron microscopy methods and detection of the resulting in-gap states by photoemission spectroscopies, becomes feasible [226,228,251]. As advances continue in the development of electron microscopy methods it is to be hoped that future studies will provide insight not only on the presence and local concentration of oxygen vacancies, but also provide direct information on local relaxations [227].

High, but localized, concentrations of $V_{\text{O}}$ are a focus of current attention due to their possible role in the development of the two-dimensional electron system at the SrTiO$_3$/LaAlO$_3$ interface [251]. This work has led to the development of growing insight on the possible role of $V_{\text{O}}$ from experimental photoemission studies and theoretical approaches that include a combination of DFT with dynamical mean field theory. These studies identify the dual role of $V_{\text{O}}$ contributing electrons to the itinerant electron system, but also giving localized in-gap states with Ti 3d $\epsilon_g$ character [214,251]. While the approaches taken in this context include strong correlation effects, the resulting picture is similar to that which emerges from accurate hybrid functional DFT in the normal weak-coupling defect-state picture, namely the association of two very different electron states to the vacancy, a donor state electron and a small electron polaron with $\epsilon_g$ character [225].

Optical emission, absorption, and optical conductivity methods have exhibited sensitivity to oxygen deficiency, however, further studies that correlate all observable transitions for a systematic set of samples with varying oxygen content would be valuable. These studies may include both thin film and bulk samples subjected to the appropriate thermal and atmospheric treatments, guided by defect chemistry to provide a suitable range of oxygen deficiency [220]. Where appropriate, transport measurements yielding conductivity and carrier mobilities would also be desirable. Every effort should be made to employ more than one relevant experimental method to study the same sample set. For example, it may be possible to prepared samples demonstrating the presence of the expected optical transitions characteristic of $V_{\text{O}}$, but at sufficient concentrations to enable electron microscopy or photo-emission spectroscopy measurements to also be performed.

Further careful electron magnetic resonance studies are also required. For example the EPR center assigned to the $V_{\text{O}}$ in LaAlO$_3$ has been incompletely characterized, and reports of Ti$^{3+}$ centers related to oxygen deficient in titanate oxide perovskites require further investigation. If candidate centers can be established then modern hyperfine spectroscopy methods, combined with $^{17}$O enrichment, should yield local structural information. Correlation with optical and other...
methods would then be possible.

Positron lifetime measurements in perovskites can detect and identify both the A-site and the B-site cation vacancy defects, the characteristic positron lifetime values for these defects are clearly separated due to their rather different open volume sizes. However, typically the concentration of these defects are above the saturation trapping limit in thin film or ceramic oxides [230]. This still allows the nature of the dominant positron trapping vacancy defects to be identified and if there are several such defect types present an indication of relative concentration may also be possible. However, the absence of positrons annihilating from perfect lattice, the reduced bulk lifetime component, prevents the determination of absolute vacancy concentrations [230,234]. Measurements of both positron lifetimes and positron diffusion lengths should extend the concentration range for which absolute values can be estimated, and should be explored. Currently there are no direct measurements of vacancy defect specific trapping coefficient values for oxide materials and estimates of concentrations are being made using plausible values obtained from semiconductors and theory. The demonstrated utility of positron annihilation methods for studying vacancy related defects in thin film oxides provides a strong motivation for further fundamental measurements of trapping coefficient values [230]. Again, it would be highly desirable to perform studies that combine positron annihilation measurements with other methods that have provided evidence for sensitivity for cation vacancies, such as optical spectroscopy. For samples which exhibit sufficiently large concentration of cation vacancies, combined experiments with electron microscopy methods should also be considered. It would also be desirable design experiments to investigate the acceptor behavior of cation vacancies by combining defect spectroscopy with electrical characterization.

In near-stoichiometric perovskite oxides the results of defect spectroscopies support the expectation that vacancies are the primary point defects, but the situation is more complex for larger non-stoichiometry [224]. A combined positron annihilation and electron microscopy study demonstrated that for PLD grown SrTiO3 thin films moving into the region of assumed Ti deficiency rapidly resulted in a more complex situation; VT were observed, but with the onset of SrO layer inserting Ruddlesden-Popper (RP) phase defects VR were again detected [234]. In addition, vacancy cluster defects were also detected in this Ti deficient region. Further work is required to gain a deeper understanding of the defect content of nonstoichiometric perovskite oxides thin films.

Of central importance for the development of electronic application of complex oxides will be the suppression of technologically relevant point defects to concentration levels required to maximize doping efficiencies and carrier mobilities. Significant advances have been made in the case of homoepitaxial SrTiO3 grown by PLD and by MBE [70,210,212]. Positron lifetime measurements on La-doped hybrid MBE grown layers demonstrated that cation vacancy defect concentrations were below the saturation trapping threshold, annihilations from positrons delocalized in perfect lattice were detected [211]. Further advances in the growth of exceptional quality oxide thin films can be anticipated. The impurity ion content of the films did not prevent the attainment of excellent doping efficiencies and carrier mobilities, however, control of these levels may come relevant as film quality further improves. Further reduction in impurity ion content, however, may depend on an improvement in the quality of the available starting materials.

Finally, it should be noted that while progress is underway in advancing experimental characterization of point defects in perovskite titanate oxides it would be timely to extend these studies to model strongly correlated oxides, for example SrVO3. It should also be noted that advances in methods that enable operando experiments probing defects in operational devices is anticipated. Operando X-ray absorption spectroscopy and TEM have been performed [252,253], but there is a need for methods enable defect identification, electrically detected magnetic resonance can provide direct evidence on defects in operational devices [254].

3. Concluding remarks

Point defect characterization of thin film perovskite oxide materials has identified the presence of cation vacancies, both A-site and B-site vacancies. Substantive spectroscopic evidence for the oxygen vacancy defect exists, although unambiguous identification based on local structural information has yet to be established. Vacancy cluster defects have also been detected in perovskite oxide thin films. Significant advances in the growth of high quality complex oxide thin films have been made in recent years, and there is clear evidence defect densities can be suppressed to sufficiently low levels to enable efficient doping and excellent carrier mobilities.

Developments in electron microscopy of exotic states at oxide interfaces: Cryogenic imaging and advanced detectors

Ismail El Baggari a, Lena F. Kourkoutis b,c,*

aDepartment of Physics, Cornell University, Ithaca, NY 14853, USA
bSchool of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA
cKavli Institute of Science, Cornell University, Ithaca, NY 14853, USA
*Corresponding author.
E-mail: lena.f.kourkoutis@cornell.edu (L.F. Kourkoutis).

Abstract

Oxide interfaces produce exotic and functional properties such as ferroelectricity, superconductivity, and metal-insulator transitions. In particular, structural coupling at interfaces and charge transfer stabilize new electronic phases not found in the bulk. As an atomic-scale probe, aberration corrected scanning transmission electron microscopy (STEM) coupled with electron energy loss spectroscopy (EELS) provides crucial information about interfacial quality, cation intermixing, and composition. More importantly, STEM now directly visualizes and quantifies structural order parameters with picometer precision. It has mapped unusual ferroelectric states and structural distortions in heterostructures and superlattices, advancing experimental and theoretical understanding of emergent order at oxide interfaces. Further, atomically-resolved EELS has enabled mapping of subtle interfacial charge transfer, valence changes, and bonding.

In this review, we discuss two parallel developments in electron microscopy with promising applications for oxide heterostructures. First, the development of low temperature capabilities enables high-resolution mapping of previously inaccessible, low temperature electronic phases, including high-temperature superconductors, charge-ordered states and metal-insulator transitions. Hitherto, most efforts have been limited to room temperature measurements due to stage instabilities upon cooling, precluding understanding of exotic states that emerge at cryogenic temperatures. Second, the development of segmented and pixelated detectors enables new imaging modes that are sensitive to a variety of degrees of freedom including magnetic, electric, and strain fields arising at interfaces. By expanding both the range of accessible electronic phases and the sensitivity to local fields, atomic-resolution electron microscopy will help disentangle the complex couplings and emergent states in oxide heterostructures.

Keywords: Scanning transmission electron microscopy; Electron energy loss spectroscopy; Oxide interfaces; Structural distortions; Cryo-STEM; Pixel array detector

1. State of the art

Oxide heterostructures allow the realization of unconventional electronic states and tailored functional properties including ferro-
electricity, magnetism, and confined electron gases [255,256]. Key to the emergence of the rich properties and functionalities in these systems is the possibility of imposing structural ground states not found in the bulk. For instance, epitaxial strain, achieved by growing materials on substrates with different lattice constants, may change the lattice spacing, modulate the amplitude of structural distortions, and even stabilize new crystal symmetries [257]. Alternatively, interfacial connectivity may cause thin films to adopt structural distortions found in the underlying substrate [258]. The induced structural changes collude with the charge, orbital, and spin degrees of freedom, allowing the creation and manipulation of novel, exotic states of matter. Interfacial charge transfer further enriches the phase diagrams of oxide interfaces, by forming metallic, magnetic, and even superconducting states confined at the interface [255,256].

Parallel to remarkable improvements in atomic-precision materials growth, high resolution scanning transmission electron microscopy (STEM) has been instrumental for understanding the structure and bonding in heterostructures. In STEM, high-energy electrons are focused into a small probe to perform scattering experiments that generate a variety of signals. Electrons that have scattered to high angles may be collected by an annular detector, forming high-angle annular dark field (ADF) images as the probe is scanned across the sample. The contrast in high-angle ADF comparatively robust and scales to first order with the atomic number of the scattering potential; heavy atoms appear bright, light atoms appear dark, and oxygen is usually invisible. Annular bright field (ABF)-STEM is a more recently developed imaging technique that uses part of the bright field disk and allows the detection of oxygen columns. In addition to imaging, electron energy loss spectroscopy (EELS) determines composition and bonding with atomic resolution. Mapping composition and valence has been especially useful for detecting cation diffusion, oxygen vacancies, and electronic changes at interfaces.

Today, sub-angstrom resolution STEM imaging is routinely achieved thanks to the advent of commercially available aberration-correctors. In addition to resolving individual atomic columns in crystals, ADF-STEM can be used to determine their positions with picometer precision [259,260]. Consequently, it is now possible to directly map layer-by-layer lattice parameters and strain in oxide heterostructures.

A tantalizing capability in ADF is the direct measurement of functional structural order parameters. Epitaxial ferroelectric systems have especially benefited from numerous atomic-scale studies. In an oxide ferroelectric, the central atom (B-site) offsets relative to the center-of-mass of the A-site cage (Fig. 16(a), center), breaking inversion symmetry and producing polarization. By directly measuring the picometer-scale offset in ADF data, the polarization vector can be determined at each atomic site. These measurements have unraveled unconventional ferroelectric behavior in superlattices including the smooth rotation of polarization vectors, the formation of flux-closure structures, and polar vortices (Fig. 16(b)) [261,262].

Oxygen octahedral rotations are another powerful order parameter which mediates magnetic exchange and bandwidth. Using chemical pressure, epitaxial strain, or interfacial coupling, oxygen rotations may be controlled [257,258]. ABF-STEM is well suited to visualize and quantify both oxygen and cation distortions (Fig. 16 (c)), as long as imaging artefacts are carefully accounted for [263]. Recent ABF measurements have shown, for instance, that the growth of La$_1$$_x$Sr$_x$MnO$_3$ (LSMO) on the ferroelectric BaTiO$_3$ induces novel distortions that depend on the relative thicknesses of the two materials [264]. A thick ferroelectric layer suppresses octahedral rotations and induces a polar distortion in the LSMO layer (Fig. 16(c), (d)). As a result, an unusual polar ferromagnetic state emerges in the ultrathin LSMO layer, in agreement with theoretical predictions (Fig. 16(e)) and transport [264]. The quantification of interfacial distortions elucidates novel ground states and helps guide materials-by-design endeavors.

Electron energy loss spectroscopy (EELS) has been used to great effect to visualize interfacial electronic reconstructions at atomic

Fig. 16. Functional atomic displacements designed and characterized at the atomic scale. (a) Crystal structure of a cubic ABO$_3$ perovskite (top), polar displacement (center), and octahedral rotations and tilt (bottom). (b) ADF-STEM measurement of picometer-scale polar displacements (yellow arrows) reveals polar vortices in layered PbTiO$_3$/SrTiO$_3$ superlattices. Adapted from [262]. (c) ADF-STEM images of La$_1$$_x$Sr$_x$MnO$_3$ (LSMO) grown on BaTiO$_3$ (BTO), showing polar displacements (top) and octahedral tilts (bottom) of oxygen columns in the LSMO layer. (d) ADF measurement of metal-oxygen bond angles for thick (N = 32, top) and ultrathin (N = 4, bottom) LSMO. For N = 4, rotations are suppressed and polar distortions emerge. (e) Density functional theory calculations of octahedral tilt as a function of number of LSMO layers. For ultrathin LSMO (N=4, blue), octahedral rotations are suppressed. The N = 10 superlattice exhibits phase separation. Monte Carlo calculations and transport (not shown) show that the N = 4 superlattices exhibit an unusual polar ferromagnetic state. Adapted from [264]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).
resolution. By mapping the fine structure of EELS edges, small valence changes and charge transfer associated with emergent electronic behavior at interfaces can now be quantified [265,266]. For example, the evolution of the screening mechanism in LSMO grown on STO as it undergoes an insulator-metal transition has been mapped by tracking the Mn valence (Fig. 17(a)) [267]. The charge transferred and its spatial extent are quantitatively determined as a function of doping; close to the transition, the transferred charge is less than 0.05 electron and is confined to 2 unit cells, in agreement with theoretical models (Fig. 17). In a more recent example, charge accumulation has been observed and quantified at domain boundaries in BiFeO₃ [127]. These localized, charged interfaces and boundaries not only host unusual electronic states, but also have potential for device applications.

2. Current Challenges and Future Prospects

The overwhelming majority of atomic-resolution STEM measurements have been performed at room temperature, as stage stability is crucial for obtaining high resolution, high signal-to-noise data. Even in cases where emergent phenomena occur at low temperatures, STEM and EELS techniques have been limited to room temperature characterization. Achieving cryogenic capabilities is paramount to accessing and probing low temperature phases including high-temperature superconductivity, charge density waves, multiferroics, metal-insulator transitions, and magnetic transitions. Further, quantitative measurements of electric and magnetic fields at interfaces are essential for understanding the complex couplings driving emergent states. However, conventional imaging techniques are insensitive to momentum information, which precludes the detection of atomic scale and nanoscale fields in the specimen. There are many other limitations being addressed by developments in theory and instrumentation. In this section, we focus on new developments in low temperature STEM-EELS and in universal detectors which extend the range of detectable information.

Atomic-scale imaging at low temperature is necessary to address a variety of open questions, from quantifying structural distortions that stabilize unusual ferromagnetic phases [268], to characterizing interfacial behavior in unconventional superconducting heterostructures [269], to visualizing metal-insulator transitions [270]. At present, cryogenic STEM is a fledgling technique, as holder designs have lagged behind advances in resolution and precision. That said, cryogenic STEM has been demonstrated on cobalt oxides, albeit with limited resolution and stability [271]. Recently, sub-angstrom resolution (0.78 Å) imaging has been demonstrated in aberration-corrected STEM near 95 K using a liquid-nitrogen side-entry holder (Fig. 18(a)) [272]. More importantly, picometer-scale lattice displacements associated with charge ordering in a manganite are measured and mapped under cryogenic conditions (Fig. 18(a)). The role of the lattice in a variety of low temperature electronic phases can now be quantified with high resolution and precision.

To go beyond high resolution cryogenic imaging to spectroscopic mapping is significantly more challenging. A spectrum with sufficient signal-to-noise ratio requires a dwell time of a few of milliseconds, orders of magnitude larger than typical imaging dwell times. Thus, low temperature EELS experiments have focused on collecting area-averaged spectra or line profiles. For example, the Co valence near 85 K has been determined in LaCoO₃ as it undergoes a spin reconstruction [271]. A more recent example involves FeSe/SrTiO₃ which displays an unusually high superconducting transition...
temperature driven by interfacial effects. Using line profiling across the interface, thickness-dependent charge transfer at 10 K has been observed [273]. However, the tendency of emergent states towards nanoscale inhomogeneity and spatial modulation requires atomically-resolved spectroscopic maps at low temperatures. In addition to ongoing improvements in holder stability, more efficient detectors hold great promise for atomic-resolution EELS mapping under challenging conditions. Unlike indirect, scintillator-based detectors, direct electron detectors collect electron-hole pairs and, consequently, improve the detector quantum efficiency, the point spread function, and the signal-to-noise ratio [274]. The improvements are especially critical for EELS applications that require low-dose or rapid-acquisition such as when operating in cryogenic conditions or when performing monochromated measurements. Even at a relatively short dwell time (2.5 ms/pixel), the signal-to-noise ratio of the spectra recorded using a direct electron detector camera remains high, allowing, for instance, atomic-resolution elemental maps of a LSAM/STO interface near liquid nitrogen temperature (Fig. 18(b)) [275]. Future studies will undoubtedly map fine structure changes associated with exotic orderings of spin, charge, and lattice. A radically new imaging approach involves the deployment of segmented and pixelated detectors (Fig. 19) which extend the range of measurable quantities in the electron microscope. These detectors benefit from greatly improved dynamic range and read-out speed, and have allowed new imaging modes including measurements of fields in materials (Fig. 19(b), (d)). The detection of momentum-dependent scattering is particularly useful, as deflections or intensity redistribution in the diffraction pattern may be associated with the presence of strain, magnetic, or electric fields in the specimen. Conventional detectors such as ADF and ABF integrate out such momentum information. In contrast, segmented detectors, such as quadrant detectors, can distinguish small horizontal and vertical shifts in intensity due to the presence of local potentials, a technique known as differential phase contrast (DPC) [276–278]. For instance, magnetic skyrmions in FeGe1-xSx have been reconstructed with high resolution (Fig. 19(b)) [279]. In another application, both ferroelectric domains and atomic electric fields in BaTiO3 have been measured using DPC [276]. Pixelated detectors (Fig. 19(c)) offer even more flexibility since they can collect the full diffraction pattern at each scan position, allowing 4D STEM (2 scan coordinates and 2 momentum coordinates) [280–282]. The collected diffraction patterns can be analyzed and integrated in many ways, forming a variety of conventional (ADF, ABF) and unconventional (center of mass, phase) images post hoc [280,282]. Center of mass imaging (Fig. 19(d)), for instance, can detect deflections of the electron beam that may arise from polar and ferroelastic domains [282]. A present challenge involves disentangling the various fields affecting the shifts and intensities of recorded diffraction patterns. Indeed, crystal tilt, polarization, and magnetic fields combine in a complex way, especially at epitaxial interfaces where elastic and electromagnetic fields couple strongly. That said, careful and clever decompositions of the various components can be achieved by considering symmetry, momentum dependence, simulations, and complementary measurements of diffraction patterns. These novel imaging capabilities will elucidate emergent textures arising at oxide interfaces, including ferroelectric vortices, magnetic skyrmions, and modulated phases.

3. Concluding remarks

Electron microscopy and spectroscopy have advanced significantly beyond the ability to resolve atoms. Complex lattice behavior can be mapped with picometer sensitivity, and subtle charge transfer processes at interfaces may be quantified. In the future, it will be possible to perform microscopy across a range of temperatures, from liquid helium temperature to far above room temperature, to map the entirety of phase diagrams, probe phase transitions, and track the evolution of exotic states with atomic resolution. Developments in detector technology and theoretical understanding of new imaging modes open the possibility of probing not only the atomic structure, but also the magnetic, electric, elastic, and polar fields arising at interfaces. These recently demonstrated advances hold great promise for understanding the couplings underpinning the complexity and functionality of oxide heterostructures.

Acknowledgements

This work was supported by PARADIM, a National Science Foundation Materials Innovation Platform (Grant DMR-1539918), and by the Air Force Office of Scientific Research (Award number FA 9550-16-1-0305).

Resistive switching oxides for data storage
Regina Dittmann *
Peter Grünberg Institute 7, Forschungszentrum Jülich GmbH & JARA-FIT, 52425 Jülich, Germany
*Corresponding author.
E-mail: r.dittmann@fz-juelich.de

Abstract
A large variety of oxides can exhibit a reversible, redox-process
based non-volatile change in electrical resistance upon electrical stimulus, a phenomenon known as resistive switching. Resistive switching cells are therefore considered as non-volatile resistive random access memory (ReRAM). The most commonly used materials are amorphous or polycrystalline binary oxides such as TiO2, HfO2 and Ta2O5. However many complex oxides such as titanates, man-ganites and ferrates have also been reported to show resistive switching. The development of ReRAM has progressed rapidly since the mid-2000s in both industry and academia. Very promising properties such as scalability down to the nm regime, fast writing (100 ps), low energy consumption (sub pJ/bit), and high performance with respect to endurance and retention has been demonstrated so far which makes ReRAM a promising candidate for future emerging non-volatile memories. The first mass-production of low-inegration density ReRAM cells started in 2013 and ReRAM cells are currently under discussion to be employed in embedded memory applications and as future storage class memory.

Keywords: Memory devices; Memristors; Resistive switching; RRAM; ReRAM

1. Working principle and requirements for data storage applications

A large variety of oxides can exhibit a reversible, non-volatile change in electrical resistance upon electrical stimulus, a phenomenon known as resistive switching or memristive devices [283]. Resistive switching devices are often composed of a metal oxide layer sandwiched between two electrodes as sketched in Fig. 20(a). In the simplest case, after an initial forming process, a reversible change between a low resistance state (LRS) and a high resistance states (HRS) takes place and can be interpreted as a switch between the logical “1” and “0”, respectively. The state of the resistive switching cell is detected by applying a read voltage which is insufficient to modify the resistance state. Resistive switching cells can therefore be employed as non-volatile resistive random access memory (ReRAM or RRAM). It is important to note that in many ReRAM cells, the device resistance can be switched to multiple values in between HRS and LRS, displaying an analogue behavior. This analogue switching behavior can be utilized for multilevel storage, i.e. to store more bits per cell. Multilevel storage serves as the basis for analogue computing as required in neuromorphic computing.

Among metal oxides, two different types of memristive systems show the so called bipolar switching, illustrated in Fig. 20, namely electrochemical metallization (ECM) cells also called conductive bridge memories (CBRAM) (Fig. 20(b)) and valence change mechanism (VCM) cells (Fig. 20(c)) [284]. ECM cells operate by an electrochemical dissolution of an active electrode metal such as Ag or Cu, a drift of cations through a metal ion conductor and a formation of a metal nanofilament. Since the metal ion conductor not necessarily has to be an oxide, we will restrict ourselves in the following on VCM cells. The VCM is typically found in metal oxides that show sufficient ion mobility. The migration of these ions changes the local stoichiometry and, hence, lead to a redox-reaction accompanied with a valence change of the cation sublattice and a change in the electronic conductivity. This valence change usually takes place within small filaments. The most commonly used materials are amorphous or polycrystalline binary oxides such as TiO2, HfO2 and Ta2O5, however many complex oxides such as titanates, manganites and ferrates have been reported to show resistive switching. Complex perovskite oxides, are less attractive for memory applications due to the potential incompatibility of the materials with complementary metal-oxide-semiconductor (CMOS) technology and, in particular in their crystalline form, due to the high processing temperatures required.

In order to compete with Flash memory, ReRAM should display characteristics such as $R_{HRS}/R_{LRS} > 10$, Write voltage $< 5$ V, Read voltage 0.1-0.5 V, scalability to $< 22$ nm and/or 3D stacking, switching speeds below 100 ns regime, low energy consumption, high endurance (write cyclability) of at least $10^7$ and retention times of 10 years. More details about requirements for other types of memories can be extracted from references [285,286]. It is important to note that the combination of requirements for ns switching and a retention time of 10 years sets a voltage-time dilemma, namely that a ratio of the write voltage to the read voltage of only ten needs to lead to an acceleration of the switching kinetics of retention time to writing time of about $10^{16}$.

2. State of the art

The development of VCM type devices has progressed rapidly since the mid-2000s in both industry and academia. one of the reasons is that many amorphous or polycrystalline binary oxides with excellent switching behavior such as Ta2O5 and HfO2 have already been implemented in the running CMOS lines as gate-oxides.

With respect to scalability, planar cell sizes down to 10 nm × 10 nm have already been demonstrated for HfOx/Hf ReRAM [287]. By employing a sidewall electrode geometry, operating HfO2-based ReRAM cells with an area of 1 nm × 3 nm have been successfully fabricated recently [288].

According to the thermally assisted ionic motion in combination with the small distances which have to be overcome to move ions in the spatially confined interface region (see Fig. 20(c)), the write speed is typically in the order of tens of ns. In dedicated studies, ReRAM devices have been observed to switch as fast as 100 ps [289] and potentially even faster. As a result of the temperature assisted field accelerated motion of ions, a sufficient non-linearity of the switching kinetics in order to solve the voltage-time dilemma is provided for the filamentary switching systems [290].

With respect to low power consumption, it has been demonstrated for Ta2O5 based nanodevices that they can switch at sub 2 ns times under sub 2 V with less than 10 μA, resulting in a sub pJ/bit operation

![Fig. 20. (a) Sketch of a metal-oxide-metal cell, exemplary current-voltage curve of a bipolar switching cell and illustration of the memory function; (b) sketch of the cross section a ECM cell; (c) sketch of the cross section of a VCM cell in the vicinity of at the active electrode.](image-url)
energy [291]. Although reliability is currently one of the most severe obstacles for the commercialization of ReRAM, a variety of groups have demonstrated cells fulfilling the requirements mentioned above. By performing temperature accelerated lifetime tests, a retention time of 10 years has been demonstrated for a large variety of materials, e.g. in HfO₂ based cells [292]. Typically an endurance of 10⁶–10⁷ cycles is reported, however, a best performance of 10¹²² cycles has been demonstrated for Taox based VCM cells [293]. However, one has to note that an endurance and retention trade-off was identified for ReRAM [294]. Therefore, best performance in both retention and endurance has not been realized so far.

Another crucial issue for the commercialization of ReRAM is the cell-to-cell and cycle-to-cycle variability. According to the stochastic process of filament formation during forming and switching, variability is an inherent problem for filamentary ReRAM cells. Moreover, the variability deteriorates with decreasing cell and filament size, resulting in a scalability and variability trade-off. In particular the occurrence of tail-bits can be regarded as road-block for the scaling of ReRAM devices. For more detailed information, we recommend the comprehensive performance table of ReRAM devices including numerous material combinations published over the last years which can be found in reference [295].

Fig. 21 presents the industrial development of the storage capacity of ReRAM in comparison with the competing non-volatile data storage technologies, namely CBRAM, phase change memory (PCM), spin-torque magnetic random access memory (STT-MRAM) and NAND-Flash. Although impressive advances have been obtained so far, the storage capacity of ReRAM is still below most of the competing memory technologies. However, driven by the superior power consumption of ReRAM in comparison with Flash memory, Panasonic released in 2013 the world’s first mass-production of an 8 bit microcomputer with 180 nm node, 64 Kbyte embedded ReRAM for portable healthcare products (see red point 4 in Fig. 21) [296].

3. Future prospects and challenges of ReRAM

Based on the future societal challenges and the related requirements in the field of information technology (IT) such as ubiquitous computing, storage and processing of “big data” in conjunction with the internet of things (IoT), dramatic changes in the way systems store and access data have to take place. The classical cache/memory/storage hierarchy is rapidly becoming the bottleneck for large systems. Fig. 22 illustrates the role of the different types of current memory devices arising from their costs and level of performance and storage density. There is a huge performance gap between dynamic random access memory (DRAM) and primary storage devices like hard disc drives (HDD), and even solid state drives (SSD). The “holy grail” for the memory industry would be to come up with a new memory, the so-called ‘storage class memory’ (SCM) filling the gap between DRAM and primary storage, from both a performance and a price point perspective.

ReRAM has primarily been considered as Flash replacement in the past. However, as a result of the current breakthrough in 3D NAND Flash technology with densities up to terabit on a single chip (see Fig. 22), companies are considering ReRAM rather for future SCM than for Flash replacement.

With respect to the future role of ReRAM technology, it is important to note that all stand-alone memory approaches discussed above will continue to undergo the most aggressive pitch scaling and cost reduction pressure among all other possibly data storage applications. The chances that a 2D ReRAM approach will be able to compete on the market is becoming increasingly unlikely and highly sophisticated 3D approaches such as the 3D X-point technology of Intel and Micron based on PCM [298] will have to be developed for ReRAM in order to compete.

Therefore, the majority of IT related companies are considering ReRAM rather for embedded memory applications than for stand alone memories. One successful example mentioned above is the 180 nm node memory chip of Panasonic employed for portable healthcare products [296]. In the meantime Panasonic has succeeded with the fabrication of prototype ReRAM in the 40 nm node and is striving to enter the market for IoT application in the near future [299].

Within this context it is important to note that the leading foundry TSMC, according to their latest press-release [300], announced to offer embedded ReRAM in combination with 22 nm FinFET technology in 2019. If this holds true, it would be a considerable break through for embedded memory products based on ReRAM technology. Furthermore, smart systems with power-supply-independent non-volatile memories combined with sensors and cognitive systems could be highly promising novel IoT applications for ReRAM technology in the future. Moreover, ReRAM could be employed for neuromorphic computing due to its analogue-type switching ability. Intensive research is currently conducted towards the use of ReRAM as synapses or neurons in future neuromorphic circuits. One example is the self-
learning music composing chip which has been fabricated by IMEC in HfO2 ReRAM technology [301]. Since this topic is beyond the scope of the article, the reader is directed towards the article 'neuromorphic computing' in this special issue.

4. Concluding remarks

ReRAM technology has strongly advanced over the past 5 years. Sub-ns switching, high integration densities and first approaches for 3D stacking have been demonstrated for prototypical ReRAM systems and no fundamental limits to use ReRAM for highly integrated data storage applications have been identified so far. However, there exist several trade-offs such as between retention and endurance and between scaling and variability which could be eliminated in the future by exploring new materials and/or new material combinations. Though one should keep in mind that industrial groups have focused so far on materials which have already introduced in their CMOs-lines, such as HfO2 and Ta2O5. Therefore, a considerable threshold of performance improvement has to be overcome in order to justify the introduction of novel, more complex materials. Besides this, a large amount of processing and integration issues, in particular with respect to 3D integration, will have to be addressed in the future for ReRAM to become competitive with both conventional and alternative emerging data storage technologies.

Acknowledgements

Funding from the DFG (German Science Foundation) within the collaborative research center SFB 917 ‘Nanoswitches’ and the W2/ W3 program of the Helmholtz association is gratefully acknowledged. We furthermore acknowledge funding by the ToBE CoST action (MP1308).

Oxides for data storage and processing: Ferroelectric tunnel junctions

Vincent Garcia *, Manuel Bubes

Unité Mixte de Physique, CNRS, Thales, Univ. Paris-Sud, Université Paris-Saclay, 91767 Palaiseau, France

*Corresponding author.

E-mail: vincent.garcia@cnrs-thales.fr (V. Garcia).

Abstract

Ferroelectric tunnel junctions are two-terminal devices in which a ferroelectric layer of a few unit-cells is sandwiched between two electrodes. In these devices, polarization reversal leads to large modifications of the tunnel resistance, allowing a non-destructive readout of the information stored by the polarization direction. These devices possess attractive properties for applications as nonvolatile memories. In addition, exploring the fact that polarization usually reverses by the nucleation and propagation of domains, these junctions can also exhibit a memristive behavior associated to non-uniform configurations of ferroelectric domains. Such ferroelectric memristors can be used as artificial synapses in neuromorphic architectures. Here we review recent progress and future challenges with ferroelectric tunnel junctions.

Keywords: Electron tunneling; Nanoscale ferroelectrics; Electroresistance; Memristors

1. Context

Compared to other existing non-volatile technologies on the market like Flash memories, ferroelectric random access memories (FeR-AMs) possess key advantages such as fast writing speed, low power operation, and high read/write endurance [302]. In FeRAMs, the memory element is a 100-nm-thick ferroelectric layer, which polarization orientation is switched by applying an external electric field. However, the capacitive readout of the polarization orientation limits the scalability of FeRAMs up to gigabit densities [303]. In addition, the information needs to be rewritten after capacitive readout. Alternatively, ferroelectric diodes, in which the current flowing through the ferroelectric depends on the polarization orientation at the ferroelectric/electrode interface, have been considered to promote a non-destructive resistive readout of the information [304]. As the ferroelectric layer thickness is large enough, the limited readout current prevents the miniaturization of ferroelectric diodes. In ferroelectric tunnel junctions (FTJs), a ferroelectric tunnel barrier is sandwiched between two electrodes [305,306]. This layer is a few-unit-cell-thick so that quantum-mechanical electron tunneling becomes possible [307]. The ferroelectric polarization reversal produces large changes of the tunnel transmission via electrostatic effects at the ferroelectric/electrode interfaces (Fig. 23). The induced tunnel electroresistance (TER) or OFF/ON ratio can be as large as $10^4$ at room temperature [308,309]. This non-destructive readout of information, together with the much larger readout current densities than in ferroelectric diodes, could lead to a new generation of high-density ferroelectric memories.

2. State of the art

While the concept of FTJs was proposed in the early seventies by Eski [310], it did not receive much attention for decades as the critical thickness of ferroelectrics was believed to be well beyond the nanometer range [311]. From the beginning of the century, the development of modern techniques for the growth of epitaxial oxide thin films lead to the demonstration of ferroelectricity in ultrathin thin films [312,313]. In 2003, resistive switching was first demonstrated across ultrathin films of PbZr0.52Ti0.48O3 sandwiched between SrRuO3 and Pt, with resistance contrasts of 400% at room temperature [314]. In 2008, the same authors managed to monitor independently polarization reversal and resistance variations across the ferroelectric layer and showed that the two switching events were not correlated [315]. In 2009, three groups independently demonstrated that polarization reversal could induce large variations of the tunnel resistance on bare surfaces of ferroelectrics (as illustrated in Fig. 24) [316–318], using scanning probe microscopy techniques derived from atomic force microscopy (AFM). This joint discovery triggered active research on solid-state FTJs showing large TER associated to the reversal of ferroelectric polarization [308,309,319–321]. In addition, the ferroelectric tunnel barrier can be used to trigger interfacial phase transitions if strongly
correlated oxides are used as electrodes. Alternatively, ferromagnetic electrodes combined to a ferroelectric tunnel barrier, enable a non-volatile control of the interfacial magnetic properties. As electron tunneling is highly sensitive to such interfacial changes, this opens avenues towards the non-volatile control of electronic and spin-tronic responses in multifunctional devices [321].

Thus, FTJs have potential applications as binary non-volatile memories, taking advantage of the non-destructive readout of polarization and simpler device architecture than conventional FeRAMs. Another degree of freedom is the ferroelectric domain structure of the tunnel barrier that can be exploited for analog devices such as memristors [322]. Indeed, considering that in ferroelectrics, polarization reversal usually occurs by the nucleation and propagation of domains [323] and that the domain size scales with the square root of the film thickness [324], non-uniform configurations of ferroelectric domains should in principle be achievable in FTJs. Using BaTiO3 and BiFeO3 FTJs, the interplay between the analog response of the junctions and the ferroelectric domain configurations was investigated: the progressive resistance evolution between ON and OFF is accompanied by the reversal of the ferroelectric polarization from up to down through the progressive nucleation of domains in several zones with limited propagation (Fig. 25) [308,325]. This was recently exploited to achieve

![Fig. 24. Direct evidence for ferroelectricity-related TER with ultrathin BaTiO3 films. Parallel (a-c) PFM phase and (d-f) C-AFM resistance maps of four ferroelectric stripes for BaTiO3 films, 1, 2, and 3 nm thick. The corresponding resistance profiles are displayed underneath [316].](image)

![Fig. 25. Switching of resistance and polarization in a 180-nm-wide BiFeO3-based FTJ. PFM images show a progressive switching from downward to upward (1–4) as resistance switches from OFF to ON under negative voltage pulses of 100 ns. Reversibly the ON to OFF switching is accompanied by polarization switching from upward to downward (4–7) with positive voltage pulses. Adapted from [308,327].](image)
3. Future challenges

Although experimental evidence suggests that TER is undoubtedly associated with ferroelectric polarization switching [316–318,320], there are still some fundamental issues to be addressed. Indeed, large TER in oxide-based tunnel junctions cannot be quantitatively interpreted by simple electrostatic models involving partial screening of polarization charges at the interfaces [306,307]. Instead, it necessitates complex descriptions containing interfacial dielectric layers [328] or doped-semiconducting layers [329]. For example, experiments with BaTiO3 or BiFeO3 tunnel barriers generally show an OFF (ON) state for polarization pointing toward the oxide electrode (Co or Pt electrode) [308,320,329,330]. This is counterintuitive as it indicates that the effective charge screening is better at the oxide interface than at the simple metal interface. It could suggest the presence of an interfacial dielectric layer at the metal/ferroelectric interface [330].

Moreover, most of the experimentally-reported FTJs are using ferroelectric barriers made of oxide perovskite thin films [321]. To maintain ferroelectricity in nanometer-thick oxide films, sophisticated experimental approaches such as strain engineering and careful control of epitaxial growth are generally required, which inevitably results in a complex fabrication process. En route toward more processable FTJs, organic ferroelectric materials have also been considered as tunnel barriers [331]. Organic FTJs may also exhibit different electronic transport properties from their inorganic counterparts because of the weak Van der Waals interfacial bonding with metal electrodes. Tian et al. fabricated submicron organic FTJs using PVDF as a tunnel barrier on silicon substrates, in which the TER can be quantitatively described by standard electrostatic models [332]. In parallel, research efforts are concentrating on the integration of oxide-based FTJs on silicon [157,333] which would favor the realization of non-volatile memories or artificial neural networks combining ferroelectric devices with CMOS technology.

In conclusion, FTJs are promising devices to be used as memristors in neuromorphic architectures and as non-volatile memory elements. The devices show large OFF/ON ratios [309,327,330], low write energies [320], high uniformity [320,334] and can be switched at relatively high speed [320,334]. In contrast to the cycling performance of capacitive ferroelectric memories (> 1014 cycles) [335], endurance in FTJs has only been demonstrated up to a few 106 cycles [334], after which the switching becomes less deterministic, indicating a strong pinning of domain walls, probably due to creation of rearrangement of defects, such as oxygen vacancies. Additionally, the low resistance state requires higher voltage pulse amplitudes than before cycling, which could be related to an increased internal field by the movement of oxygen vacancies [335] which destabilizes the ON state. A possible means of slowing down the migration of oxygen vacancies and therefore enhancing this result is doping the ferroelectric material to reduce their mobility [335]. Besides, although submicron FTJs were realized by several groups [319–321,327] suggesting no limitation in downsizing them [327], their high-density integration requires further investigations to ensure switchable polarization for junctions with diameters smaller than 100 nm. Finally, FTJs also appear as a great fundamental tool to investigate minute changes at the interfacial electronic or magnetic properties or to probe domain dynamics in ultrathin ferroelectrics taking advantage of the high sensitivity of electron tunneling.

Acknowledgements

We acknowledge financial support from the ERC Consolidator grant “MINT” (contract number n° 615759) and the ANR project “FERROMON”. This publication has received funding from the European Union’s Horizon 2020 research innovation programme under grant agreement 732642 (ULPEC project). This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (Labex NanoSaclay, reference: ANR-10-LABX-0035).

Oxides for data storage: Ferroelectric RAMs

Uwe Schroedera,*, Thomas Mikolajicka,b

aNaMLab gGmbH, Noethnitzer Straße 64, 01187 Dresden, Germany
bChair of Nanoelectronic Materials, TU Dresden, 01062 Dresden, Germany
* Corresponding author.
E-mail: Uwe.Schroeder@namlab.com (U. Schroeder).

Abstract

The very low power programming operation together with long data retention capability make ferroelectrics very promising materials for non-volatile data storage applications. Current products on the market are limited by the properties of the ferroelectric PbZrTiO3 material and their non-compatibility with CMOS processing. Here, current ferroelectric memory cells are not scaling below 100 nm technology node mainly due to the very high physical thickness of the ferroelectric material and the non-solved issue of maintaining ferroelectricity at the sidewalls of 3D structures. These state of the art memory cells will be compared to hafnium oxide based devices, a material well known in today’s CMOS processing environment. In comparison to classical ferroelectrics, HfO2, when doped with a variety of different dopants, shows ferroelectric properties in a much thinner thickness range. Therefore, further scaling of ferroelectric memory devices could be enabled. This paper reviews the current status of hafnium oxide based memory devices, compares their properties to products on the market and describes a possible road-map for the future.

Keywords: PZT; Hafnium oxide; FeRAM; FeFET; Memory

1. Ferroelectric memories based on perovskite materials

First attempts failed to realize memories based on cross-point arrangements of ferroelectric BaTiO3 in the 1950s [336,337]. Since the late 1980s, ferroelectric properties utilized in a one transistor-one capacitor (1T1C) memory cell have been [338] as a possibility to store information for non-volatile memory (FeRAM) applications. Traditional ferroelectrics are characterized by a field driven switching, fast read/write access time (< 50 ns), non-volatile retention (up to 10 years), and possibility to endure a higher number of read and write cycles (up to 1014 cycles). The low write energy, fast read/write access time and high read/write cycle endurance are specific advantages over currently used flash memory devices (write/erase times in the 10s ms range, typically 105 write/erase cycles). The main reason for the preferred usage of flash memories is their much better scalability. Accordingly, a higher memory array size could be realized (many Gb for flash vs. few Mb for FeRAM per single chip). Current FeRAM products use a Pb(Zr,Ti) O3 (PZT) perovskite based ferroelectric. Other materials were screened for this application e.g. BaTiO3 (BTO), SrBi2Ta2O9 (SBT), and BiFeO3 (BFO) [339–343] but could not demonstrate better memory properties. Typically, crystallization anneals above 600°C are needed to reach the ferroelectric phase. Moreover, ferroelectric materials show a severe degradation when annealed in hydrogen containing ambient [344]. Since embedded planar capacitor FeRAM cells using PZT are integrated...
in a complementary metal oxide semiconductor (CMOS) flow at the 130 nm node, the high crystallization anneals and the hydrogen sensitivity require both changes of the front-end and back-end of line (BEOL) CMOS flow.

In current products, a planar capacitor structure with a typically 70 nm thick PZT film is used [342,345]. For future scaling of the FeRAM cell below 130 nm technology node a reduction of the physical thickness of the ferroelectric (FE) material is required which is hard to achieve [346]. For scaled capacitors a 3D structure is necessary to keep the capacitor area constant when a smaller footprint is available. A PZT layer with this thickness would not fit into scaled 3D capacitor structures below 130 nm node. Moreover up to now, it was not possible to achieve FE properties in PZT at the sidewalls of 3D capacitors in scaled down structures [347,348] in contrast to a polar behavior for SBT in low aspect ratio devices [349].

An alternative approach that promises better scalability is the integration of the ferroelectric into the gate stack of a MIS transistor leading to a ferroelectric field effect transistor (FeFET) [350]. However, the high permittivity of perovskite materials leads to a depolarization field if a material with low permittivity like an interface oxide or a depletion layer is connected in series. As a result, long retention times of the stored polarization state are very difficult to achieve in scaled down FeFET devices using perovskites [351]. Although, first non-volatile retention values can be reached, but a 10 × higher field needs to be applied to a HfO2 based dielectric compared to PZT to switch polarization direction.

This improves retention for HfO2, but degrades the lifetime of the device resulting in an earlier breakdown of the dielectric during field cycling. In literature, endurance values of typically $10^9$-$10^{10}$ field cycles [355] are reported for doped HfO2, which need to be improved in relation to up to $10^{14}$ cycles for PZT based devices [341,361]. In contrast to these values for doped HfO2, the same structure with anti-ferroelectric Si doped HfO2 or Hf$_x$Zr$_{1-x}$O$_2$ showed drastically improved endurance behavior. Recently, a new concept of an anti-ferroelectric RAM (AFE-RAM) has been introduced [362], which used the enhanced cycling performance of an AFE ZrO2 based capacitor. Here, a non-volatile memory performance could be demonstrated by shifting the

### Table 1
Comparison of current and future 1T1C FeRAM and 1T FeFET memory cell concepts.

<table>
<thead>
<tr>
<th></th>
<th>FeRAM PZT [341]</th>
<th>FeRAM HfO$_2$ [355]</th>
<th>AFE RAM ZrO$_2$ [362]</th>
<th>FeFET HfO$_2$ [367]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarization $P_r$ [$\mu$C/cm$^2$]</td>
<td>$-20$</td>
<td>$-20$</td>
<td>$-10$</td>
<td>$-20$</td>
</tr>
<tr>
<td>Endurance [cycles] w/o fatigue</td>
<td>$10^{15}$-$10^{15}$</td>
<td>$10^9$</td>
<td>$&gt;10^{12}$</td>
<td>$10^9$</td>
</tr>
<tr>
<td>Coercive Field $E_c$ [MV/cm]</td>
<td>$0.1$</td>
<td>$1-1.4$</td>
<td>$0.8$</td>
<td>$1-1.2$</td>
</tr>
<tr>
<td>Technology Node [nm]</td>
<td>130 planar device</td>
<td>30 3D device</td>
<td>30 3D device</td>
<td>30 Planar/3D</td>
</tr>
<tr>
<td>Memory size</td>
<td>4-8 Mb</td>
<td>$&gt;4$ Gb</td>
<td>$&gt;4$ Gb</td>
<td>eNVM 3D NAND</td>
</tr>
<tr>
<td>Retention [years @85°C]</td>
<td>$&gt;10$ [359]</td>
<td>$&gt;10$ (MW90%) [631]</td>
<td>$10$ (MW 50%) [363]</td>
<td>$10$ (MW 50%) [360]</td>
</tr>
</tbody>
</table>

In 2011 [353] many of the issues of traditional ferroelectric memories can be solved. The material can be deposited by atomic layer deposition (ALD) and ferroelectric properties within a 3D capacitor structure are already proven [354]. Depending on the dopant material in HfO$_2$ [355] the ferroelectric properties can be reached in a different doping regime (e.g. Si [385], Al [385], Gd [385] $<3-30$cat%, La $<3-15$cat% [356], Zr $<30-70$cat% [357]) and after different crystallization temperature anneals (e.g. Zr, Gd $<500$ °C, Si, Al, La, Sr $<800$ °C). Accordingly, different non-volatile memory device applications with different thermal budget requirements may require different dopants in HfO$_2$.

For a 1T1C FeRAM memory cell in a back-end application dopant materials resulting in a low crystallization temperature like Zr or Gd are beneficial. If the material needs to withstand higher temperatures during the further processing as it is typical in a gate first ferroelectric field effect transistor doping with silicon is a suitable option [358]

Table 1 shows a comparison of ferroelectric properties for a PZT based perovskite material in relation to HfO2 or ZrO2 based ferroelectrics. Similar remanent polarization values can be reached, but a 10 × higher field needs to be applied to a HfO2 based dielectric compared to PZT to switch polarization direction.

This improves retention for HfO2, but degrades the lifetime of the device resulting in an earlier breakdown of the dielectric during field cycling. In literature, endurance values of typically $10^9$-$10^{10}$ field cycles [355] are reported for doped HfO2, which need to be improved in relation to up to $10^{14}$ cycles for PZT based devices [341,361]. In contrast to these values for doped HfO2, the same structure with anti-ferroelectric Si doped HfO2 or Hf$_x$Zr$_{1-x}$O$_2$ showed drastically improved endurance behavior. Recently, a new concept of an anti-ferroelectric RAM (AFE-RAM) has been introduced [362], which used the enhanced cycling performance of an AFE ZrO2 based capacitor. Here, a non-volatile memory performance could be demonstrated by shifting the
polarization hysteresis due to introduction of an internal bias field caused by two electrodes with different workfunction. This allows switching between a polarized and a non-polarized state, which reduces caused by two electrodes with different workfunction. This allows reduction of the retention behavior is determined with a reduced memory window (MW) after 10 years [363]. But, for the new HfO2 and ZrO2 based memory cell concepts not all devices criteria are characterized yet. Specially, the imprint behavior might be critical due to the high amount of trap sites in the electrode/dielectric interface region and within the HfO2 itself [364].

The discovery of ferroelectricity in HfO2 have also significantly changed the prospects of ferroelectric field effect transistors [365].

The retention issue is solved by the lower permittivity and the higher coercive field [366]. Already in 2012, first scaled devices using 28 nm technology have been demonstrated [367] and a full integration of memory arrays together with CMOS logic was shown in 2016 [358]. Current devices are mainly limited by low endurance values caused by charge trapping in the interfacial SiO2 resulting in a closure of the MW [354,360].

3. Concluding remarks

Ferroelectrics are ideal candidates for low power nonvolatile memories. However, the integration challenges of perovskite materials have limited their usage to niche applications so far. The discovery of ferroelectricity in doped hafnium oxide, a standard material for state of the art integrated circuits, has brought new fuel to this technology. Fig. 26 gives a schematic view of the historic development together with a roadmap for new ferroelectric memory devices.

Acknowledgements

This work was supported in part by the EFRE fund of the European Commission, by the Free State of Saxony (Germany), and by funding of the Deutsche Forschungsgemeinschaft (DFG project: Inferox)

Alternative logic concepts using oxide-based electronic devices
Stephan Menzel a,*, Anne Siemon b

aPeter Grünberg Institut (PGI-7), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
bInstitut für Werkstoffe der Elektrotechnik (IWE 2), RWTH Aachen University, 52066 Aachen, Germany
* Corresponding author.
E-mail: st.menzel@fz-juelich.de (S. Menzel).

Abstract

The increasing demand for data-storage and for processing big data drives the miniaturization of CMOS logic and conventional data storage. As CMOS scaling reaches its physical limits, alternative logic devices are extensively studied. Moreover, computing big-data suffers from data being transferred between CPU and main memory in conventional von Neumann computing architectures. Thus, alternative logic devices that can facilitate near-memory or in-memory computing architectures are of high interest. In these architectures, also called processing-in-memory architectures, a light-weight processing unit is located close to the memory. This proximity leads to a very fast data transfer. In an ultimate computing architecture the discrimination between data and storage could become blurred if devices are used that offer both logic functionality and data storage capability.

This article gives an overview of alternative logic devices and concepts based on oxide electronic devices. The discussion focuses on logic concepts based on bipolar resistive switching devices (BRS), which offer logic-functionality and data storage capability at the same time. Resistive switching devices consist of a simple metal/insulator/metal structure, where the insulating material is typically an oxide. In these devices the data is stored in form of different atomic configurations that encode different resistance states. By applying a bipolar voltage scheme the resistance can be switched between a high resistive (HRS) state and a low resistive state (LRS). Based on these devices different logic-in-memory concepts will be presented and future prospects and challenges for this emerging technologies will be discussed.

Keywords: Logic-in-memory; Non-von-Neumann architectures; Beyond CMOS; Resistive switching; ReRAM

1. State of the art

A variety of alternative logic devices based on oxide electronics have been proposed in literature (e.g. in [368–371]). In general, one can distinguish between transistor-like approaches, threshold-like approaches (mainly all-spin logic devices) or hybrid nonvolatile passive memory logic devices. These devices use different state variables such as spin, ferroelectric polarization, atomic configurations etc. to process the logic functions. When discussing alternative logic devices one important question is whether the alternative device is supposed to replace current CMOS logic completely or should it only complement CMOS for specific tasks. Either way there are five criteria that are needed to be fulfilled for all logic approaches [372]:

1. First of all, logic concepts should be able to implement a full set of Boolean functions, e.g. AND/NOT or OR/NOT, or IMP/NOT (functional completeness).
2. Secondly, a nonlinear transfer characteristic is required to obtain a good signal to noise ratio.
3. In order to refresh smeared out signals the devices should enable power amplification.
4. Moreover, an output signal should not influence the input signal of a logic gate. Thus, an efficient feedback suppression is desired.
5. Finally, inputs and outputs of the logic devices should be compatible, thus a transfer of information is possible. This requirement is referred to as concatenability.

If these criteria are met, an arithmetic logic unit (ALU) can be designed using the alternative logic approaches. An ALU normally includes at least a full-adder, a multiplexer, an AND gate and a NOT gate.

One of the most important still unanswered question in this field is how efficient these implementations are. Nikonov et al. revised several alternative logic concepts for replacing CMOS and provided a suitable benchmark, which included estimates of standby power, speed and energy for a wide set of circuits including an ALU [368]. Ferroelectric-based logic devices were identified as promising class of nonvolatile devices, but suffering from speed limitations. Spin-tronic circuits are good candidates for low standby power applications and they can also offer high speed operation [373]. Nikonov et al. further concluded that interconnects will dominate switching energy and delay. The latter result was also found when considering extremely scaled memory arrays [374].

Instead of replacing CMOS completely, hybrid computing architectures have been proposed. In these hybrid architectures mostly passive nonvolatile logic devices are operated using CMOS devices as active devices. Since passive devices could not meet the power amplification criteria and the concatenability criteria, these functions are taken over by the CMOS part. The benefit of these approaches is, however, that the memory devices are able to store data and to perform logic operation at the same time. Thus, logic-in-memory concepts can be realized.
Recently, several logic-in-memory concepts based on resistive switching oxide BRS devices have been discussed, i.e., MAGIC [375], CRS-logic [376], Scouting logic [377] and MRL [378]. These logic concepts can be classified according to their working principle. The most common working principles are depicted in Fig. 27 and are based on (a) a voltage divider effect, (b) writing, (c) current detection, (d) a CMOS-like working principle or (e) threshold gates. In Fig. 27 the inputs of the example logic gates, p and q, have a green background and the outputs are marked by the red background. Here, a logical zero 0 is either decoded as HRS, if it is a resistance, or as a low voltage level. A logical one 1 is defined as LRS or high voltage level.

Fig. 27(a) shows the circuit of an IMPLY gate [379] as an example for a voltage divider structure. Two different voltages $V_{COND}$ and $V_{SET}$ are applied to the two BRS devices, P and Q. Here, $V_{SET}$ is higher than the switching voltage $V_{VRS→LRS}$, which is itself larger than $V_{COND}$ ($V_{SET} > V_{VRS→LRS} > V_{COND}$). Since the applied voltages create a voltage drop over the resistance $R_Q$, which depends on the input resistances of device P $R_P$ and device Q $R_Q$, the voltage drop over Q can be above or less than $V_{VRS→LRS}$. Depending on this voltage, $R_Q$ may be switched, thus the result of the IMPLY function is resembled by $R_Q$. In Fig. 27(b) the structure of the CRS-logic [376] is depicted as an example for the Writing-style logic family. The inputs are the former resistance state $Z$ of the BRS device and the applied voltages $V_P$ and $V_Q$, which are set according to the desired logical inputs. Similar to writing data, the voltages $V_P$ and $V_Q$ are applied to the device. By using different schemes for applying $V_P$ and $V_Q$, Boolean functions can be implemented in several sequential steps. Fig. 27(c) shows the scouting logic circuit [377] as an example for a current-based approach. Here, the resistance states of the BRS devices are used as inputs. The bitline current is a function of these inputs due to a simultaneous read-out of several devices connected to the same bitline. This current is compared to a reference current I_{Ref} by an amplifier. The output of this comparison is the result of the desired function. Depending on $I_{Ref}$ different Boolean logic functions can be achieved. In Fig. 27(d) the MRL [378] approach is depicted as an example for CMOS-like approaches. In this approach the applied voltages represent the inputs. Depending on the applied voltages and the arrangement of the BRS devices, the devices switch under different conditions. Due to the anti-serial connection of two BRS devices, one device is in the LRS and the other one is in the HRS. Thus there will be, similar to CMOS, either a low ohmic path from the output to GND or from the output to $V_{DD}$. Fig. 27(e) shows a typical structure of a threshold-gate logic. Here, the inputs can be the applied voltages or the resistance states of the BRS devices or both. The applied voltages are weighted by the BRS resistance states, so that the potential of the threshold gate input is above or below its threshold and the output adjusts accordingly. The simplest threshold gate is an inverter as depicted in Fig. 27(e).

Table 2 provides the categorization of the different logic approaches based on their working principles. Moreover, it summarizes if the approaches are array-compatible.

As discussed in Section 1, the basic components of an ALU are an adder, a multiplexer, an AND gate, and a NOT gate. Fig. 28 shows the experimental realization of all these components except for the multiplexer in CRS-logic on small arrays [385]. The multiplexer function is realized by the CMOS control unit, which coordinates the memory arrays (cmp. Fig. 28(a)). Thus, a simple ALU using the CRS-logic is proven to be possible. The efficiency of such an ALU, however, is still questionable and heavily depends on the CMOS control unit.

2. Future prospects and challenges

While some of the logic-in-memory concepts have been already demonstrated experimentally (CRS and IMPLY logic), most of the proposed approaches have been only discussed on a conceptual level or by simulation. The validation of a logic-in-memory concept is limited by the accuracy of the underlying resistive switching device model. In some cases the resistive switching device was modeled as a binary device that switches from one state to the other when the applied voltage exceeds a threshold voltage. This approximation covers the dynamics of a real resistive switching device insufficiently. In fact, no real threshold exists but the switching dynamics are highly nonlinear [386]. A small change in the applied voltage leads to a large change in switching time.

Table 2 Overview of different logic-in-memory approaches based on resistive switching devices.

<table>
<thead>
<tr>
<th>Logic approaches</th>
<th>Logics</th>
<th>Array-compatible</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Voltage divider</td>
<td>Magic [375]</td>
<td>Yes</td>
</tr>
<tr>
<td>(b) Writing-style</td>
<td>IMPLY [379]</td>
<td>Yes</td>
</tr>
<tr>
<td>(c) Current-based</td>
<td>Snider [380]</td>
<td>No</td>
</tr>
<tr>
<td>(d) CMOS-like</td>
<td>Extended CRS-logic [371]</td>
<td>Yes</td>
</tr>
<tr>
<td>(e) Threshold logic</td>
<td>Pershin et al. [381]</td>
<td>Yes</td>
</tr>
<tr>
<td>(f) MRL</td>
<td>Vourkas et al. [382]</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2 provides the categorization of the different logic approaches based on their working principles. Moreover, it summarizes if the approaches are array-compatible.
This means that a resistance switching is still possible at low voltages, which could lead to unwanted state drifts in partially-biased cells [387]. Hence, it is important to develop realistic device models and then validate the different approaches using these models.

As many experimental phenomena in particular switching variability are physically not understood, the validation of logic concepts should not be based on simulations alone. It is indispensable to proof the functionality experimentally. In a first step this could be done in a small circuitry consisting of only few devices (see Section 1), but finally the experimental proof should be given on array size to consider all parasitic effects. Having in mind the cycle-to-cycle and cell-to-cell variability, in particular logic-in-memory approaches based on a voltage divider concept could fail in large-scale circuits. Also sneak paths existing in large arrays could disturb current-based logic approaches. Besides the proof of concept on an array level, it is necessary to consider the whole hybrid computing system. Here, the overhead in the CMOS control unit is an important factor, which is not investigated enough till now. Thus in future, suitable benchmarks should include the required control logic.

Recently, Le Gallo et al. demonstrated a mixed-precision computing approach [388]. The hybrid architecture consists of a high-precision computing von Neumann machine and a low-precision memory computing system. This system has been employed to solve linear equations with mixed-precision. In this example the memory array performs a matrix-vector multiplication. The used memory array consists of phase-change memory devices, which in fact do not employ oxide materials. Matrix vector-multiplications, however, have been demonstrated using oxide-based resistive switching devices for neuromorphic applications [389,390]. Already today, the novel logic devices demonstrated in this example show a highly efficient acceleration of a specific task and proved to be worthwhile for further investigations.

3. Concluding remarks

In this paper different logic-in-memory approaches using oxide-based resistive switching devices are reviewed. The approaches show promising prospects for future hybrid-computing structures. However, further research is required to evaluate whether hybrid-computing will be only used to accelerate specific tasks or is efficient even for more general computing tasks.

Acknowledgements

This work was supported in parts by the Deutsche Forschungsgemeinschaft (DFG) under Grant SFB 917 ‘Nanoswitches’ and the COST Action MP1308 ‘TO-BE (Towards Oxide-Based Electronics)’.

High-k dielectrics for CMOS and emerging logic devices
Marco Fanciulli * Department of Materials Science, University of Milano Bicocca, Milano, Italy MDM Laboratory, IMM-CNR, Agrate Brianza, Italy *Corresponding author. E-mail: marco.fanciulli@unimib.it

Abstract

High dielectric constant (high-k) materials development has been pivotal in Moore’s scaling of CMOS logic to address short-channel effects leading to source-drain leakage. High-k dielectrics technology is crucial to CMOS scaling down to the ultimate node. In addition, the maturing material science and technology of high permittivity materials found several other applications in emerging logic and memory devices for classical and quantum information processing within von-Neumann and non Von-Neumann schemes, as well as in other application areas such as spintronics, energy harvesting and production, sensors, and neuromicroelectronics. This paper focuses on a brief description of the state of the art and future prospects of high-k dielectrics for devices with logic functionalities.

Keywords: High-k dielectrics; CMOS scaling; Emerging logic devices

1. State of the art

The reduction of the gate length in a bulk CMOS, as determined by Moore’s law, leads to short channel effects (SCE) when the edge of the depletion region at the drain is so close to the source that source-drain leakage increases and the transistor threshold voltage decreases. For bulk
In CMOS, a solution is to increase the doping level in the channel in order to reduce the depletion width at the drain junction. By doing so, the capacitance coupling between gate and channel as well as the fraction of voltage falling across the semiconductor channel decrease slowing down the transistor switching speed. High doping level in the channel results also in a lower carrier mobility. The reduction of the thickness of the gate oxide (SiO$_2$) increases the gate capacitance making the oxide capacitance higher than the channel capacitance providing a larger share of the applied gate voltage to the semiconductor. The problem arises when the thickness of the dielectric falls below ~1 nm where direct tunneling, leading to gate leakage, undermines the effort to achieve better device performances. The situation is quite common along the Moore’s law road where an improvement in one direction often implies a worsening in another. The introduction of high-k dielectrics as gate oxides allows, maintaining the same required gate capacitance, a thicker dielectric, thus avoiding, in principle (see discussion later on), gate tunneling. The tunneling probability depends, indeed, not only on the physical thickness of the dielectric, but also on the effective mass of the carriers passing through it and on the energy barrier heights separating the semiconductor channel from the gate contact [391–393]. In this context one can introduce a figure of merit $f = \frac{k}{m_0}$, where $k$ is the dielectric constant of the gate insulator, $m_0$ is the effective tunneling mass, and $E_b$ is the energy barrier height formed at the gate dielectric/substrate interface. A larger $f$, at a given equivalent oxide thickness, results in a more significant reduction of the tunneling gate leakage [393].

The introduction of high-k dielectrics as gate oxides is therefore motivated by the control of the saturation current, the sub-threshold current, and the gate current. Within the gradual channel approximation, the drain current can be written as [394]

$$I_D = \frac{W}{L} \mu C_{inv} \left(V_G - V_T - \frac{V_D}{2} \right) V_D$$

(1)

where $W$ is the width of the transistor channel, $L$ is the channel length, $\mu$ is the channel carrier mobility, $C_{inv}$ is the capacitance den-sity associated with the gate dielectric when the underlying channel is in inversion, $V_G$ and $V_D$ are the voltages applied to the transistor gate and drain, respectively, and $V_T$ is the threshold voltage. The saturation current $I_{D,sat}$ is obtained when $V_G - V_T = V_D$ in Eq. (1) and is determined the ON state of the transistor. The OFF state is determined by the sub-threshold drain leakage current [395]:

$$I_{D,subt} = \frac{W}{L} C_{dep} V_T \frac{V_D^{1+\gamma}}{V_D^{1+\gamma} - V_T^{1+\gamma}} \left(1 - e^{-\frac{V_D}{\gamma V_T}}\right)$$

(2)

where $C_{dep}$ is the capacitance of the depletion region, $V_T$ is the ther-mal voltage (= kT/q), and $\gamma = 1 - C_{dep}/C_{inv}$ is the sub-threshold parameter. From $I_{D,subt}$ it is possible to define the sub-threshold slope (ss) as

$$ss = \frac{1}{\ln(10)} \frac{\partial I_{D,subt}}{\partial V_G}$$

(3)

This parameter indicates how sharp is the switching off of the transistor by reducing $V_D$ and must be as small as possible. A reference target value is ss = 60 mV/dec. Reduction of ss can be achieved by decreasing $C_{dep}$ or increasing $C_{inv}$. All these parameters call, upon CMOS scaling, for the introduction of a gate oxide with a dielectric constant higher than that of SiO$_2$. Another important device parameter is the natural length [396,397]:

$$\lambda_n = \frac{f_g}{\sqrt{\varepsilon_{g,t}}}$$

(4)

where $g$ indicates the equivalent number of gates ($g = 1$ or bulk MOSFET, $g = 2$ for double gate MOSFET, …), $\varepsilon$ the dielectric constant, and $t$ the physical thickness. This parameter represents the length of the region of the channel that is controlled by the drain. A device is free of short-channel effects if the effective gate length of a MOSFET is larger than 5–10 times the natural length [396,397] This requirement clearly indicates that a thin channel (SOI), multiple gates, and high dielectric constant materials allow scaling avoiding SCE.

Table 3 gathers some selected requirements for the materials and the device at technology nodes envisaged for the period 2013–2027 [398].

When addressing the search for the optimum high-k material to be used in CMOS scaling, however, several other factors must be taken into account in addition to the dielectric permittivity:

- large band gap
- high band offsets
- large effective mass in the dielectric (tunneling)
- channel mobility comparable to that of SiO$_2$
- thermal and chemical stability in contact with semiconductor (silicide/silicate formation)
- thin and stable interfacial SiO$_2$ (or native oxide)
- scalable equivalent oxide thickness EOT < 1 nm
- compatibility with gate electrode material
- density of interface states comparable to that of SiO$_2$
- low lattice mismatch and similar thermal expansion coefficient with Si
- negligible capacitance-voltage hysteresis (< 20 mV)
- good reliability (no charge trapping, high breakdown voltage, etc.)

Unfortunately, the requirements are not compatible with each other. A high dielectric constant means, in general, a low band gap and a small effective mass and therefore high tunneling. In addition, the high-k implies soft LO phonons which lead to a reduction of the channel mobility. In general, with the increase of the metal atomic number, the metal ionic radius increases, but the cohesive force decreases in the metal oxide, leading to a large dielectric constant. The band-gap energy decreases with increasing dielectric constant. A narrow bandgap results in smaller energy band offsets with Si. A band offset less than 1.0 eV may lead to an unacceptably large leakage current, see Table 4. TiO$_2$ and barium strontium titanate (BST), showing profoundly higher permittivity than SiO$_2$, were reported not to be thermally stable with silicon substrates, but even more important are the following issues: (i) TiO$_2$ has revealed almost zero CBO leading to high leakage current in MOS [399]; (ii) in general very high dielectric permittivity causes field induced barrier lowering (FIBL) which degrades short channel effects of MOS transistor [400]. Interface states also play a significant role in determining the gate performance. From the microscopic point of view, the dom-inant defects at the high-k/silicon interface have been found to be the Pb centers, well known in the Si/SiO$_2$ system. This is due to the fact that at the high-k/silicon interface a sub-silicon oxide (SiO$_x$) is almost always present and that the localized wave function of
the dangling bond at the interface – the Pb center – is not sensitive to the chemistry of the high-k layer standing on top [401].

High-k gate dielectrics revealed a stronger scattering of the channel-nanocharges compared with SiO2. This increased interaction has been attributed to remote effects such as remote phonon scattering and remote coulomb scattering. Fischetti et al. [402] reported a theoretical-investigation showing that for softer materials (and the high-k dielectrics have softer modes leading to the higher dielectric constant) with LO phonon energy around 50 meV the strong electron-phonon coupling degrades the mobility. Later it was argued [403] that remote phonon scattering had a weaker effect than predicted by Fischetti et al. and that remote Coulomb scattering with oxide fixed charges could explain the mobility degradation. However, a very high concentration of charged defects, not consistent with experimental data, was needed to explain the observed mobility degradation. Although remote scattering mechanisms play an important role, additional processes, not yet completely identified, must be assumed to determine the mobility in MOSFETs with high-k dielectrics [404].

Hafnium-based gate dielectric materials, first introduced by Intel in its 45 nm high volume manufacturing process in 2007, are still used in current technology. The large dielectric constant, originating from a large polarization of the Hf-O molecule, and the band-gap contributed to establish HfO2 and HfO2-based materials (silicates, incorporation of N or Al in HfO2) as the most promising high-k gate dielectric on silicon. Improvement of HfO2 characteristics as gate dielectrics has also been achieved stabilizing the tetragonal and the cubic phases which show a larger dielectric constant compared to the monoclinic phase. A review of the different properties and applications of Hf-based dielectrics can be found in Ref. [405].

Rare earth oxides [406] and in particular lanthanum-based ternary oxides may play a major role in meeting the ITRS requirements for scaling beyond the 14 nm node.

Device performance, according to Eq. (1) can be improved upon scaling by including high-k gate dielectrics (C(k)), but also exploiting high-mobility (µ) channel materials such as Ge and/or GaAs which can be integrated on the silicon wafer. This is the current trend for high performance devices. Before moving to new channel materials the band structure of silicon can provide the last opportunity for this material in terms of mobility. Strain removes the sixfold degeneracy of the conduction band of silicon reducing scattering mechanisms and increasing the charge mobility. It must be pointed out that the interface between strained silicon and oxides revealed different band offsets and a reduction of the density of interfacial states.

In contrast with silicon, high mobility semiconductors such as Ge, GaAs, In0.53Ga0.47As do not have an electrically good native oxide. The development of high-k dielectrics therefore naturally matches the introduction of high-mobility channel materials provided that efficient passivation procedures are developed. The interface between high-k and high-mobility materials is characterized often by a large density of interfacial defects which may lead to the so called pinning of the chemical potential making impossible to drive the transistor [407]. For this reason, Ge and GaAs are proposed as p – and n – channel materials respectively. Remarkably the microscopic nature of these defects is much less known compared with the Si/SiO2 or Si/ high-k cases. Only recently, for example, the Ge dangling bond at the Ge/oxide interface, analogous to the well-known Pb center at the Si/oxide interface, has been revealed by electrically detected magnetic resonance [408]. It should be pointed out that high-k materials appropriate for Si do not necessarily perform well on Ge or III-Vs, and vice versa.

Interface passivation is one of the most important challenges facing the integration of high-k dielectric on high-mobility channels. Ge and III-Vs channels reveal a density of interface states in the 1013 cm−2 range, much larger than the Dg of the reference Si/SiO2 forming-gas treated interface typically below 1010 cm−2. In the specific case of Ge, recent progress has been achieved regarding interface passivation adopting either ultrathin Si capping layer, or thermally grown GeO2 or GeON interfacial layers, or PH3 or H2S sur-face treatments. Few high-k dielectrics, in particular rare-earth oxides (La2O3) deposited at high temperature or subjected to post deposition annealing, have also been provided to improve Ge surface passivation (Dg in the 1011 eV−1 cm−2 range), as com-pared to HfO2 or ZrO2. However, an estimated value of the dielectric constant of about 9 and no evidence for an interfacial layer were explained in terms of a reaction between La2O3 and Ge substrate to form a low k and leaky La–Ge–O germanate over the entire film thickness, limiting gate scalability [409].

The most used strategies for passivating III-V channels are based on sulfoxides or on a very thin amorphous or crystalline Si layer as an interfacial control layer between GaAs and SiO2, Si3N4, and HfO2 [410]. Low interface trap density was demonstrated at Bell Labs already in 1996 by depositing Ga2O3(Gd2O3) on GaAs in ultra-high vacuum [411]. Atomic layer deposition (ALD) of Al2O3 on GaAs has also shown good results in terms of leakage current and breakdown electric-field characteristics [412], while the reactivity of the As-capped In0.53Ga0.47As(001) surface and the electrical quality of the interface with ALD grown Al2O3 were found to be dependent on the surface reconstruction [413]. Ge-based passivation has also been demonstrated on III-V substrates [414].

Bulk MOSFET scaling, despite the introduction of new materials such as the high-k dielectrics, cannot fulfill the challenging Moore’s law requests. New architectures, based on CMOS compatible pre-cessing, not affected by the intrinsic limits of bulk MOSFETs, have been proposed. Among these, multi-gate devices (Fin-FET, and Nanowire-FET with a gate all around geometry) are already substitute-bulking MOSFET exploiting enhanced electrostatic channel control and fully depleted SOI (FDSOI). In this context the control of the quality of the high-k dielectrics and of the interface with the semiconductor in terms of defects and native oxide interfacial layer formation appears as a challenging task.

A short review of the state of the art of high-k dielectrics for logic devices must stress the central role played by the main deposition method adopted also at the industrial level: ALD. This deposition technique, due to several key characteristics (high-conformality, low growth temperature, good control of the stoichiometry, opportunity for surface engineering and selective growth), in the last twenty years has imposed itself as a key enabling technology for the integration of high-k dielectrics not only for CMOS scaling, but also in several other evolutionary as well as revolutionary devices with logic, but not only, functionality.

2. Future prospects

With CMOS transistors reaching their fundamental physical limits due to scaling, various kinds of alternative memory and logic devices, going “Beyond CMOS” have been proposed. These nanode-vices take advantage of quantum mechanical phenomena, such as tunneling, charge properties, such as spin, quantum confinement effects (coulomb...
blockade, spin blockade), and ballistic transport characteristics. Some of these devices, for example tunneling-FET (TFET), or single electron transistors (SET), exploit quantum mecan-ical phenomena yet providing classical information processing with low power consumption. TFETs based on vertical silicon nanowires, are attracting a lot of interest due to their capability of providing low off current, ss not limited by KTs and a weak T dependence. Due to the indirect band gap of silicon TFETs exhibit low drive currents and the introduction of more complex heterostructures or III-V materials has been pur-suied as well. In any case for all these devices, high-k dielectrics play an important role as gate oxides as well as mitigate some of the problems related to the reduced size of the nanowire, for example the dielectric mismatch [415] between the semiconductor and the surrounding dielectric leading to an increase of the dopant activation energy and therefore to a reduction of the doping efficiency.

HfO2, ZrO2 and mixed Hf1-xZr2O revealed ferroelectric properties attributed to the non-centrosymmetric orthorhombic crystal form. Ferroelectricity was recently exploited in p-FET showing ss < 60 mV/dec using these dielectrics [416].

Among the novel, scalable, and energy-efficient computing technologies, many non-charged logic devices show intriguing opportunities. Spintronic/nanomagnetic devices, with electric field induced low energy and fast switching characteristics, in which the switching of magnetization by spin torque (ST) has been demonstrated to have relatively high switching energy which can be low-ered by exploiting magneto-electric (ME) switching achieved by placing other materials (mostly oxides) adjacent to the ferromagnetic layers, have been recently realized [417,418]. Several oxides play a crucial role in spintronic devices for classical information pro-cessing such as magnetic tunnel junctions and the major break-through in this field in 2005 was based mainly on the optimization of the ferromagnetic layer (Fe)/oxide(MgO) interface [419,420]. Although one can envisage magnetic-tunnel-junction (MTJ) devices fully based on oxides [421], so far this approach did not lead to functional properties comparable with those based on the Fe/MgO/Fe architecture.

Two dimensional materials (2D) are emerging in nanoelectronics. FET and TFET based on these materials are attracting a strong interest. The integration of oxides, in most cases deposited by ALD, is a key step for the final functional properties of these devices. Al2O3 capping of silicene for example unable the ex-situ characterization as well as the realization of the first transistor based on this material [422]. Transistors based on two-dimensional layered semiconductors such as transition metal dichalcogenides (2D-TMD: MoS2, WS2, MoSe2, . . .) have been also demonstrated. The lack of surface dangling bonds on one side has the positive effect of avoiding surface roughness scat-tering, however on the other side poses some problems in the integra-tion of high-k gate insulators which can be solved by introducing interfacial engineering processing steps which may lead to detrimental effects on the electrical properties of the 2D layer [423]. For Mo-and W-based dichalcogenides, in addition, MoO3 and WO3, respectively, are not good insulators and may even act as dopants. In this context a recent ad- vancement is the demonstration that some lay-ered 2D semiconductors such as HfSe2 and ZrSe2, have band gaps between 0.9 and 1.2 eV (bulk to monolayer) and technologically suitable high-k native oxides, HfO2 and ZrO2, respectively [423].

The application of high-k dielectrics in conventional and innova-tive non-volatile memories cannot be addressed here due to space limita-tions. It is worth mentioning that the realization of memristors based on metal-oxide-metal (MIM) structures integrated with con-ventional CMOS logic allows the development of neuromorphic com-putation as well as intriguing hybrids circuits for neuroelectronic applications possibly leading to neuro-prosthetic devices [424].

Last but not least the development of silicon qubits, either based on quantum dots or isolated donors in silicon, aggressively addressed in the last decade, relies on the well-established high-k dielectric tech-nology [425]. For qubits the interface between the dielectric and the semiconductor as well as the defects in the oxides play a crucial role in determining the (electron) spin coherence time and the requests for this class of devices are even more demanding than those imposed by the ultimate envisaged CMOS technology node. Another intrinsic source of decoherence for qubit schemes based on semiconductors or super-conductors are the dissipative two-level systems (TLS) in the amorphous dielectrics which micro-scopic nature has yet to be established [426].

3. Concluding remarks

The development of oxides with high dielectric constant, vigorously started at the end of the last century, was mainly motivated by the scaling of bulk CMOS logic and led to advancements in mate-rials science, technology, and device physics. This technologically driven effort revealed several additional remarkable opportunities in other areas such as non-volatile-memories, spintronics, quantum information processing, sensors, energy production and harvesting, and neurolec-tronics. Obviously the material optimization will depend on the specific application, and even for CMOS scaling the best solution may differ for high-performance or low-power devices. The main requirements for the high-k growth (i.e., conformality, low growth temperature) contributed to establish ALD as one of the lead-ing deposition methods, particularly when amorphous or nanocrystals-talline materials are required. Although several challenging tasks are ahead of us the maturing of these class of materials is clearly central in the development of ultra-scaled CMOS devices (FinFET, NW-FET, TFET) as well as emerging classical (SET, memristors) and quantum information processing (several qubit schemes) devices.

Oxide nano-electronics for neuromorphic computing
Julie Grollier*
Unité Mixte de Physique, CNRS, Thales, Univ. Paris-Sud, Université Paris-Saclay, 91767 Palaiseau, France
*Corresponding author.
E-mail: julie.grollier@cnrs-thales.fr

Abstract
This short article reviews the current progress in producing oxide-based artificial arrays of nano-synapses and neurons, then describes the future challenges towards assembling large-scale inte-grated neuromorphic computing systems.

Keywords: Neuromorphic computing; Memristors; Artificial synap-ses; Artificial neurons

1. Introduction

Most neural networks today are running on classical computers. Unfortunately, these hardware implementations are inefficient in terms of speed and energy consumption because of the von Neu-mann’s bottleneck. Achieving real-time, low power learning requires building parallel hardware networks of millions of interconnected artificial synapses and neurons. Oxide nanoelectronics can be help-ful in this context. Indeed, oxide-based nanodevices can be fabri-cated on top of the CMOS plane, as already done in the fabrication process of Resistive Random Access Memories (ReRAM) (see Section 14 of this Roadmap). This opens the possibility to bring mem-ory close to processing and compute with low power. In addition, oxide-based nanodevices can be fabricated in huge numbers on silicon: the most recent ReRAM prototypes feature several billions of memory cells [427]. Oxide-based nanodevices are also highly tun-able. By changing materials and inter-faces, their response can be designed, leading to superconductivity,
ferroelectricity, ferromag-netism or a combination of those [428]. This multifunctionality allows envisioning fabricating nano-synapses and nano-neurons with similar materials on-chip. Challenges include identifying the most promising among these effects towards low power reliable nano-synapses and neurons.

2. Synapses

Oxide resistive switching cells enable the implementation of artificial nano-synapses. These nano-resistors with memory, also called “memristors”, are non-volatile and tunable (Fig. 29(a) [429,430]). As such, their conductance can store the analog synaptic weights of artificial neural networks [431]. Large matrix networks of memris-tors, called crossbar arrays (Fig. 29(b)), can be fabricated to connect neurons fabricated in CMOS in the plane below the arrays. It has been estimated that the connectivity can in principle reach brain-like densities of 10,000 memristive synapses per neuron by stacking several crossbars on top of each other and carefully designing the electrodes [432–434]. Furthermore, as illustrated in Fig. 29(c), memristive crossbar arrays physically achieve the dot-product operation which is at the core of the inference process [432,433,435].

The resistive variations of most oxide memristors today rely on the electrical-field induced formation and subsequent manipulation of a nanometer-thick conductive filament (oxygen vacancies or metallic ions if electrodes are chemically active) in an insulating matrix sandwiched between two metallic electrodes [436]. Depending on the materials and prevailing physical effect leading to resistive switching, oxide memristors take different names (Conductive Bridge RAM, Electro-Chemical Cells, RedOx etc) as defined in Section 14 of this Roadmap.

Oxide memristors will need improvements to overcome their competitors including technologically-mature chalcogenide-based phase change materials [437,438]. Indeed, the filamentary nature of oxide memristors is a challenge for fabricating high density matrix suitable for applications. The highest hurdle is finding ways to reliably form filaments in every virgin cell of large crossbar arrays. Reducing the dispersion of ON and OFF resistance states and switching voltages is the second. For this purpose, researchers are designing forming-free oxide memristors in which resistance changes are due to a uniform drift of oxygen vacancies across the device section instead of forming a filament [439,440].

Fig. 29. Artificial synapses (a) Typical resistance versus voltage dependence of a bipolar non-volatile memristor (b) Schematic of a memristive crossbar array (c) Principle of inference with memristors.

The average cyclability of oxide memristors is also significantly lower than the one of phase change devices [441]. In order to gain in endurance, and possibly in switching speed, it can be interesting to design memristive devices which do not rely on large ionic motion to induce resistance variations but leverage instead electronic effects. It has for instance been demonstrated that the electronically-induced phase transition in strongly correlated electron systems gives rise to synaptic-like responses [442,443]. Alternativa-vely electrically-induced polarization switching in ferroelectric tunnel junctions induces large analog resistance variations thanks to the nucleation and propagation of ferroelectric domains in the thin active tunnel barrier [325]. Both types of artificial synapses are promising, but their compatibility with CMOS processes remains to be demonstrated before using them for applications.

Neuromorphic computing with memristors is still a nascent field. Experimental demonstrations of supervised learning with memristor arrays are recent. Networks of tens to hundreds of oxide memristors have been trained to recognize small images or electroencephalography signals [389,444–448]. In these demonstrations, learning was performed with supervised techniques such as back-propagation [449], and the memristor values were modified one by one by applying the required write pulses. Neurons were emulated using standard electronic equipment, such as arbitrary waveform generators or Field Programmable Gate Arrays. On-chip learning with co-integrated memristors and CMOS neurons remains to be demonstrated. Additionally, it has been highlighted that, while device to device variability is not an issue for neuromorphic comput-ing [450], noise and drifts in the same memristor can be detrimental to performance [451]. More problematically, non-linearities in the conductance versus voltage characteristics of memristors are strongly degrading the quality of learning with back-propagation [451,452]. Strategies have been devised to deal with this issue but they are slowing the process and costly in terms of energy consumption [452]. Fabricating memristor devices with above-threshold lin-ear characteristics through material engineering is therefore a relevant issue for implementing deep learning on chip [453].

Whereas the most powerful learning algorithms today are supervised, the future probably resides in unsupervised learning, through which the system learns by itself to detect the correlations in the input data, without the need of previous human-based labelling for error minimization. Memristor synapses could play a major role in this field, since they also allow a direct implementation of unsu-pervised learning rules [454,455]. In particular, it has been demon-strated that oxide memristors enable one of the most promising unsupervised learning rule today, called spike-timing dependent plasticity (STDP) [456]. STDP has been demonstrated experimentally in single devices with a wide variety of memristor types, including filamentary, uniform oxygen vacancies drift and ferroelectric oxide memristors [439,456–460]. The learning rule can be achieved by carefully engineering the neural pulse shapes [454,455], by leveraging the intrinsic stochasticity of the device [458] or by exploiting the different relaxation times of multiple state variables governing resistive switching [461,462]. Recently, learning through STDP has been achieved in small scale hardware neural networks combining memristor synapses with externally-emulated neurons [463,464]. Large scale simulations have highlighted that with a much larger number of memristors, typically 105 to 106, handwritten digit recognition or classification of car lanes in a video can be achieved [450,465]. Finally, alternative types of unsupervised learning rules are interesting to explore [466,467]. In the future, other handles can be used to enhance the plasticity of oxide nanodevices, using for example light, magnetic fields or other types of stim-uli [468].

As a conclusion, oxide memristive synapses open a path towards real-time supervised and unsupervised learning. Several challenges will have to be faced to achieve this goal. Devices need to be stabil-ized and their performances improved, notably in terms of power consumption and cyclability. Accurate modelling of the multi-faceted resistive switching mechanisms in these devices will have to be established, in particular in the form of compact models [469]. It will have to be demonstrated experimentally that crossbar arrays can connect neural layers composed of thousands of neurons in an all-to-all manner [432,470,471]. For this purpose, sneak path issues will need to be
Fig. 30. Artificial neurons (a) Typical resistance versus voltage dependence for a unipolar volatile memristor. The zone of negative differential resistance (NDR), is illustrated with a dashed line (b) Typical circuit to create oscillations (c) Typical resulting relaxation oscillations.

overcome, for example by engineering materials to include selector devices below each memristor [472]. Further-more, high connectivity will require going 3D, for example by stacking several crossbars on top of each other [434,473]. Finally, electronic chips interconnecting plastic memristive synapses and CMOS neurons will have to be built to tackle real-scale and real-time pattern classification [474].

3. Neurons

Oxide materials also offer a path for replacing micrometers-wide CMOS neurons with nanoscale electronic structures. Indeed, when thermal effects dominate, the same polarity of applied voltage or current allows switching back and forth between the ON and OFF states [429,441]. This kind of memristor, illustrated in Fig. 30a, is volatile, and called unipolar, in contrast to bipolar devices which require consecutive applications of opposite-sign programming pulses in order to switch back and forth. Unipolar oxide memristors display negative differential resistance (Fig. 30a), which, complete-ment of internal or external capacitances and resistors, can be used to give rise to electrical oscillations for a constant applied voltage V (circuit illustrated in Fig. 30b). Neuron-like relaxation oscillations and spiking (illustrated in Fig. 30c) have been evidenced in a number of oxide materials such as NbO$_2$ [475,476], VO$\_2$ [477], TaO$_3$ and TiO$_2$ [478]. The frequency is set by the time scale of the charge and discharge of the capacitor in the circuit [478]. Interestingly, these relaxation oscillators can exhibit chaotic dynamics, which can be useful to avoid getting stuck in local minima while learning [479]. Similarly to biological neurons, they also have the ability to synchronize [477,478], a key property which can be exploited for computing [480]. At the single device level, the challenge will be to improve the cyclability of these devices (for example, at 10 Hz, an endurance of $10^6$ cycles corresponds to a lifetime of about one day for the oscillator) as well as their frequency stability [481]. Similarly to the case of synapses, purely electronic memristors could be a promising route for progressing in this direction. For example, electronic induced switching in Mott insulators has been shown to mimic leaky integrate and fire neurons [482]. The next step will be to interconnect several of these neuron-like oscillators with oxide synapses [479,483]. Searching for solutions enabling both kinds of memristive behaviors (non-volatile and bipolar for synapses, volatile and unipolar for the neurons) with the same or compatible materials is an interesting path to build ultra-dense artificial neural networks. Finally, computing and learning through the coupled dynamics of these oscillators remain to be demonstrated.

4. Concluding remarks

Electronic oxide nanodevices can implement both neurons and synapses, opening the path to ultra-dense neuromorphic chips capable of online learning. Challenges on the way can be found at the device level (decrease power consumption, increase lifetime), at the array level (fabricate dense 3D crossbar arrays of these elements) and at the system level (cointegration with CMOS).

Acknowledgements

This publication has received funding from the European Union’s Horizon 2020 research innovation programme under grant agreement 732642 (ULPEC project).

Possible future quantum technologies based on correlated nanoelectronics

Guanglei Cheng$^{a,b,c}$, Jeremy Levy$^{b,c,*}$

$^a$CAS Key Laboratory of Microscale Magnetic Resonance and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China

$^b$Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA

$^c$Pittsburgh Quantum Institute, Pittsburgh, PA 15260, USA

$^*$Corresponding author.

E-mail: jlevy@pitt.edu (J. Levy).

Abstract

The advancement of oxide growth and device fabrication techniques has enabled the novel phenomena of correlated oxides to be controlled at nanoscale dimensions. This emergent field of correlated nanoelectronics has a great potential for applications in quantum technologies. Here we first review the status of quantum coherent transport in correlated oxides, with a focus on SrTiO$_3$-based electron systems. Then we outline possible routes towards quantum computing and quantum simulation.

Keywords: Correlated nanoelectronics; Quantum technology; Quan-tum computing; Quantum simulation

1. State of the art

Quantum-mechanical effects, e.g., tunneling through gate barri-ers in transistors, are conspiring to end a four-decade run of unprecedented growth in semiconductor-based technologies. At the same time, other quantum-mechanical effects, including quan-tum superposition and quantum entanglement, can potentially be utilized as the basis for new quantum technologies: quantum com-puting, quantum simulation, quantum state transfer, quantum sens-ing and quantum communica-tion. A principal challenge is to preserve quantum coherence which drives the core functionalities of quantum devices. Multiple platforms for realizing quantum technologies, including superconducting circuits, trapped ions, semiconductor quantum dots and nitrogen va-canies in diamonds, have been developed. Correlated nanoelectronics, which combines the para-digms of semiconductor nanoelectronics and correlated electronic materials (principally complex-oxides), brings many novel proper-ties to the table, and may be able to provide capabilities not present in other solid-state platforms.

Complex oxides possess nearly all properties of semiconductors as required for quantum technologies. Heterostructures can be engineered with atomic precision. One of the most widely studied sys-tems, SrTiO$_3$ (STO), has proven to be a highly promising material platform for
implementing correlated nanoelectronics. As shown in Fig. 31, STO can be patterned in nanoscale (quantum) devices reminiscent to semiconductors either by traditional lithography techniques or using a conductive-atomic force microscope (c-AFM) tip [484]. The latter method can routinely achieve sub-10 nm resolution. In the other hand, the inherent electron-electron correlations add rich features (e.g., magnetism, ferroelectricity, electron pairing and superconductivity) to the quantum devices. Such a combination makes STO a unique system for developing future quantum technologies.

Currently in STO-based electron systems, ballistic quantum transport in electron wave-guides and nanowires have been demon-strated [485,486]. Quantum Hall phases have been reported in δ-ped STO [487,488], which shows that the electron system can be engineered to have really high mobility and low carrier density. Quantum transport can be controlled at the single-electron limit using single-electron transistors (SETs) [489,490]. Strongly correlated phases have been identified in which electrons undergo pairing without superconductivity at the LAO/STO interface [490]. Superconducting quantum interference devices (SQUIDs) have been created using STO-based heterostructures, opening an avenue for all-oxide superconducting electronics [107]. Spin-to-charge conversion through the inverse Edelstein effect is observed, which is useful in probing spin states in spin-based quantum computing [36,491].

2. Current and future challenges

The DiVincenzo criteria [492] describe the basic requirements for building a universal quantum computer. Foremost in current stage, one needs a very robust qubit that can be easily scaled. So far, no plausible qubit has yet been demonstrated in correlated oxides. Several realizations are possible and might be more promising than the semiconductor count parts:

(i) Spin based qubits. Electron spins in c-AFM sketched quantum dots in principle can be used for spin qubit. The strong spin-orbit coupling at the LAO/STO interface can be also used to manipulate spins [493]. The advantages lie in the nanoscale reconfigurability of c-AFM lithography which can easily make ultrasmall dots and scale them. However, one remaining challenge is the cross-talk between local gates due to the large dielectric constant of STO at low temperatures [494]. The spin coherence time needs to be measured as well.

(ii) Topological qubits. Quantum error correction, one of the major challenges in quantum computing, requires significant physical resources (physical/ancillary qubits, realtime quantum measurement and feedback). Topological qubits can in principle provide intrinsic protection against many types of decoherence and even gate-induced errors [495]. Majorana zero modes obey non-Abelian statistics have some topological protection but are not universal. The first experimental evidence for Majorana zero modes was reported in 2012 [496], triggering a wave of interest and investment in possible applications in topologically protected quantum computation. LAO/STO nanostructures in principle possess all the ingredients necessary to create Majorana zero modes: low-dimensional superconductivity, strong spin-orbit coupling, and Zeeman interactions to remove spin degeneracy [497]. So far, no experimental signatures of Majorana zero modes in STO-based nanostructures have been reported.

Quantum simulation is another quantum technology that is distinct from quantum computation and yet offers insight into prob-lems (e.g., in materials science) that are intractable using state-of-the-art and foreseeable classical computing resources. Quantum simulators manipulate a reconfigurable quantum system to exhibit behavior that can be mapped formally (or approximately) onto a distinct physical system, to achieve insight into the underlying physical mechanisms (Fig. 32) [498]. One example would be a class of Hubbard models that have been proposed to give rise to high-temperature superconductivity.

Challenges for correlated nanoelectronics-based approaches to quantum simulation relate to incomplete understanding of the host quantum simulator (high “intellectual entropy”) as opposed to ultracold atomic quantum simulators, where the main challenge is reaching low effective temperatures. Ultracold atomic lattices presently have the best controllability and scalability, however, the operating tempera-tures are in fact too high to reach many interesting correlated phases. By contrast, a solid-state quantum simulator can reach much lower effective temperatures. In semiconductor quantum dots, small-scale Fermi-Hubbard models have been successfully demon-strated recently [499]. However, some Hamiltonians (e.g. BEC-BCS crossover) require the electron-electron interactions to be adjusted from attractive to repulsive, which is hard to realize natively in semi-conductors. In the LAO/STO nanostructures, the sign of electron-electron interaction can be tuned between repulsive and attractive [500]. Meanwhile, the high reconfigurability of c-AFM lithography allows in principle for precise engineering of a wide class of Hamiltonians.
3. Concluding remarks

Correlated oxide nanoelectronics combines the motif of semiconductor nanoelectronics with correlated oxides, and shows a promising potential for quantum technologies. Some key quantum devices including single electron transistors, nanowires and waveguides have been successfully demonstrated. More device functionalities that are absent in the semiconductor counterpart arise owing to the electron-electron correlations. Meanwhile, the c-AFM lithography technique provides a very precise and flexible means to engineer the desired Hamiltonian for quantum technologies. As a later player, the building blocks like a robust qubit and a simple solid-state quantum simulator are still ongoing experimental efforts.

Acknowledgements

G.C. acknowledges support from the Chinese 1000 Talents Plan for Young Scholars. J.L. acknowledges support from the Vannevar Bush Faculty Fellowship program sponsored by the Basic Research Office of the Assistant Secretary of Defense for Research and Engineering and funded by the Office of Naval Research through grant N00014-15-1-2847.

Epitaxial oxide barriers for magnetic tunnel junctions
Hiroaki Sukegawa*, Kazuhiro Hono
Research Center for Magnetic and Spintronic Materials, National Institute for Materials Science (NIMS), 1-2-1 Sengen, Tsukuba 305-0047, Japan
*Corresponding author.
E-mail: sukegawa.hiroaki@nims.go.jp (H. Sukegawa).

Abstract

A magnetic tunnel junction (MTJ) with a ferromagnet (FM)/oxide/ FM trilayer structure is an indispensable spintronic device for magnetic read sensors in hard disk drives and magnetic memory cells for magnetoresistive random access memories (MRAMs). The performance of an MTJ, in terms of the tunnel magnetoresistance (TMR) ratio, dramatically changes depending on the FM electrode and barrier materials, as well as their lattice-mismatch and crystal orientation relationship. Giant TMR ratios are observed in MTJs with a (0 0 1)-oriented crystalline barrier, such as MgO and MgAl2O4 (spinel), due to the spin-dependent coherent tunneling mechanism. In this chapter, we outline the recent progresses in epitaxial MTJ barrier technologies with MgO and spinel-based oxides. We also highlight the advantages of spinel-based barriers for future MTJ applications: good lattice-matching with various FM electrodes and tunability of barrier properties through compositional optimization.

Keywords: Magnetic tunnel junction; Tunnel magnetoresistance; MgO; Spinel

1. Introduction

Magnetic tunnel junctions (MTJs) are the most widely used spintronic devices that exhibit a magnetoresistance effect and are implemented as read sensors in hard disk drives and as memory cells of magnetoresistive random access memories (MRAMs) [501,502]. Since these junctions operate on the basis of magnetization reversal, they have features such as high-speed operation (ns scale), small device dimensions (on the order of ~10 nm), and infinite endurance. In addition, MTJs have been intensively investigated for applications such as highly sensitive magnetic sensors [503], non-volatile spinbased logics [504], and nanoscale microwave emitters that utilize the spin-torque oscillation effect [505]. A typical MTJ has a thin insulating layer (barrier) composed mainly of oxides that control its performance; therefore, development of new barrier materials and the associated growth technologies are important. For example, the discovery of the giant tunnel magnetoresistance (TMR) effect with a crystalline MgO(0 0 1) barrier has greatly accelerated MTJ-based device applications [420,506]. In this chapter, we introduce some state-of-the-art technologies and future challenges for oxide barriers for MTJs.

2. Current MTJ technology and future challenges

MTJ has a trilayer structure consisting of an insulator (barrier) sandwiched between two FM electrodes, as shown in Fig. 33(a). The tunnel resistance of the barrier changes with the relative magnetization between the two FM electrodes, as illustrated in Fig. 33(c): resistance change occurs when the magnetization configuration is switched between parallel (P) and antiparallel (AP) configurations under an external magnetic field or an electric current (spintransfer torque (STT) switching). This TMR effect is attributed to the spin-polarized tunneling current through FM/barrier interfaces. Assuming and are positive, the TMR ratio (%) can be defined as , where and are the resistances of the AP and P states, respectively, and and are the corresponding spin polarization of tunneling currents for FM1 and FM2 (Juliere model [507]). This suggests that a large TMR ratio will be obtained as and approach 1. Therefore, various half-metallic Heusler alloys have been employed as FM layers for MTJs; however, the large TMR degradation at room temperature (RT) restricts their implementation in practical applications [508].

In addition to the FM layers, selection of the barrier is also important since this layer determines the key features of MTJs such as a TMR ratio, value of resistance × junction-area (RA), and breakdown voltage. Typically, metal oxides have been employed as the barrier. Fig. 34 shows the development of the TMR ratio at RT for MTJs with various oxide barriers. In the 1990s, polycrystalline MTJs with an amorphous alumina (a-Al2O3) barrier were the subject of intensive study. However, the TMR ratios were limited to around 70–80% at RT because of the difficulty in obtaining sharp amorphous barrier/FM

![Fig. 33](image-url)
interfaces and the limited $P$ values (below 0.5) of typical 3d-transition FM's, such as Fe, Co, Ni and their alloys.

In 2004, large TMR ratios at RT were reported in MTJs with a crystalline MgO barrier: 220% in a textured Co$_2$Fe$_{20}$MgO/Co$_{56}$Fe$_{24}$B$_{20}$(0 0 1) MTJ [506] and 180% in an epitaxial Fe/MgO/Fe(0 0 1) MTJ [420]. The mechanism of the "giant TMR effect" in the MgO-based MTJ's is the additional spin filtering effect facilitated by the spin-dependent coherent tunneling through an epitaxial or a highly textured MgO (0 0 1) barrier [509,510]. Although the net spin polarization of bcc Fe is small ($P_{Fe}$ = 0.4 ∼ 0.45), the $\Delta_1$ Bloch states (spd symmetry) of Fe are fully spin-polarized due to the absence of the $\Delta_4$ states at $E_F$, as is observed in the [0 0 1] band dispersion of Fe in Fig. 35(a) [511]. Furthermore, as shown in Fig. 35(b), electrons with the $\Delta_1$ symmetry preferentially tunnel in the MgO(0 0 1) direction, whereas electrons in other Bloch states such as $\Delta_2$ (pd symmetry) and $\Delta_4$ (d symmetry), hardly tunnel. Therefore, the effective spin-polarization of tunneling electrons through Fe/MgO/Fe(0 0 1) becomes close to 1, leading to the giant TMR effect. Such a "pseudo-half-metallic" condition is satisfied even in CoFe(B) alloys and in some Co-based Heusler alloys. Nowadays, the CoFe/MgO/CoFe structure is used for most practical applications since the initial growth of the amorphous (a-)CoFeB layer can be achieved on any kind of underlayer [512]. On the a-CoFeB layer, a layer of (0 0 1)-textured polycrystalline MgO is grown. Upon crystallization of the a-CoFeB layer via annealing, an epitaxial relationship with the indi-vidual MgO crystals develops, resulting in a fully coherent MgO/CoFeB(0 0 1) interface. A TMR ratio as large as 60% has also been reported in a CoFeB/MgO/CoFeB MTJ [513]. A perpendicularly magnetized MTJ (p-MTJ) is a strong requirement for the reduction of the switching current density in STT-MRAMs. An MgO/CoFeB interface induces strong perpendicular magnetic anisotropy (PMA) [514], which enables the construction of p-MTJs with a large TMR ratio. Therefore, polycrystalline CoFeB/MgO/CoFeB MTJs are exclusively used in practical applications such as read sensors and STT-MRAMs.

However, further improvements to the TMR ratio and the reduction of RA are necessary to broaden the application for MTJs, parti-cularly to increase the capacity of STT-MRAMs. MgO shows a large lattice mismatch with various FM materials such as CoFe alloys, Heusler al-loys, and Mn-based alloys. This introduces many misfit dislocations near the barrier interfaces. For downsizing an MTJ cell to the order of 10 nm, the RA needs to be lowered via reduction in the barrier thickness; however, this causes degradation of device reliability (endurance) and a substantial reduction in the TMR ratios. To overcome these difficulties, many attempts have been made for the development of new oxide barrier materials. Recently, large TMR ratios exceeding 300% at RT have been reported in epitaxial MgAl$_2$O$_4$ (spinel) based barriers [515,516]. Similar to MgO(0 0 1), MgAl$_2$O$_4$(0 0 1) is predicted to exhibit the coherent tunneling through the $D_1$ states, enabling the giant TMR effect in these MTJs [517,518]. Since the mismatch between MgAl$_2$O$_4$ and typical FM materials is smaller than that between MgO and FM, it is possible to obtain a lattice-matched MTJ [519]. Fig. 36 shows cross-sectional annular dark field scanning transmission electron micro-scopy (ADF-STEM) images of the (a) Fe/MgO/Fe(0 0 1) and (b) Fe/ MgAl$_2$O$_4$/Fe(0 0 1) MTJs prepared by the sputtering method. With less than 1% bulk lattice mismatch between MgAl$_2$O$_4$ and Fe, as compared with that (∼3.8%) in a MgO/Fe(0 0 1) system, a perfectly lattice-matched barrier interface can be formed in the MgAl$_2$O$_4$ MTJ system (Outlined image, Fig. 36(d)). This high degree of lattice matching when using the MgAl$_2$O$_4$ barrier leads to excellent bias voltage dependence of the TMR ratio [520] and high breakdown voltages [521].

In addition to MgAl$_2$O$_4$, many other spinel oxides, $AB$_2$O$_4$ ($A$ and $B$: metals) are predicted to show the coherent tunneling effect due to the structural similarity [518]. Very recently, TMR ratios larger than 100% at RT were reported using various spinel-based barriers, viz. MgO/γ-Ga$_2$O$_3$ bilayer [522], MgGa$_2$O$_4$ [523], and Li-substituted MgAl$_2$O$_4$ [524]. The physical parameters of these materials, such as lattice constant and barrier height ($\alpha$ band gap, $E_g$), can be tuned by replacing the elements at the $A$ and $B$ sites of the spinel structure. In fact, RA of MgGa$_2$O$_4$ ($E_g$ = 4.9 eV) is much lower than that of MgAl$_2$O$_4$ ($E_g$ = 7.8
eV) with the same barrier thickness due to reduction in the effective barrier height [523]. Fig. 37 shows the relationship between the lattice spacing for a (0 0 1) plane and the experimental band gap of typical barriers (MgO and spinel oxides). Many of these spinels can form a stable solid solution over a wide compositional range, such that the barrier properties are continuously tunable on the basis of their composition. In the near future, MTJs are expected to be designed for specific applications by precise compositional tuning of the barrier materials.

3. Concluding remarks

Although polycrystalline CoFeB/MgO/CoFeB MTJs are dominantly used for practical MTJ devices, research on epitaxial MTJs is expected to gain increasing importance as they promise improved understanding of fundamental spin-dependent tunneling phenomena and accelerate the development of new MTJ materials. Because of the absence of grain boundaries, epitaxial and lattice-matched MTJs may overcome the characteristic device-to-device variation, even at small cell dimensions. Yakushiji et al. predicted the industrial feasibility of epitaxial MTJs grown on a Si wafer by 3-dimensional integration technology based on wafer bonding with CMOS circuits [525]. Such an epitaxial MR device grown on a Si wafer has recently been demonstrated [526]. Using such new technology, epitaxial MTJs that are difficult to grow in polycrystalline form can be implemented into practical MTJ devices. Progress in epitaxial barrier technology will be helpful to create innovative spintronic applications.

Acknowledgements

We thank M. Belmoubarik and J. Uzuhashi for their technical assistance for Fig. 36. We acknowledge support from the ImPACT Program of Council for Science, Technology and Innovation, Japan.

Magnetically ordered insulators for advanced spintronics
Matthias Althammera,b,*, Sebastian T.B. Goennenweinc,d, Rudolf Grossa,b,e
a Walther-Meissner-Institut, Bayerische Akademie der Wissenschaften, 85748 Garching, Germany
b Physik-Department, Technische Universität München, 85748 Garching, Germany
c Institut für Festkörperphysik, Technische Universität Dresden, 01062 Dresden, Germany
d Center for Transport and Devices of Emergent Materials, Technische Universität Dresden, 01062 Dresden, Germany
e Nanosystems Initiative Munich (NIM), 80799 München, Germany
*Corresponding author.
E-mail: matthias.althammer@wmi.badw.de (M. Althammer).

Abstract

Magnetically ordered, electrically insulating materials pave the way towards novel spintronic devices. In these materials the flow of magnetic excitations such as magnons results in pure spin currents. These spin currents can be driven by gradients of the spin chemical potential and/or temperature such that they can play the same role in novel spintronic devices as charge currents in tradi-otional electronic circuits. Connecting spin current based and charge current based devices requires spin to charge interconversion. This has been achieved by the spin Hall effect with an efficiency of several 10%. The recent progress in...
materials develop-ment and understanding of pure spin current physics opens up a plethora of novel device concepts and opportunities for funda-mental studies.

Keywords: Spin currents; Magnetically ordered insulators; Spintronic devices; Magnon diffusion; Ferromagnetic insulators

1. Introduction

Today’s information and communication technology is dominated by electronic and photonic devices. In sensing and magnetic storage applications, however, spin-based devices play an important role. Since many charge-based devices currently approach fundamental performance limits, it is tempting to exploit the spin degree of freedom to surpass these limits. A particular aspect is the reduction of power dis-sipation. However, since usually the spin (angular momentum) degree of freedom is carried by charge carriers, the flow of spin currents in electrical conductors is inevitably associated with the dissipative flow of charge currents. Aiming at devices with reduced dis-sipation, an obvious solution is to generate pure spin currents. This can be achieved in magnetically ordered electrical insulators (MOIs). There the charge degree of freedom is frozen out and pure spin currents are obtained by driving a flow spin excitation quanta (magnons) via a gradient of the spin chemical potential or of the temperature [529–539]. The magnons carrying the angular momentum quantum $h$ are the carriers of spin currents in these spin conductors, in the same way as electrons carrying the charge quantum $e$ are those of charge currents in charge con-ductors.

In the realm of spintronics, pure spin currents have developed into a new paradigm and are promising for a broad variety of novel device concepts [540–542]. Over the last decade extensive funda-mental research work has been dedicated to find means to efficiently generate and detect pure spin currents as well as to study their transport over length scales relevant for device concepts. In this con-text it is im-portant to mention that in contrast to charge to change in an ele-trical conductor, spin in an angular momentum conductor is not conserved. The angular momentum stored in the magnon system can leak into the phonon system. With an average magnon lifetime $\tau_m$, we can define a charac-teristic length $A_m = \sqrt{D_m \tau_m}$ for diffusive or $A_{m,ball} = v_m \tau_m$ for ballistic motion, where $D_m$ and $v_m$ are the diffusion constant and group velocity of the magnons, respec-tively. Evidently, spin current-based devices, where magnons are emitted at a source $S$ and collected at a drain $D$ should have a SD-separation much smaller than these charac-teristic length scales. The latter can be as high as several hundred nanometers in current materials (see Table 5). Moreover, for transistor-like three-terminal devices suitable implementations of a gate electrode $G$, allowing for a (dissipationless) control of spin currents are desirable. Here, we briefly review the present understanding of spin current physics, provide an outlook on the future potential of spin current-based devices, and address necessary future research directions with respect to MOIs.

2. State of the art

2.1. Foundations of pure spin current physics

Initially, the interest in MOIs was mainly driven by the spin caloritics community, facing the problem to clearly separate between contributions of charge and spin currents in spinthermogalvanic phe-nomena. The obvious solution was to freezeout the charge contribution by using MOIs. The potential of pure spin currents in MOIs for device applications was rapidly recognized and meanwhile they represent a cornerstone for modern spintronic device concepts. Since our present world is still dominated by electronics, i.e. the charge degree of freedom, there are hardly any commercial sources and detectors available for spin currents. Therefore, the research efforts in spin cur-rent physics and devices were focused on the following key tasks:

1. Generation of pure spin currents, including the efficient conver-sion of charge currents into spin currents.
2. Detection of pure spin currents, including the efficient conversion of spin into charge currents.
3. Understanding the decay of excess magnons towards thermal equi-librium.
4. Study of the diffusive and ballistic transport of pure spin currents.
5. Development of efficient tools for the control and manipulation of spin currents.

2.1.1. Generation of pure spin currents

Pure spin current generation schemes include spin pumping (driving non-equilibrium spin dynamics in MOIs by rf-fields [529–531,543–545] or surface acoustic waves [546,547]), driving spin currents by tem-perature gradients [532–534,548–551] or injecting spin currents into MOIs from a thin normal (not magnetically ordered) metal (NM) de-posited on the MOI [537–539,552,553]. The latter technique is based on the spin Hall effect (SHE) [554–557] and requires materials with strong spin-orbit coupling (large Berry phase). In this regard, the Berry phase in such materials gives rise to spin-dependent trans-verse velocities, which in turn give rise to the SHE [558]. A large variety of MOI/ NM heterostructures have been reported and large charge-spin con-version factors (spin Hall angles) exceeding 10% have been reported in literature for Pt, W, and Ta [556]. Moreover, a solid theoretical un-derstanding of the non-equilibrium spin transport across MOI/NM in-terfaces has been developed [534,550,559]. Unfortunately, the relevant materials parameters such as the spin Hall angle and spin diffusion length in the NM as well as the spin mixing conductance at the MOI/NM interface still show a significant spread in literature, most likely due to considerable variations in materials and interface properties. In Gen-eration of pure spin currents: particular, the transparency of MOI/NM interfaces for pure spin currents plays a crucial role and requires careful attention in experiments. Fortunately, MOI/NM heterostructures with an interface transparency similar to that in all metallic ferromagnet/ NM structures have been achieved, showing that a highly efficient spin current transport across the MOI/NM interface is possible [560]. Early spin pumping experiments mostly focused on the DC part of the spin transport across MOI/NM interfaces. However, more recently also the time-varying part attracted interest as it may pave the way towards high-speed spintronic devices up to the THz regime [544,561–564]. In a large number of experiments temperature gradients have been successfully used to drive pure spin currents in MOIs. In open circuits conditions a gradient of the spin chemical potential $V_\mu$ appears and – analogous to the charge Seebeck effect – this phenomenon is named spin Seebeck effect (SSE) [532–534,548–551]. Strictly speaking, the proportionality constant between $VT$ and $V_\mu$ should be denoted as spin

### Table 5

Magneton diffusion lengths $A_m$ measured in different experiments for different MOIs with different thicknesses $t_{face}$ of the MOI at $T = 300$ K. The different methods used are: spatially resolved inelastic light scattering (optical), in-ductive magnon generation and detection (inductive), lateral (lateral) and vertical (vertical) non-local electrical transport measurements.

<table>
<thead>
<tr>
<th>Material</th>
<th>$A_{m,diff}$ (nm)</th>
<th>$t_{face}$ (mm)</th>
<th>Method</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPE-YIG</td>
<td>31,000</td>
<td>100</td>
<td>optical</td>
<td>[590]</td>
</tr>
<tr>
<td>YIG, spurt.</td>
<td>10,060</td>
<td>40</td>
<td>optical</td>
<td>[591]</td>
</tr>
<tr>
<td>YIG, PLD</td>
<td>25,000</td>
<td>20</td>
<td>optical</td>
<td>[592]</td>
</tr>
<tr>
<td>LPE-YIG</td>
<td>860,000</td>
<td>200</td>
<td>inductive</td>
<td>[593]</td>
</tr>
<tr>
<td>LPE-YIG</td>
<td>9,400</td>
<td>100</td>
<td>lateral</td>
<td>[535]</td>
</tr>
<tr>
<td>LPE-YIG</td>
<td>700</td>
<td>3,000</td>
<td>lateral</td>
<td>[536]</td>
</tr>
<tr>
<td>YIG, spurt.</td>
<td>38</td>
<td>40 to 100</td>
<td>vertical</td>
<td>[588]</td>
</tr>
<tr>
<td>NiFe$_2$O$_4$, spurt.</td>
<td>3100</td>
<td>44</td>
<td>lateral</td>
<td>[576]</td>
</tr>
</tbody>
</table>
Seebbeck coefficient. However, since $V_{th}$ is difficult to measure, the spin current was converted into a charge current (leading to a voltage in open circuit conditions) using MOI/NM heterostructures (see above) and the Seebbeck coefficient has been introduced as the proportionality constant between $VT$ and the detected voltage. An important contribution of SSE experiments was to provide a deeper understanding of magnon excitations in MOIs that were previously only attainable by non-electrical transport experiments [565]. Regarding applications, the SSE may allow us to use MOI/NM heterostructures for waste energy reuse in the same way as thermoelectric devices [566]. The relevant figures of merit are presently under study [540]. While ini-tial work on the spin Seebbeck effect mostly employed ferro- and fer-rimagnetic ordered MOIs, within the last years the spin Seebbeck effect was also observed in MOIs with antiferromagnetic order $(\text{Cr}_2\text{O}_3, \text{MnF}_2)$ [567–570]. In these experiments large external magnetic fields are required to manipulate the magnetic moments of the antiferromagnet, however they proved that a pure spin current can also be generated by a thermal drive in a MOI with antiferromag-netic order.

2.1.2. Detection of pure spin currents

Spin currents can be elegantly detected via the inverse spin Hall effect (ISHE), which is the Onsager reciprocal of the SHE. This detection scheme is based on MOI/NM heterostructures where spin currents in the MOI are transferred into a thin NM layer and converted into a charge current by the ISHE. Efficient conversion requires highly spin transparent MOI/NM interfaces and NM layer thicknesses on the lengthscale of the electron spin diffusion length. Typical NM materials are $\text{Pt}$, $\text{W}$, and $\text{Ta}$, featuring spin Hall angles of 10% or more [553,556,560]. Novel two-dimensional systems approach values of 40% [571].

2.1.3. Decay of excess magnons

We already pointed out in the introduction that the relevant magnon transport length-scale crucially depends on the magnon lifetime $\tau_m$. After switching off an external perturbation, any excess population of a particular magnon with energy $\hbar \omega_m$ and wave number $k_m$ will decay towards the thermal equilibrium population by magnon-magnon and magnon-phonon interaction. The characteristic decay time $\tau_m$ will depend on both energy and wave number. The linewidth in ferromagnetic resonance experiments yields access to the uniform $k=0$ mode of spin excitations. For MOIs long magnon lifetimes (e.g. $\tau_m = 1 \mu s$ for yttrium iron garnet (YIG)) exceeding those of metallic ferromagnets by several orders of magnitude have been reported, as the usually dominant magnon-electron scattering is suppressed in MOIs [572]. The decay of finite wave number magnons can be studied by optical techniques which, however, are mostly restricted to small wave numbers [573,574]. These inelastic light scattering experiments rely on the inductive excitation of magnons via microwave antenna and spatially resolved experiments allow to investigate magnon transport in MOIs. A detailed under-standing of the magnon relaxation as a function of energy and wave number is still missing and requires further research efforts in both theory and experiment.

2.1.4. Diffusive and ballistic transport of pure spin currents

The realization of both charge-to-spin and spin-to-charge current conversion not only allowed for the study of long-distance magnon transport in MOI/NM heterostructures, but also laid the foundations for interfacing charge and spin based devices and realizing magnon based logic [535,536,575–580]. To study magnon transport in MOIs, two NM strips – one acting as the magnon source (S) and the other as the magnon drain (D) – are placed with separation $d$ on one and the same MOI as shown in Fig. 38 [581–583]. The magnons generated via the SHE in S diffusively or ballistically propagate to D, where they are converted into a voltage signal via the ISHE. The size of the voltage signal crucially depends on temperature, applied magnetic field, the SD-
including the thickness of the MOI [584], fabrication process and data analysis. When comparing and interpreting the measured magnon diffusion lengths, one has to keep in mind that several magnon branches and magnons with different wave number are contributing [585]. This is related to the fact that the magnons are generated by electron-magnon scattering at the MOI/NM inter-face, resulting in a broad excitation spectrum [559,581,582]. Hence, the measured $\lambda_{m,diff}$ represents an average value obtained for specific experimental conditions. In contrast, experiments using inductive excitation mechanisms can selectively excite certain magnon wavelengths yielding rather different propagation lengths as to the nonlocal experiments (see Table 5). Another indirect measurement method may be provided by the thermal effects associated to the magnetic excitation transport [586,587], but requires further experimental studies. Reducing the temperature to $T \leq 30$ K was found to result in a vanishingly small detector voltage. This has been interpreted as evidence for a thermally activated nature of the magnon generation process at the MOI/NM interface [536]. Meanwhile, the initial experiments have been confirmed by other groups and led to a deeper understanding of the underlying physics [535,536,575-580,588]. Interestingly, amorphous YIG thin films also exhibit long-distance magnon-transport in non-local experiments, which may allow to even use disordered spin systems as materials for magnon devices [589].

A further crucial parameter is the so-called spin convertnance describing the efficiency of the spin transfer between the electron and magnon system at the MOI/NM interface [581,582]. Very recently, means to enhance the spin convertnance have been proposed [579]. In summary, one can say that the measured magnon diffusion lengths in MOIs are promising, but more systematic experiments are required to gain a full quantitative understanding.

2.1.5. Tools for the control and manipulation of spin currents

Three-terminal, transistor-like devices are highly important for applications. Therefore, regarding the implementation of device concepts not only the generation and detection of spin currents but also the control and manipulation of spin current flow by suitable gate electrodes is required. Research in this direction is only rudimentary and requires more attention. Current approaches are mostly cener-tered on employing reconfigurable magnonic crystals [541,542]. Another pathway towards spin current control in MOIs is to employ nonlinear magnon-magnon interactions, which can be used for example to suppress a single wavevector magnon current [542]. However, this approach requires large local magnon densities, which may be achieved by making use of the magnon dispersion in magnonic crystals. Ideally, similar to charge based transistor concepts an electrical fieldcontrol magnon conduction channel would allow for very efficient magnon logic devices.

2.2. Novel magnetoresistive effects

The interplay between the electron spin accumulation generated in a NM with strong spin-orbit interaction via the SHE and the nonequilibrium magnon distribution in a MOI at an MOI/NM interface led to novel magnetoresistance phenomena, namely the spin Hall magnetoresistance (SMR) and the magnon mediated magnetoresistance (MMR), which we discuss in more detail in the following.

2.2.1. Spin Hall Magnetoresistance (SMR)

The combination of SHE and ISHE for pure spin current generation and detection in MOI/NM heterostructures led to the discovery of the spin Hall magnetoresistance (SMR) [537-539,552]. The SMR is characterized by the dependence of the NM resistance on the angle between charge current direction in the NM and the (sublattice) magnetization orientation in the MOI. In the simplest case of a collinear ferromagnet, an additional contribution to the resistance in the NM arises if the spin polarization $s$ of the spin current generated in the NM via the SHE and transferred across the MOI/NM interface is perpendicular to the magnetization $M$ in the MOI. For $s|M|$, the dissipation channel resulting from spin transfer torque and, hence, the additional NM resistance is absent. The magnitude of the SMR effect scales with the square of the spin Hall angle of the NM layer, i.e. the interconversion factor between charge and spin current. Experimental values for the relative resistance change of up to $10^3$ have been reported [553,560]. Most importantly, from these values, which can be obtained from simple magnetotransport measurements, relevant spin transport parameters can be inferred. Furthermore, SMR proved to be a powerful tool also for the investigation of non-collinear magnetic phases (e.g. helical, spin-canting and spin-flip ordering) in MOIs [594,595]. Especially MOIs with anti-ferromagnetic order have been investigated using the SMR [596-601], which further provides evidence that the SMR is sensitive to the magnetic order parameter itself (or possibly the magnetic sublattice moment orientations), but not the net magnetic moment direction. Taken things even a step further, spin currents in NM have meanwhile been used to switch the magnetization of a MOI by the spin transfer torque [602].

2.2.2. Magnon Mediated Magnetoresistance (MMR)

As discussed above, in a device structure as shown in Fig. 38 an electrical current in the source S causes a non-local voltage drop in the spatially separated drain D. Since this phenomenon is based on magnon diffusion between S and D, it is called magnon mediated magnetoresistance (MMR). While the theoretical prediction of MMR was pioneered by Zhang and Zhang [581,582] in 2012, the first experimental observation was reported in 2015 [535,536]. Although a detailed knowledge of MMR has not yet been achieved, it is evident already today that MMR has a broad potential for magnon based spintronic devices. Due to a long magnon diffusion length in MOIs lateral device schemes as illustrated in Fig. 38 are feasible with low demands on lateral dimensions. Making use of MMR, electrically controlled magnon logic circuits [578] have already been demonstrated and, very recently, evidence for the magnon transport analogue of the anisotropic magneto resistance in MOIs has been reported [577].

3. Perspectives for magnetic insulator based spintronics

Even today, a rich variety of physical phenomena has been studied in MOIs and successfully used for spintronics applications. Among them are the SMR and MMR, which are powerful tools for probing spin transport across MOI/NM interfaces and magnon propagation in MOIs. MMR already has been successfully used to implement charge-to-spin current converters and magnon-based logic devices [578]. In the following we address some important tasks and key challenges for future research in MOI based spintronics.

3.1. Magnon-based transistors

An important task will be the realization of transistor-like structures relying on magnon flow in a MOI controlled by a gate electrode. Here, the key challenge is to develop suitable gate concepts. Controlling magnon flow by an electric gate voltage would be particularly desirable in the context of electronic-magnonic hybrid circuits. Here, multiferroic materials with long magnon diffusion lengths are an option. Research in this direction is expected to also stimulate the realization of novel magnon logic devices.

3.2. Selective magnon excitation

Presently, there is little control on magnon excitation in devices relying on the SHE in MOI/NM heterostructures. The development of
concepts allowing to selectively excite magnons of specific energy and wave number are a prerequisite for coherent magnon logic circuits using a DC excitation schemes. Possible solutions for selective magnon excitation could be based on magnonic crystals, i.e. on appropriately engineering the magnonic band structure of the MOI [542].

3.3. Nonlinear phenomena

Another important research task will be the study of nonlinear effects in MOI based devices. A particularly interesting example is the nonlinear dependence of the magnon population in an MOI on the charge current flowing in a NM in devices based on MOI/NM heterostructures. First experiments show that the excitation of magnons with low damping is possible after overcoming a threshold value [605]. Similarly, the generation of magnonic superfluids using spin currents generated via the SHE has been proposed [604]. While Bose-Einstein condensation in MOIs has already been experimentally realized using parametric microwave pumping [605,606], an experimental realization of a magnonic BEC by all-electrical driving has not yet been reported. We note that a magnonic superfluid would coexist with normal magnonic excitations, leading to shortcircuiting effects of the magnon transport in MOIs. This allows for novel control schemes for magnonic transport.

3.4. Antiferromagnetic magnonics

Current experimental studies mostly focus on magnon transport in ferro- or ferrimagnetically ordered MOIs, where typical magnon frequencies are of the order of a few GHz. However, these studies and already developed device concepts can be straightforwardly transferred to antiferromagnetically ordered MOIs as already demonstrated for the spin Seebeck effect [567–570] and for the spin Hall effect [563,564]. It is expected that this leads to promising applications with an extended frequency regime. A promising example for future applications is the recent demonstration of an antiferromagnetic memory based on the manipulation by the anomalous Hall effect in Pt electrode [607]. Clearly, further studies to better understand retainability, stability, repeatability and switching speed are required for successful implementation into new device concepts.

3.5. Ballistic magnon transport

Until today, magnon based devices mostly rely on diffusive magnon transport. An important future task will be the detailed study of ballistic magnon transport. A particularly interesting aspect is the study of ballistic magnon transport in one-dimensional transport channels and the demonstration of the quantized magnon conductance in such structures. The experimental challenge is to fabricate one-dimensional magnon transport channels and to reduce the length of these channels below the magnon mean free path. To this end, a possible strategy is to use domain walls in MOIs. We note that the realization of ballistic magnon transport also is promising regarding novel device concepts.

Taken together, within the last decade the successful research on spin current transport in MOIs has led to a manifold of novel phenomena. This allowed us to get a deeper insight into the fundamental physics and paved the way towards promising applications. Over the next years, further exciting results on fundamental physics aspects and rapid progress regarding transfer into applications are expected.

1. Introduction: device and material issues

Since the introduction a decade ago of multicore design in chip manufacture, the speed of processing units has increased steadily fast. However, the existing on-chip copper interconnects – that link and transmit data between the processing cores – imposes a limitation for scalability, as the signal integrity is degraded due to the impedance of copper lines. Consequently, alternatives to conventional on-chip electrical wiring have been keenly explored over the few past years. Among them, silicon photonics is considered to be a promising solution [608], offering the ability to shift data around with improved speed, energy efficiency and immunity against electromagnetic interferences. Silicon photonics is also foreseen to drive the transition from current optical interconnects, where optics is simply used for data transmission, to a new generation of systems, where switching and routing operations are performed by optoelectronic devices [609]. Nevertheless, the plasma dispersion effect used in silicon photonics suffers from drawbacks in terms of speed, energy efficiency or optical losses due to the limitations that silicon show for the electric switching of their optical properties – e.g., the refractive index – [610]. In this context, as described below, functional metal oxides stand out for their ability to have unique optical properties and, consequently have raised a large interest for their integration into Si-compatible photonic circuitry.

As aforementioned, next generations of optical interconnects will require optical modulation and routing capabilities with high performance and, therefore, modulators and switches are key components for next-generation photonic integrated circuits. In view of their unique optical properties, the integration of functional oxides into Si-photonic chips is considered to be a strategic objective in the field. Among metal oxides, some materials stand out for their particularly remarkable properties. On the one hand, ferroelectric oxides, such as LnNbO3 and BaTiO3, display outstandingly large electro-optical properties, which can be exploited to modulate optical signals with ultra-high speeds and low losses. Alternatively, the metal-insulator transition in oxides, such as VO2, shows large electrically-induced changes of the refractive index in the range of telecommunication wavelengths, offering a promising route towards energy efficient photonic switches. Finally, novel concepts for photonic devices stem from the possibility of achieving epsilon near-zero regimes for optical communications in transparent conductive oxides. We provide here a critical survey on the current status of the integration of functional thin film oxides into Si and the progress in their optical functionalities with respect to the standards of Si photonics. Future prospects in emerging fields, such as optical networks for neuromorphic computing are also considered.

Keywords: Silicon; Photonics; Functional oxides; Electro-optics; Switching

Functionals oxides in photonic integrated devices

Gervasi Herranz a, Pablo Sanchis b

aInstitut de Ciència de Materials de Barcelona, Campus de la UAB, 08193 Bellaterra, Catalonia, Spain
bNanophotonics Technology Center, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

Corresponding author.
E-mail: gherranz@icmab.cat (G. Herranz).

Abstract

On-chip integrated photonic/electronic circuits are seen as essential for the future development of information and communication technologies. One key aspect is to add optical modulation and routing capabilities to enable optical interconnects with enhanced performance. For that purpose, modulators and switches are key components for next-generation photonic integrated circuits. In view of their unique optical properties, the integration of functional oxides into Si-photonic chips is considered to be a strategic objective in the field. Among metal oxides, some materials stand out for their particularly remarkable properties. On the one hand, ferroelectric oxides, such as LnNbO3 and BaTiO3, display outstandingly large electro-optical properties, which can be exploited to modulate optical signals with ultra-high speeds and low losses. Alternatively, the metal-insulator transition in oxides, such as VO2, shows large electrically-induced changes of the refractive index in the range of telecommunication wavelengths, offering a promising route towards energy efficient photonic switches. Finally, novel concepts for photonic devices stem from the possibility of achieving epsilon near-zero regimes for optical communications in transparent conductive oxides. We provide here a critical survey on the current status of the integration of functional thin film oxides into Si and the progress in their optical functionalities with respect to the standards of Si photonics. Future prospects in emerging fields, such as optical networks for neuromorphic computing are also considered.
mostly investigated in device architectures based either on Mach-Zehnder interferometers (MZI) or microring-resonator. Briefly, the MZI structure, shown in Fig. 39(a), is based on inducing relative phase shifts between two optical paths derived from a single source. Alternatively, in microring-resonator structures straight waveguides are placed between circularly shaped waveguides, as depicted in Fig. 39(b). At resonance, light is passed through the loop from the input waveguide, so that its intensity rises steeply due to constructive interference, which is detected at the output waveguide. The advantage of ring resonators is that they minimize the footprint, but this reduction comes at the expense of optical bandwidth that can only be increased by using more complex structures or switching techniques. Conversely, MZI structures have a wider optical bandwidth, but commonly have larger footprint and consume more electrical power, which limits the scalability.

It is also important to distinguish between modulation and switching functionalities. The electro-optic modulator is a device that changes the properties of an optical wave, usually the amplitude or phase, by means of an electrical signal. The most relevant parameters are the electro-optical bandwidth, modulation efficiency and insertion losses. Electro-optical bandwidths above 40 GHz are required for datacom applications that eventually could enable data bit rates up to 100 Gb/s depending on the modulation format. The modulation efficiency is normally quantified in terms of the V.L product, which relates the length and driving voltage required to achieve a \( \pi \)-phase shift. Commercial modulators have typically values around 4 to 8 V cm while insertion losses are around 5 dB. However, low driving voltages below 3 V are mandatory for allowing the co-integration with CMOS electronics. Lower insertion losses are also needed for efficiently integrating many optical functionalities into a single chip. In that context, ferroelectric oxides have grabbed the attention of scientists and engineers because of the large linear electro-optical coefficients intrinsic to their non-centrosymmetric crystal structures and their feasibility to be integrated with silicon photonic devices [611].

Optical switches are used to route an optical signal from an input port to an output port. The optical bandwidth of the switch determines the maximum data rate of the optical signal that can be routed to adjacent output ports with minimum insertion losses and crosstalk. Optical routing of signals with rates above 100 Gb/s are expected to meet the future needs of data centers. On the other hand, the switching speed is related with the time required to change the state of the switch by applying an electrical excitation or, in other words, the routing of the high-speed optical signal from one output port to another. The requirement in terms of switching time depends on the application. However, fast responses in the order of the nanosecond are pursued to reduce latency in data communications. An additional relevant property is the power of the electrical signal used to change the switching state, i.e. the power consumption of the optical switch, which will determine, together with the footprint, the capacity to scale the number of ports. Power consumptions in the milliwatts range or below are expected for enabling large switching matrices. In that case, strong electronic correlations in metal oxides and transparent conducting oxides are currently promising approaches for electro-optical switching applications.

All the aforementioned prospects for novel integrated photonic devices rely on the successful monolithic integration of functional oxide nanostructures on top of silicon. Since the physical properties of oxides depend critically on the structural quality, one of the most challenging aspects is the achievement of high-quality epitaxial growth. The reason behind this difficulty is both the structural dissimilarity between oxides and silicon and the formation of SiO\(_2\) or silicate layers during first stages of growth, which can prevent good epitaxy [612,613]. Fortunately, solutions to these problems have been provided by breakthroughs on thin film growth, exemplified by the achievement of high-quality SrTiO\(_3\) or yttria-stabilized zirconia (YSZ) buffer layers that have enabled the successful integration of functional oxides on silicon [613]. Undoubtedly, the integration of functional oxides into Si-photonicics has greatly profited from this progress.

Another important aspect is related to the methods for the thin film deposition of materials. In this respect, there have been successful advances using either physical methods -e.g., pulsed laser depositions, sputtering or molecular beam epitaxy- or chemical routes, such as chemical vapor or atomic layer deposition. Among the latter, chemical solution deposition methods offer promising routes towards low-temperature processing methods that are compatible with the integration into standard microelectronic processes in CMOS technology and have potential for large scalability [613].

In the following, an overview is given on the current status and future challenges of functional oxides for silicon photonic applications.

2. Current and future challenges

Ferroelectric oxides. Today commercial electro-optical modulator devices used in fiber-optics communications are mainly based on the Pockels effect i.e., the change of the refractive index induced by an electric field, so that the optical phase is modulated according to the applied electric signal. Having access to such electro-optic effect is highly desirable due to the inherent ultra-fast speed, optical lossless performance and high linearity for enabling multilevel modulation formats. Unfortunately, the Pockels effect is not present in silicon due to its centrosymmetric diamond crystal structure. Therefore, alternative approaches are currently being investigated such as applying stress to the silicon waveguide to break the crystal symmetry or the integration of active materials into the silicon platform. In the latter, ferroelectric oxides are excellent candidates due to their high Pockels coefficients. LiNbO\(_3\) is the first evident option as the technology is mature and commercially available and, therefore, its material properties are very well controlled. Modulation bandwidths up to 50 GHz with modulation efficiencies below 6.5 V.cm have been demonstrated in hybrid LiNbO\(_3\)/Si photonic structures [614]. However, LiNbO\(_3\) cannot be grown on top of silicon and bonding techniques are required for enabling the integration with silicon photonics [615], which remains as the main future challenge. These considerations suppose a serious roadblock for the development of integrated electro-optic modulators based on LiNbO\(_3\).

More recently, barium titanate (BaTiO\(_3\)) has appeared as a
promising alternative to LiNbO3 [611]. Indeed, the Pockels coefficient in bulk BaTiO3 crystals (r_{BTO} > 1000 pm·V^{-1}) is several orders of magnitude higher than LiNbO3 (r_{LiNbO3} ≈ 30 pm·V^{-1}) [616]. These interesting properties have stimulated the analysis of BaTiO3-based photonic devices. Along this line, Mach-Zehnder modulators with Si-compatible BaTiO3 waveguides, shown in Fig. 40(a), have predicted values of V_rL as low as 0.27 V·cm for c-axis oriented BaTiO3 and TE polarization by rotating the optical waveguide to an optimum angle [617]. The reality, however, is that the experimental values of the Pockels coefficients measured in thin films are much smaller than those reported for bulk crystals [618]. Electro-optical modulation based on a hybrid BaTiO3/Si waveguide structure was experimentally reported with an estimated effective Pockels coefficient up to 213 pm·V^{-1}, voltage length product of V_rL = 1.5 V·cm, and modulation bandwidths in the gigahertz regime but large propagation losses of 44 dB/cm [619], much larger than the propagation losses below 2 dB/cm of state-of-the-art silicon waveguides. We see, therefore, that the measured voltage length product V_rL is within the range of commercial modulators, but the appearance of high propagation losses is a serious drawback that needs to be tackled. Fortunately, recent studies show that optical losses depend strongly on the material processing and appropriate CMOS-compatible procedures can be developed to obtain low-loss waveguide structures with absorption losses around 6 dB/cm, which are much closer to the standards of silicon photonics (~2 dB/cm) [145].

On the other hand, as aforementioned, there is still a large room to improve the optical properties of Si-integrated thin films and, in particular, the electro-optic coefficients. In this respect, a key aspect is to have an accurate and reproducible control of the ferroelectric domain orientation during the film growth, since it has been demonstrated that domain orientation has a strong impact on the modulation performance [617,620]. It has been revealed that the domain orientation is influenced by competing compressive stress from epitaxy and tensile stress from thermal expansion [612] or the variation of oxygen pressure during growth [621]. Additionally, a comprehensive recent study has shown the strong dependence of the Pockels effect in BaTiO3 thin films on the microstructure [618]. More specifically, the main aspects that have been identified as essential to maximize the electro-optic response are related to the reduction of the film porosity, which reduces the effective electric field inside the ferroelectric layer and the increase of tetragonality c/a, where c and a are the lattice parameters, which is generally used as an indicator of symmetry breaking. A reduction of tetragonality may be caused by defects that restore locally centrosymmetry or to the formation of multidomains with random orientations. Therefore, achieving dense, tetragonal, epitaxial films remains the big challenge for the next future to enhance the electro-optic properties towards bulk values. Very recently, a high Pockels coefficient of 923 pm·V^{-1} and modulation speeds up to 50 Gbit/s have been demonstrated in two complementary photonic and plasmonic based modulating devices. The obtained results confirm the promising potential of BaTiO3 for enabling novel photonic components at both micro and nanoscales with superior electro-optical performance [622].

**Metal-insulator transitions in oxides.** Because of the metal-to-insulator transition, vanadium dioxide (VO2) has raised recently the interest for plasmonics, metamaterials, and reconfigurable photonics. The most striking feature of this compound is that above the transition temperature it behaves as a metal, but as soon as the temperature is reduced below the critical value the resistance changes abruptly by several orders of magnitude. Consequently, this sharp transition entails a large modulation of the complex refractive index, particularly at telecommunication wavelengths [623], which allows the design of ultra-compact devices. Hybrid VO2/Si photonic waveguide structures have been successfully demonstrated mainly with one input/one output (1 × 1) configuration and exploiting the change in the imaginary part of the refractive index [624]. Electro-absorption modulation with large extinction ratios above 10 dB have been achieved with small active lengths of only 1 µm [625]. Such an ultra-compact modulation device allows for scaling up the integration of modulators arrays and reduces the size mismatch between photonics and electronics. The main challenge is the ultimate modulation speed that can be achieved. The pure electronic process in the VO2 phase transition is inherently very fast and switching times lower than 2 ns have been experimentally measured in VO2/Si devices [626]. Nevertheless, a Joule based thermal process has been shown to be necessary for completing the transition, which is required to maximize the modulation depth but would prevent its use for high-speed modulation. Therefore, strategies to minimize the impact of the slow thermal component are mandatory for enabling practical modulating devices.

As mentioned in the introduction, a desirable target is to develop optical switches with low power consumption, low insertion losses and compact footprint as key parameters for scalability. Very recently, efforts in this direction have been investigated and 2 × 2 photonic switches have been designed using microring resonators based on the hybrid VO2/Si waveguides depicted in Fig. 40(b), with footprint below 50 µm² and a high optical bandwidth that could support a data throughput above 500 Gbit/s [627]. However, insertion losses are too high (> 1 dB) for enabling large-scale switching matrices. The reduction of the imaginary part of the VO2 refractive index at the insulating state or novel disruptive designs of the VO2/Si photonic waveguide structure are potential approaches for minimizing insertion losses. On the other hand, the impact of the external circuitry on the power consumption when electrically switching the VO2 has been proven critical and power consumption reductions up to 90% have been demonstrated reaching values around 5 mW [628].

The integration of VO2 films on hybrid silicon waveguides faces also material challenges. The most essential issue is related to the structural mismatches of VO2 and Si, which can degrade the optical performance. Moreover, due to the multiplicity of valence states of vanadium, there are several stable oxides of different chemical composition, including VO2 polymorphs, V2O5 and V2O3, among many other phases. This has stimulated the research of the optimum growth conditions for high
optical performance of VO2 films on silicon platforms, with promising results [629]. These encouraging perspectives may enable future devices based on VO2 competitive with actual silicon photonic switches. Furthermore, the potential of hybrid VO2/Si devices to operate as photodetectors with high sensitivity at telecom wavelengths has also been shown with responsivity in excess of 10 A/W [625].

Transparent conducting oxides. Several transparent conducting oxides, and in particular indium tin oxide (ITO), have gained an increased interest during the last years for pushing forward the limits of electro-optical functionalities. As with VO2, the main outstanding property is the ability to achieve an ultra-large change of the complex refractive index (Δn > 1) but, in this case, by changing the carrier concentration under the application of an electric field [630]. Furthermore, an epsilon-near-zero (ENZ) regime at telecommunication wavelengths can be achieved and exploited to significantly enhance the variation of the optical mode properties. In the ENZ condition (i.e., permittivity ε ≈ 0), the material undergoes a transition between a dielectric (low absorption) and a metallic response (high absorption). The electric field is largely enhanced with respect to the adjacent materials which improves the modulation efficiency when a high optical confinement across the ENZ material is also achieved [631].

Along this line, most of the proposed approaches for electrooptical modulation are based on the control of plasmonic modes through a charge-accumulation layer in metal-oxide-semiconductor (MOS) structures. High modulation depths (2.71 dB/µm) have been demonstrated [632]. However, a trade-off between modulation depth and insertion losses is still present due to the high losses inherent to plasmonic mode propagation. Therefore, pure dielectric approaches based on replacing metal layers by semiconductor dielectric materials, basically doped silica, have also been designed with the aim of minimizing losses [633]. One of the main challenges is still to demonstrate high-speed modulation as an ultra-fast time dynamics linked to the carrier concentration effect is expected. Conversely, a non-expected resistive switching performance has been reported that would not be useful for implementing modulating devices but could open a new way for enabling optical switching with a bistable performance [634]. Additional transparent conducting oxides, such as those based on doped zinc oxide, are also being considered as an alternative to ITO.

3. Concluding remarks

In the previous section, we have described the current status and future challenges on the research of some functional oxides and their integration into Si-photonic devices to achieve optical modulation, switching and even photodetection capabilities with unique performance. These devices could enable a transition to a new generation of optical interconnects for datacom applications but they could also expand applications in other fields, such as sensing. Furthermore, the demonstration of successful active photonic devices opens up fascinating novel avenues into other emergent fields, such as non-volatile optical memories, quantum photonics or neuromorphic computing.

For instance, the above-mentioned Mach-Zehnder interferometers (Fig. 39(a)) work by controlling the interference of beams travelling along waveguides, so that the relative amplitude and phase of the interfering beams in the output waveguides can be modulated. This process enables the possibility of controlling and programming the transmission response through a mesh of interconnected arrays of MZI nodes. This complex structure should allow developing a remarkably wide range of optical functionalities. Among the most interesting prospects, arrays of interconnected MZI modules can enable linear optical operations that can mimic neuromorphic operations using optical networks. An impressive step towards this paradigm has been realized recently in silicon integrated photonic meshes comprising hundreds of optical components in millimeter-sized chips, demonstrating key aspects of an optical neural network processing [635].

At this point is where some functional oxides may have a decisive role for further development of Si-integrated optical neuromorphic computing. In particular, it is known that brain-like computing requires highly connected networks of nonlinear elements that are “trained” to perform operations. For that purpose, integrating materials with non-volatile physical properties is an important advantage. In this line, recent research has shown the potential to reconfigure the strength of interconnected networks of oxide-based ferroelectric tunnel junctions [327]. Similarly, other oxides with non-volatile functionalities – as in ferroelectrics or resistive switching materials – may provide a playground to implement physically and modulate the connection strengths in a network. With this in mind, and given the impressive progress in the integration of electro-optic oxides into silicon platforms, the possibility of further development into optical neural networks is particularly enticing.

Acknowledgements


Recent concepts and future opportunities for oxides in solar cells

Anders Hagfeldt*, Monica Lira-Cantu**

*Laboratory of Photomolecular Science, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
**Catalan Institute of Nanoscience and Nanotechnology (ICN2), CSIC and The Barcelona Institute of Science and Technology (BIST), Campus UAB, Bellaterra, E-08193 Barcelona, Spain

*Corresponding author.
E-mail: monica.lira@icn2.cat (M. Lira-Cantu).

Abstract

We are moving towards a sustainable society powered by renewable energy where solar photovoltaics is one of the most important players. In the past few years, emerging photovoltaic (PV) technologies have observed an exponential increase in power conversion efficiencies (PCE) with halide perovskite solar cells above 23%, tandem photovoltaics reaching 27% or dye sensitized solar cells for indoor lighting at the impressive 28.9% PCE mark. These technologies could potentially complement the commercial silicon solar cells but they must also reduce the number of solar cells in the manufacturing process and be more environmentally friendly. Novel materials such as perovskites and organic photovoltaics (OPV) have shown potential for replacing silicon-based solar cells due to their low costs and long-lifetimes. However, the steady progress of Si-integrated photonic devices based on oxides indicates that traditional photovoltaic materials are not enough to meet the current demand for solar energy. New materials and technologies such as organic photovoltaics (OPV) and dye sensitized (DSSC) cells can be used to improve the efficiency of solar cells and reduce the environmental impact. In addition, new materials and technologies such as perovskites and organic photovoltaics (OPV) can be used to improve the efficiency of solar cells and reduce the environmental impact. In conclusion, the development of new materials and technologies is crucial for the continued growth of the solar energy industry.
1st generation PVs refers to Silicon solar cells, 2nd generation refers to thin film inorganic PVs like CIGSe, CdTe, amorphous and microcrystalline Silicon solar cells.

Keywords: Semiconductor oxides; Photovoltaics; Perovskite solar cells; Stability; Barrier layers

1. Introduction

The application of photovoltaic (PV) technologies for the conversion of solar energy into electricity can contribute to the increasing global energy demand and the international goals of renewable energy utilization. Power conversion efficiency (PCE) is a critical parameter for PV cost reduction: higher PCE will always translate into less expensive PV system. Third generation PV technologies refer to organic (OPV), dye sensitized (DSSC) and halide perovskite (PSC), tandem and multi-junction solar cells (TSC) and all-oxide solar cells (AOSC). Among them, PSCs have recently demonstrated an impressive increase in efficiency, currently at 23.3%, with the development of new halide perovskite light harvester materials. Semiconductor oxides constitute a fundamental part of the technology providing powerful charge transport layers, light absorbers, conducting metal electrodes, nanostructures for light management and also recombination layers in multijunction and tandem solar cell technologies (Fig. 41). Oxides have demonstrated to have versatile hosts structures which allow the customization of their electronic, ferroelectric, or optical properties with the feasibility to be fabricated by low-cost and scalable fabrication methods. Their outstanding ability to preserve or improve solar cell characteristics permits their application in low-weight, flexible and semi-transparent PV devices and printed electronics. Doped and undoped binary semiconductor oxides provide enhanced stability to solar cells which is a significant step towards the commercialization of the technology. Moreover, complex semiconductor oxides are currently emerging as materials with novel properties that can influence solar cell performance. Among these properties are magnetism, ferroelectricity or piroelectricity (among others), that collectively with the most classical binary materials, can deliver novel and innovative features.

Rather than a review of the recent literature, the aim of this perspective document is to provide a brief critical analysis of the state of the art and future prospects of the application of oxides in emerging photovoltaics. For more detailed information on these oxides, the reader can consider recent published reviews[636–642]. In this work, very specific information is given including graphs with solar cell efficiency values. Among the many different functionalities that oxides can provide to solar cells, in this work, we specifically focus on the application of oxides in solar cells. Transition metal oxides are standard materials applied nowadays in many thin-film optoelectronic devices and are compatible with the manufacturing techniques for large-scale, large-volume, flexible and disposable/reusable devices. These stringent requirements are more relevant for emerging photovoltaics (e.g. OPV, DSSC, PSC) where large-scale, large-volume techniques require low fabrication costs. Besides the many advantages of oxides, there are a few important properties which make oxides very attractive for their application (and future commercialization) of solar cells: their low temperature synthesis methods, their low cost (in comparison with the most applied organic semiconductors) and the stability towards harsh environment (e.g. humidity) which permits to enhance solar cell lifetime.

2. State of the art

Transition metal oxides are standard materials applied nowadays in many thin-film optoelectronic devices and are compatible with the manufacture techniques for large-scale, large-volume, flexible and disposable/reusable devices. These stringent requirements are more relevant for emerging photovoltaics (e.g. OPV, DSSC, PSC) where large-scale, large-volume techniques require low fabrication costs. Besides the many advantages of oxides, there are a few important properties which...
2.2. Oxides as intermediate layers in tandem solar cells

Multijunction devices, such as tandem or triple cells, have attracted great attention due to the potential to maximize the use of the solar spectrum, while reducing energy loss. Their importance resides on the efficiencies beyond the single-junction Schockley-Queisser limit which maximum achievable theoretical efficiency exceeds that of single-junction devices. In this configuration, two or more solar cells with different active layers are stacked and connected in series. In tandem solar cells, metal oxides can be find as thin films working as interlayers, in some cases acting as the recombination layer between adjacent sub-cells (facilitating efficient recombination of the charge carriers) or as window layers (as highly conductive transparent oxide films, with minimal parasitic adsorption protecting the absorber layers from damage during fabrication or for device stability) [647,659].

Fig. 3 shows a schematic representation of a tandem solar cell (Fig. 43(b)) and examples of oxide layers acting as intermediate layers in different tandem solar cells (Fig. 43(a)). One of the most promising tandem solar cells applying emerging photovoltaics is the silicon/PSC device with a record efficiency of 27.3%. The solar cell applies MoOx oxide as the interlayer. Other tandem configurations include sub-cells made of copper indium gallium (di) selenide (CIGS/PSC), two OPVs (OPV/OPV) or Quantum dot solar cell and OPV (PbS-CQD/OPV) as shown in Fig. 43. The interlayer can be made of only oxides as observed [643,658] or as window layers (as highly conductive transparent oxide films, with minimal parasitic adsorption protecting the absorber layers from damage during fabrication or for device stability) [647,659].

Fig. 42. Application of oxides as transport layers in solar cells. (a) Schematic representation of a PSC. (b) SEM image of a PSC applying TiO2 as the ETL. (c) Band alignment of the PSC shown in (a) and (b). (d) Selected examples of transport layers made of metal oxides and the power conversion efficiency obtained when applied as electron (ETM) or hole (HTM) transport materials in emerging solar cells. DSSC: Dye sensitized solar cells; PSC: Perovskite solar cells; OPV: Organic solar cells. Detailed information can be found in [636], CuGaO2 in OPV, DSSC and PSC [650,651,653–655], CuCrO2 in PSC [648–650,652], PMO: polyoxometalate [645], TaWOx in PSC [656], FeTiOx in PSCs [657].
for the indium tin oxide and molybdenum oxide interlayer (MoOx/ITO) or the ITO/SnO2/ZTO (ZTO: ZnSnOx). The combination of an oxide and a semiconducting polymer is also possible, for example ZnO and sodium pentachlorophenate (ZnO/PCP-Na) or MoOx and PCDTBBT.2

2.3. Oxides as light harvesters

Oxides applied as the main light harvesting material in solar cells constitute what is known as all-oxide solar cells (AOSCs), a solar cell technology currently in expansion. We can identify two types of oxides for the indium tin oxide and molybdenum oxide interlayer (MoOx/ITO) or the ITO/SnO2/ZTO (ZTO: ZnSnOx). The combination of an oxide and a semiconducting polymer is also possible, for example ZnO and sodium pentachlorophenate (ZnO/PCP-Na) or MoOx and PCDTBBT.2

2 PCDTBBT: Poly[N-9’-heptadecanyl-2,7-carbazole-alt-5,5-(4’,7’-di-2-thienyl-2’,1’,3-benzothiadiazole)], Poly[9-(1-octylnonyl)-9H-carbazole-2,7-diyl]-2,5-thiophenediy1-2,1,3-benzothiadiazole-4,7-diyl]-2,5-thiophenediyl].
doped CuO₂ and ZnO₃GeO₄ results in PCE above 8% (Fig. 44) and the multilayer heterojunction solar cell of the type Ga₂O₃Al₂O₃Na-CuO₂ results in 6% efficiency (see Fig. 44). Among binary and ternary oxides, ZnO has remarkable properties such as low processing cost, its capability for being syn-thetized at low temperatures, its wide direct band gap (−3.37 eV) and high electron mobility (−120 cm²/Vs). Bulk CuO₂ solar cells with ZnO layers have resulted in PCE efficiencies of 4.31% where both the CuO₂ bulk and ZnO are doped with Na and Cu₂, respectively. Some ternary oxides applied as heterojunction with CuO₂ are CuGaO₂, ZnGa₂O₄, MgInO₃, ZnSnO₃, ZnGeO₂, Zn²FeO₂, CuInO₂, AgInO₂ among others. Multicomponents of two or more binary compounds such as ZnO, MgO, Ga₂O₃, Al₂O₃, InOₓ and SnO₂ being the largest efficiency observed for ZnGa₂O₄ (5.36%) with an open-circuit voltage of Vₜ = 0.81 V (Fig. 43). Fig. 44 shows the power conversion efficiency for AOSC applying oxide-oxide heterojunctions. For more details see [636,637].

Photoferroics include photoactive oxides with ferroelectric (FE), antiferroelectric (AFE) [661] and multiferroic (MF) (ferroelectric and magnetic) properties. A ferroelectric material is characterized by its spontaneous electric polarization. Their light-to-energy conversion characteristics have resulted in the observation of what is known as the anomalous photovoltaic effect (APE) and the bulk photovoltaic effect (BPE). In the APE, the photovoltage is larger than the band gap of the corresponding semiconductor with voltages that can reach thousands of volts. In the BPE, the giant photovoltages are observed in single crystals, and are due to the non-centrosymmetric structure of the oxide. Up to now, the development of the large photovoltage is limited by the very low current densities observed, which results in too low power conversion efficiencies for a photovoltaic device.

The most studied ferroelectrics are oxides with the perovskite structure. This is a very large family with composition ABO₃, where A and B each represent a cation element or mixture of two or more such elements or vacancies. Lithium niobates (LiNbO₃) are an important type of oxides exhibiting ferroelectricity. These are characterized by their trigonal paraelectric structure, and are very closely related to the perovskites oxides and materials such as lithium tantalite. Among non-perovskite ferroelectrics, the layered ferroelectric oxide based on Bi₂O₃ slabs and perovskite-like blocks, are of particularly interest. Ferroelectric photovoltaics is nowadays much in vogue due to the above band gap voltages observed in BiFeO₃ and the development of a ferroelectric semiconductor band structure engineering to obtain ferroelectric with bandgap within the visible range. Fig. 45 shows some examples of AOSCs applying FE and MF oxides as absorbers. In this graph, only the very few examples showing power conversion efficiencies above 1% are shown (for a recent detailed review on these oxides with efficiencies below 1% see [636,637]). Although solar cells made with photoferroic oxides are characterized for very low current densities and thus, low power conversion efficiencies, values up to 8% have been recently reported. To achieve this, the oxide band gap was tuned by doping and the final device was made as multijunction stacking three identical subcells: Back in 2011 the oxide Bi₂FeCrO₆ (BFO), grown on NbSrTiO₃, showed a record high power conversion efficiency of 6%. The photocurrent and the photovoltage were Jsc = 1 mA/cm² and Vₜ = 0.6 V. Piezoresponse force microscopy revealed a prominent ferroelectric imprint and it was linked with strong inner electric field driving charges generated via photoelectric effect toward the electrodes. In 2014, the same authors reported a complete mapping of bandgap tuning of the BFO layers to improve itsphotovoltaic performance under AM 1.5G irradiation (100 mW/cm²) [662]. Following spin-density functional analysis, authors assigned the band in the visible range to charge transfers between Cr and Fe mixed d orbitals, e.g., Fe/Cr cationic ordering. This was effectively tuned during the PLD growth adjusting parameters such as the growth time. The PCE improved by decreasing the growth rate, reaching an impressive maximum value of 3.3%. The photocurrent and the photovoltage were Jsc = 11.7 mA/cm² and Vₜ = 0.79 V. They ascribed the increased Jsc to an enhanced light absorption and reduced recombination due to narrower direct bandgap. The optimized stack of three BFO layers further increases the power conversion efficiency up to 8.1%. Thus, the key in photoferroic oxides seems to reside in the optical band gap tuning and the multilayerjunction configuration.

2.4. Enhanced stability of solar cells applying oxides

Organic semiconductors are the most applied barrier layers in OPV, DSSC or PSC. The high hygroscopic and acidic nature of poly (3,4-ethylene dioxythiophene):poly(styrenesulfonic acid), PEDOT: PSS, has demonstrated to be the cause of limited solar cell lifetimes. Another issue is its low work function (WF) (about 5.0−5.1 eV) which limits the photoactive materials to be applied and forms imperfect ohmic contacts at interfaces in some solar cells [637]. Moreover, organic semiconductor are also sensitive to oxygen-containing atmospheres and UV light. Semiconductor oxides like NiO, MoO₃ or VO₂ are some examples of high WF oxide interlayers introduced on solar cells as replacement of PEDOT: PSS. Their high WF and higher stability to moisture has resulted in highly stable organic and perovskite solar cells (see Fig. 46). Among the many properties of some perovskite oxides, and of high relevance for solar cell stability, is their good electron mobility and inferior UV photocatalytic ability. La-doped BaSrO₃ (LBSO), for example, is a ternary perovskite oxide with exceptionally high electrical mobility of 320 cm² V⁻¹ s⁻¹ at room temperature. It also shows high UV resistant properties if compared with the most studied metal oxides in emerging solar cells, e.g., TiO₂ or SnO₂. These properties make this oxide a good candidate to work as an ETM. Recent studies have
demonstrated that the replacement of TiO$_2$ by LBSO in halide perovskite solar cells provides good resistance to UV light and thus, enhanced device stability [663]. Another example is the ferroelectric oxide PbZrTiO$_3$ (PZT). Under UV light the inhibition of poling results in the decrease of the reversible remnant polarization. Nevertheless, more recent studies report the enhancement of the remnant polarization under UV light [664]. Understanding the effect of the UV light on FE oxides could contribute to unveil similar enhancement in transport properties and solar cells stability [665].

Fig. 46 shows some examples of the stability enhancement in conventional and carbon-based perovskite solar cells and also on organic solar cells, when semiconductor oxides are applied. The graph shows the performance loss (decrease in efficiency) after stability analysis for several solar cells. The more evident improvement is the replacement of PEDOT by an oxide. The solar cell applying PEDOT shows a compete degradation after stability analysis but the device applying the semiconductor oxide demonstrates excellent stability (Figs. 46(c) and 46(i)).

The interfacial modification of the oxide is also advantageous for device stability as shown by the modification of ZnO with polyethylenimine ethoxyalted (PEIE) molecule (Fig. 46(a) and (b)). Other examples is the printable version of the PSC, called the Carbon-based PSC, it has recently shown 10,000 h (above one year) stability under constant irradiation. This is the longest ever reported lifetime for a PSC. This solar cell applies semiconductor oxides as ETM (e.g. TiO$_2$) and inert oxides as scaffolds (e.g. Al$_2$O$_3$ or ZrO$_2$) where the halide perovskite is infiltrated. It is consider that the excellent stability observed for this solar cell is in part due to the application of a interfacial modifier, the aminovaleric acid iodide AVA, which can functionalize the TiO$_2$ oxide reducing surface defects and electron/hole recombination. But most important is the functionalization of the inert oxide scaffold by AVA, which seems to be the key for the long-term device stability [674].

3. Future prospects/current and future challenges

Among the many aspects to consider for the development of novel oxides for high efficient and stable solar cells, here we discuss three of them: (1) synthesis of oxides, (2) defects and (3) multilayer junctions.

3.1. New methods for the synthesis of highly crystalline oxides at low temperature

Innovation in printed optoelectronics (including printed photovoltaics) require the fabrication of thin films by low cost methods (e.g. solution processing) which includes the use of low sintering temperatures for device fabrication. The synthesis of oxides at low or mild temperatures is easily made for binary oxides and some ternary oxide compounds. Some examples are the synthesis at low temperature by sol-gel methods of binary oxides like V$_2$O$_5$ [668] or MoO$_3$ [675]. The synthesis of doped and ternary oxides like TaWO$_4$ [656], CuCrO$_2$ [648], CuGaO$_2$ [651] or LaBaSnO$_3$ [663] have also been engineered in order to be obtain at temperatures compatible with solar cell fabrication requirements of low temperature. Nevertheless, the synthesis of more complex oxides (i.e photoferroics) requires high sintering temperatures (usually above 600 °C) which limits the substrate to be applied. In solar cells, a complete device requires that at least one electrode to be transparent. This restricts the possible useful substrates to transparent conducting oxides (TCO), which nowadays are limited to ITO and FTO on glass (although other oxides can be applied). These two TCO maintain their good conductivity properties at temperatures not above 450 °C and 550 °C respectively. Nevertheless, some oxides, especially complex oxides, like ferroelectric/multiferroic oxides, require temperatures above 600 °C in order to preserve their ferroelectric properties. To solve this problem, Calzada, et al., has developed synthesis methods compatible with plastic substrates, at temperatures below 400 °C. Specifically, the PhotoChemical Solution Deposition (PCSD) technique that combines the preparation of light-sensitive sol(solution)-gel precursors and their irradiation with UV-light to produce a metal – O – metal skeleton, from which an appreciable reduction in the temperature of crystallization of the functional metal oxide is achieved [676]. Other examples report on the application of short sintering times (up to 10 min) at temperatures of 600-650 °C for the synthesis of ferroelectric oxides (e.g. PZT) deposited on glass/FTO substrates [637]. This method allows the synthesis of the oxide with the desired crystalline structure and ferroelectric properties without damaging the conductivity of the glass/TCO electrode. Another possibility is the synthesis of the oxide as nanostructured material (nanoparticles, nanorods, etc.) at superior temperatures and the posterior preparation of the oxide electrode on TCO substrates applying pastes. This methodology is currently applied for the preparation of the mesoporous TiO$_2$ electrodes employed in DSSC and PSCs. This method requires the synthesis of highly crystalline mesoporous oxides. However, some solar cell applications require the use of dense thin films with high crystalline quality. Recently, a dense layer of La-doped BaSnO$_3$, an oxide which usually requires temperatures above 1000 °C for its preparation, was synthesized through a crystalline superoxide-molecular cluster (CSMC) colloidal solution. The crystalline LBNO perovskite phase was successfully converted into a pure perovskite BSO phase by annealing for 30 min at 200 °C, which requires a much shorter time and lower temperature than conventional routes. The oxide was applied as ETM in efficient PSCs showing 1000 h stability, which is superior to the 500 h reported for the same PSC applying the TiO$_2$ oxide [663].

3.2. Defects in oxides and interface engineering

Defects in oxides, like oxygen (O$_{vac}$) and metal vacancies (M$_{vac}$), can determine the final properties of a solar cell, especially its PCE and long-term stability. Defects largely govern the electronic and optical properties of oxides, these are charge centres that control the donor or acceptor character of the material and can be potential binding sites for anchoring molecules. The transport properties of oxide semiconductors depend on defects caused by deviation in stoichiometry and the presence of impurities. Electron transport semiconductor oxides, like TiO$_2$, are characterized by an oxygen vacancy (O$_{vac}$)-mediated conductivity. In TiO$_2$, high concentration of O$_{vac}$ translates in higher electric conductivity, but the transport of electrons between O$_{vac}$ defect sites mediates the transport of photogenerated holes. The fractal pathway involves empty O$_{vac}$ defect sites which conduct electrons to holes, leading to undesirable charge recombination in solar cells. In hole-transport oxides, for example NiO, transport properties depend mostly on metal vacancies and the conduction of electricity takes place through positive hole conduction. Thus, treatments given to the oxide surface affect in great extend its photochemical behaviour. For example, oxygen plasma treatment in TiO$_2$ (an ETM) reduces its conductivity, while under the same conditions the conductivity of NiO (a HTM) increases. Due to the detrimental effect that defects have in solar cell response, different methods have been proposed to passivate or eliminate them. For example, the application of organic interfacial modifiers with anchoring groups specifically selected to bond with oxides [677] or the application of coatings of secondary oxides, like Al$_2$O$_3$ applied to supress surface defects, avoid interfacial recombination, and enhance device stability [678]. Defects on these oxides are affected by the synthesis method, the doping level, the posttreatment atmosphere, the annealing temperature, the passivation and functionalization of the surface, among others.

3.3. Multilayer junction and solar cell configuration

High efficient solar cells, e.g. halide perovskite solar cells and organic solar cells, utilize electron and hole transport layers at both sides of the light harvester material for the effective extraction of charges.
However, in AOSC applying complex oxides like photoferroics, the application of homojunctions (single oxide layer) is a common practice, as well as the addition of a single transport layer adjacent to the light absorber. Still, multilayer junctions have demonstrated the charge extraction through the addition of an electron and/or hole transport layer. For example, the analysis of the ITO/PbZrTiO₃/Pt solar cell results in negligible response with efficiency at 0.008%. The addition of a Cu₂O hole transport layer between the PbZrTiO₃ and the Pt electrode resulted in an efficiency increase up to 0.57% [679]. In another example, the efficiency of ITO/ Bi₂FeCrO₆/NaSrTiO₃ solar cell, initially at 0.5%, was effectively increased up to 2% by the addition of a NiO layer between ITO and the Bi₂FeCrO₆ [680]. The most efficient AOSCs up to date correspond to a multijunction made of ITO/ Bi₂FeCrO₆/Bi₂FeCrO₆/Bi₂FeCrO₆/ NaSrTiO₃ with 8.1% efficiency [662], MgF₂/Al-ZnO/Ga₂O₃/Al₂O₃/ Na-Cu₂O with 6.25% or MgF₂/Al-ZnO/(Ga₀.₉₇₅Al₀.⁰₂₅)₂O₃/Cu₂O:Na with 6.1% [681]. In all these cases, the solar cells are the result of incorporating multilayers to the final device. In some cases, a multilayer is used to enhance light adsorption [662], in others the multilayer is the result of adding electron or hole extraction layers to enhance charge extraction [682]. Also, high relevance is the application of doped oxides with good electrical conductivity as the metal electrodes for example the use of Al-doped ZnO or Na-doped Cu₂O, among others.

4. Concluding remarks

This work is a critical survey of the state-of-the-art and future prospects of the application of different metal oxides in third-generation solar cells (e.g. dye sensitized, organic, halide perovskite, all-oxide solar cells). We analysed the application of a broad spectrum of oxides, from binary and ternary oxides to more complex oxide compounds. We analysed the influence of the type of oxide applied on the power conversion efficiency of the final solar cells. We analysed the beneficial effect of replacing organic semiconductors, like PEDOT:PSS, by semiconductor oxides in emerging solar cells. Especially important is the enhancement observed on the stability of solar cells under ambient atmosphere and moisture for the long-term stability under irradiation conditions. We briefly commented on the need for careful control of surface and bulk defect in oxides and the need to passivate oxygen vacancies to avoid charge recombination and the loss of device stability. Several options can be applied for the passivation of defects; one of them is the functionalization of oxide surfaces with organic functional molecules. We have also commented on the importance of applying oxide interfaces as hole or electron extraction layers (or both) in all-oxide solar cells, especially in ferroelectric/antiferroelectric and multiferroic oxides. We believe there are new possibilities for the enhancement of the efficiency in these solar cells when applied as multi junction devices. Finally, the ferroelectric and multiferroic nature of these oxides could permit the development of new functionalities (for example protection from UV light to enhance device stability) and unseen photovoltaic response due to the presence of remnant polarization (which can permit new forms of charge transfer under selected conditions).

Acknowledgements

We thank the financial support from the SNSF NRP70 “Energy Turnaround” and GRAPHENE project supported by the European Commission Seventh Framework Programme under contract 604391 is gratefully acknowledged. To the Spanish MINECO through the Severo Ochoa Centers of Excellence Program under Grant SEV-2013-0295; for the grant ENE2016-79282-C5-2-R and the OrgEnergy Excellence NetworkCTQ2016-81911-REDT. To the Agència de Gestió d’Ajuts Universitaris i de Recerca for the support to the consolidated Catalonia research group 2014SGR-1212 and the Xarxa de Referència en Materials Avançats per a l’Energia (Xarmae). To the CERCA Programme/Generalitat de Catalunya and to the European COST Action StableNextSol project MP1307.

All-oxide heterojunction solar cells

Rubén Tamayo, Albert Calleja*

OXOLUTIA S.L., Avda. Castell de Barberà, 26, Tallers 13, Nau 1, 08210 Barberà del Vallès, Barcelona, Spain

*Corresponding author.
E-mail: acalleja@oxolutia.com (A. Calleja).

Abstract

Solar cells based on oxides are being studied worldwide to lower the cost of photovoltaic solar energy or finding new market niches like powering devices in the age of the Internet of Things or tandem cells which potentially allow higher exploitation of the solar spectrum. If we exclude solar cells based on ferroelectric oxides, p-type copper (I) oxide or Cu₂O stands as the first choice as absorber, with a bandgap of 2.1 eV and a theoretical efficiency as high as 20%. Typical n-type oxide is ZnO but more sophisticated phases are explored with more suitable characteristics. Doping of the Cu₂O with alkaline or alkaline-earth elements also is a fruitful research area which is developing fast in order to decrease resistivity while retaining high mobility and carrier concentration. In fact, the record efficiency value is 8.1% for pulse-laser deposited (Zn, Ge) O₂ on Na-doped Cu₂O obtained from thermal oxidation of copper sheets. Different synthesis approaches are described for the Cu₂O. However, thermal oxidation of copper metal, electrodeposition and magnetron sputtering appear as the most used. On the other hand, chemical solution approaches are also promising for cost reduction by using liquid-type depositions like inkjet printing followed by suitable thermal treatments. All-oxide solar cells also include Transparent Conductive Oxides (TCOs). Significant effort is also being devoted to improve them. In this article, these topics are discussed presenting altogether recent results that show that all-oxide heterojunction solar cells continue to progress in order to become technically and economically feasible in several photovoltaic applications.

Keywords: Cu₂O-based solar cells; Transparent coating oxides; Tandem solar cells; Inkjet printing; Internet of things

1. State of the art

The increasing energy demand during the last decade has led to the need of seeking alternatives to conventional silicon solar cells [683]. Metal oxides are materials widely employed in solar cells as transparent electrodes and electron-selective layers. However, their use as absorber layers is still under research because of the low efficiencies in all-oxide solar cells reported to date. Because of the abundance, chemical stability and the possibility to manufacture by roll-to-roll (R2R), metal oxides are promising materials for low-cost solar cells. Different absorber oxides such as Co₃O₄, BiFeO₃, Fe₂O₃ and CuO have been studied so far, being Cu₂O the most investigated. Cuprous oxide (Cu₂O) is a p-type semiconductor with a direct band gap of 2.1 eV and because it is non-toxic and presents a theoretically energy conversion (η) of around 20%, Cu₂O is considered a rising material for thin film solar cells. In the 70–80% Cu₂O Schottky junctions were deeply studied, showing efficiencies of around 1–2% [684]. It was found that a copper-rich layer is formed at the Cu₂O/metal interface because of Cu₂O reduction and subsequent interdiffusion phenomena, causing a barrier height and low conversion performance [685].

A solar cell can also be made by joining two regions with p-type and n-type doping. In the case of Cu₂O most reported solar cells have been...
made by depositing different n-type semiconductor upon the absorber layer, forming a p-n heterojunction. However, there are some papers reporting Cu$_2$O-homojunction solar cells, which we will comment later on. The n-type semiconductor acts as electron-selective layer and reduces the electron-hole recombination that takes place in the Cu$_2$O/top-electrode interface. In addition, the energy barrier originated as a consequence of the band alignment is expressed as a conduction-band offset ($\Delta E_{CB}$) and is critical for current and voltage losses. However, $\Delta E_{CB}$ can be tuned by using proper tern-ary materials [686], modifying the conduction/valence band position.

Cuprous oxide has been widely studied by several deposition techniques leading to electrical and structural properties in a large range. Electrodeposition and thermal oxidation are the most common techniques employed to date. By electrodeposition an aqueous solution contains a copper salt precursor such as Cu$_2$SO$_4$ or Cu(CH$_3$COO)$_2$ in acid lactic. The pH is adjusted to basic values of 12–13 by adding NaOH, LiOH or KOH. In the thermal oxidation pro-cess Cu$_2$O is obtained from the annealing of a Cu sheet at high tem-perature (800–1000 °C) in Ar/N$_2$ atmosphere. Both methods differ in the achieved electrical properties. Typically by electrodeposition resistivity values of 10$^5$–10$^6$ Ω·cm, mobility of 0.4–1.8 cm$^2$/V·s and hole-concentration of 10$^{12}$–10$^{15}$ cm$^{-3}$ [687], respectively, are obtained while values of 10$^4$ Ω·cm, 10$^3$ cm$^2$/V·s and 10$^1$–10$^2$ cm$^3$/V·s are achieved by thermal oxidation [688]. Differences in resistivity and mobility can be attrib-uted to the crystalline size characteristic of each technique.

1.1. Doping Cu$_2$O

Cuprous oxide is considered an intrinsic p-type semiconductor confirmed experimentally and by ab initio calculations. Despite the origin of p-type behaviour is still under debate, it is mainly attributed to the formation of Cu vacancies in the lattice that act as acceptors. Because Cu$_2$O has an intrinsic high electrical resistivity (10$^5$–10$^6$ X cm), significant attempts have been made to decrease it by doping. To achieve interesting efficiencies, much lower resistivity values are required while maintaining high electrical mobility. In particular, Minami et al. [689] studied the sodium incorporation to a thermally oxidized copper foil from a post-annealed treatment with Na$_2$CO$_3$ powder. They reported a decrease of resistivity from 10$^9$ to 10$^7$ X cm and an increase of hole concentration from 10$^{13}$ to 10$^{18}$ cm$^{-3}$. The Na atom acts as donor and is incorporated at interstitial sites of the Cu$_2$O lattice associating with an oxygen vacancy. In consequence, a copper vacancy is released and a hole in the valence band is generated.

Another attempt to dope was carried out with nitrogen by Ishizuka et al. [690]. They prepared Cu$_2$O films in an rf-sputtered chamber controlling the O$_2$ and N$_2$ flow rates and temperature by which the Cu$_2$O single phase is stable. By varying the N$_2$ flow rate at a fixed temperature of 400 LC the optimum resistivity and hole concentration were 15.2 X cm and 10$^{15}$–10$^{17}$ cm$^{-3}$, respectively. It is expected that the N atom incorporates to the Cu$_2$O lattice at oxygen interstitial sites acting as acceptor.

Some authors have claimed n-type conductivity in chlorine-doped Cu$_2$O by electrodeposition. A theoretical study demonstrated that for non-doped Cu$_2$O, there is no intrinsic defect source causing n-type conductivity [691]. In addition, Nian et al. [692] found experimentally that the cause of this behaviour could be explained in terms of surface defects as nanocavities during the deposition. These defects could allow absorption of Cu$^{2+}$ ions because of intrinsic Cu vacancies causing an inversion layer and hence a p-n conductivity transition. Biccari et al. [693] studied the chlorine doping on Cu$_2$O obtained from the thermal oxidation of a Cu foil. The doping was performed using two methods: the classical one in which a chlorine vapour from an MgCl$_2$ source is introduced during the oxidation process [684]; and evaporation of a CuCl layer before the oxidation. They reported an improvement of resistivity and carrier concentration with 200 X cm and 8.6 × 10$^{13}$ cm$^{-3}$, respectively. Indeed, Hall effect measurements showed p-type conductivity for all Cl-doped Cu$_2$O samples.

1.2. An overview of Cu$_2$O-based solar cells evolution

In Fig. 47 a plot with reported efficiencies of all-oxide Cu$_2$O-based solar cells by different deposition techniques is presented. The common material architecture is also shown, formed by a Cu$_2$O absorber layer, typically with a thickness of 1–200 μm and an n-type emitter semiconductor layer in which light shines on, usually with lower thicknesses of 50–200 nm. A MgF$_2$ layer above the TCO is usually employed as anti-reflective coating. The p-n heterojunction is processed with a transparent coating upon the n-type material which acts as top electrode and a metal contact below the Cu$_2$O as back contact. Several top and back electrodes have been employed since the band alignment is crucial for the performance of the device. Furthermore, metals with high work functions are required, such as Au and Ag. As can be seen in Fig. 47, higher efficiencies are obtained by thermal oxidation of Cu at high temperature.

Fig. 47. (a) Plot showing all-oxide Cu$_2$O-based solar cells reported efficiencies by different deposition techniques for the Cu$_2$O layer. (b) Scheme of a typical Cu$_2$O-based solar cell.
Mittiga et al. [694] proposed an ITO/Cu2O/Au solar cell with 1.07% of efficiency by annealing a Cu sheet at high temperature in 2006. Then they inserted a ZnO layer by ion beam sputtering reaching 2% of efficiency and an open circuit voltage of 0.59 V. This enhancement could be achieved because of the lower electron affinity of ZnO in comparison to ITO. A performance improvement of 3.83% was presented by Minami et al. [688] with similar layer materials and the same method for obtaining Cu2O. Instead of ITO they used AZO as top electrode, which has a better band alignment with ZnO (lower ΔEcb).

It was argued that the recombination of carriers in a Cu2O-based solar cell is originated because the inverse of absorption coefficient (α−1) is greater than the minority carrier diffusion length (Ld). For this reason, different nanostructures, such as nanowires, nanotubes, nanorods and nanoparticles have been proposed as n-type semiconductor to improve charge separation and increase the p-n contact area. Electrodeposited ZnO is the most versatile electron transport material used in terms of growth and shapes. However, all Cu2O based solar cells containing nanostructured interfaces have shown poor efficiencies. For instance, Yuhas et al. [695] developed a solar cell with Cu2O nanoparticles and ZnO nanowires by drop casting with 0.035% of efficiency. Another attempt was performed by Jia et al. [696]. In this case they deposited ZnO nanorods by hydrothermal method obtaining 0.88% of efficiency. A study of an electrodeposited Cu2O/ZnO solar cell suggest that the low performance (Voc) when using nanowires can be attributed to a low built-in-potential achieved at the heterojunction for small distances between adjacent nanowires [697].

In 2013 Minami et al. [689] suggested that the open circuit voltage (Voc) and fill factor (FF) could be improved by reducing energy discontinuities at the interface, originated by the difference in electron affinities of p-n-type semiconductors. They proposed Ga2O3 as n-type semiconductor, reaching 5.38% of efficiency and 0.8 V of Voc. Since sodium was confirmed to improve the electrical properties of Cu2O, an AZO/Ga2O3/Cu2O:Na/Au solar cell with 5.53% of efficiency was achieved [689]. It was demonstrated the importance of optimizing the heterojunction by selecting an appropriate selective layer and the electrical properties of Cu2O.

Electrodeposited Cu2O-based solar cells with reasonable high efficiencies have also been reported. For instance, Lee et al. [699] reached a value of 3.97% and 1.2 V of Voc by depositing Ga2O3 by ALD. Because ALD is a conformal deposition technique, the n-type layer is totally in contact with the p-type Cu2O layer. This Voc, which is the highest reported, could be explained because of lower surface defects due to a better p-n contact. It is well-known that lower mobility and hole concentration are usually obtained by electrodeposition to the detriment of current density (Jsc) and FF. Thus, an electrodeposited n-doped Cu2O-based solar cell should present a low performance.

The use of ternary compounds has demonstrated an important improvement in terms of Voc and efficiency (Table 6) because the cation/anion ratio might improve the band alignment reducing ΔEcb. In such a way, an efficiency enhancement of 6.25% was reached by depositing n-type Ga0.975Al0.0252O3 by PLD onto thermally oxidized Na-doped Cu2O [700]. Recently, n-type Zn0.38Ge0.62O has been deposited by PLD leading to 8.1% of efficiency in a AZO/Zn0.38Ge0.62O/Cu2O:Na/Au solar cell, Voc and Jsc up to 1 V and 10 mA/cm2, respectively. This result corresponds to the best Cu2O-based solar cell to date [701].

Table 6 Performance summary of Cu2O-based solar cells with different n-type ternary compounds. Values obtained from references [700,701,712].

<table>
<thead>
<tr>
<th>Solar cell architecture</th>
<th>α (%)</th>
<th>Voc (V)</th>
<th>Jsc (mA/cm2)</th>
<th>FF (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZO/Zn0.38Ge0.62O/Cu2O/Au</td>
<td>8.10</td>
<td>1.1</td>
<td>~10</td>
<td>~65</td>
</tr>
<tr>
<td>AZO/Ga0.975Al0.0252O3/Cu2O:Na/Au</td>
<td>6.25</td>
<td>0.84</td>
<td>10.8</td>
<td>69</td>
</tr>
<tr>
<td>AZO/Ga0.975Al0.0252O3/Cu2O:Na/Au</td>
<td>6.10</td>
<td>0.84</td>
<td>10.95</td>
<td>66</td>
</tr>
<tr>
<td>AZO/Ga0.975Al0.0252O3/Cu2O:Na/Au</td>
<td>5.42</td>
<td>0.84</td>
<td>10.11</td>
<td>64</td>
</tr>
<tr>
<td>AZO/Al-Ga-Mg-Zn-O/Cu2O:Na/Au</td>
<td>5.40</td>
<td>0.96</td>
<td>10.6</td>
<td>53</td>
</tr>
<tr>
<td>AZO/ZnGa2O4/Cu2O/Au</td>
<td>3.56</td>
<td>0.81</td>
<td>10.22</td>
<td>65</td>
</tr>
<tr>
<td>AZO/Zn0.38Ge0.62O/Cu2O/Au</td>
<td>4.29</td>
<td>0.80</td>
<td>10.99</td>
<td>51</td>
</tr>
<tr>
<td>AZO/ZnSnO3/Cu2O/Au</td>
<td>3.64</td>
<td>0.63</td>
<td>9.93</td>
<td>59</td>
</tr>
<tr>
<td>AZO/ZnSnO3/Cu2O/Au</td>
<td>2.92</td>
<td>0.57</td>
<td>9.99</td>
<td>51</td>
</tr>
<tr>
<td>AZO/Zn-Sn-O/Cu2O/Au</td>
<td>2.65</td>
<td>0.55</td>
<td>7.37</td>
<td>65</td>
</tr>
<tr>
<td>AZO/MgIn2O3/Cu2O/Au</td>
<td>2.58</td>
<td>0.50</td>
<td>7.82</td>
<td>66</td>
</tr>
<tr>
<td>AZO/GaIn2O3/Cu2O/Au</td>
<td>2.22</td>
<td>0.52</td>
<td>8.13</td>
<td>65</td>
</tr>
<tr>
<td>AZO/CuGaO2/Cu2O/Au</td>
<td>1.49</td>
<td>0.33</td>
<td>8.09</td>
<td>56</td>
</tr>
<tr>
<td>AZO/CuIn2O3/Cu2O/Au</td>
<td>1.24</td>
<td>0.44</td>
<td>5.04</td>
<td>56</td>
</tr>
<tr>
<td>AZO/CuIn2O3/Cu2O/Au</td>
<td>0.75</td>
<td>0.29</td>
<td>5.35</td>
<td>49</td>
</tr>
<tr>
<td>AZO/AgIn2O3/Cu2O/Au</td>
<td>0.12</td>
<td>0.24</td>
<td>1.19</td>
<td>45</td>
</tr>
<tr>
<td>AZO/ZnGeO2/Cu2O/Au</td>
<td>0.03</td>
<td>0.35</td>
<td>0.21</td>
<td>40</td>
</tr>
<tr>
<td>AZO/ZnSnO3/Cu2O/Au</td>
<td>0.01</td>
<td>0.39</td>
<td>0.04</td>
<td>48</td>
</tr>
</tbody>
</table>

1.3. Transparent coating oxides

Transparent Coating Oxides (TCO) are inorganic materials employed in electronics as flat-panel displays, touch-panel controls, electrochromic mirrors/windows and light-emitting diodes. Particularly, in solar cells they are used as the top electrode to collect the electron-carriers generated at the p-n junction. An appropriate TCO may be highly electrical conductive and have a low absorption coefficient in the visible range. A good tool to analyse the quality performance is given by [702]:

$$\sigma = \frac{1}{R_s \times \ln(T+R)}$$

where σ is the electrical conductivity in siemens, α is the optical absorption coefficient, Rs is the sheet resistance in ohms per square, T is the transmission and R is the reflectance for a given wavelength. The higher the α/σ ratio, the better the TCO performance. Different considerations must be kept in mind when choosing a TCO, such as physical, chemical and thermal stability, etchability, deposition temperature, uniformity, toxicity and manufacturing cost. Because electrical and optical performance depend on electronic properties of a semiconductor, it is also important to consider the band gap, the carrier concentration and the work function [703]. In addition, the deposition technique influences on the properties, being chemical vapour deposition (CVD) and sputtering the most common used techniques. Indeed, they can also be obtained by evaporation, pulsed laser deposition (PLD), sol-gel, chemical bath deposition and electroplating.

SnO2, ZnO, In2O3 are the most common oxides used as TCOs in solar cells. These materials have a band gap in the range of 3–4 eV, hence they are intrinsically transparent and insulators at room temperature. To achieve good electrical properties, the oxide requires to be doped by impurities (i.e. F, A13+, Sn4+, In3+, Ga3+, Sb5+) in substitutional sites of the lattice usually originated by oxygen vacancies. As a consequence of increasing the carrier density in a range of 1016–1021 cm−3, the Fermi level displaces into the conduction band and the semiconductor becomes n-type degenerated.

In2O3-based TCOs currently offers the lowest electrical resistivity. In particular, Sn-doped In2O3 (ITO) is widely employed in display applications. In2O3-ZnO (IZO) alloys have been proved to be transparent conductors with remarkable thermal stability. Nevertheless, indium is a rare and expensive element which increases the production cost of these TCOs. Al-doped ZnO (AZO) has been com-monly used in Cu2O/ZnO-based solar cells because of its low resistivity (8.5 × 10−3 Ω cm) and its band alignment with ZnO. Another example is F-doped SnO2 (FTO) which has received attention for its low cost associated to the raw materials and also for the good electrical properties and thermal stability it presents.

2. Future prospects and current and future challenges

Cu2O-based solar cells still show efficiencies significantly below its theoretical value, which hamper their practical use in photovoltaics. For this reason, it is early to anticipate any substitution of other commercially well-established PV cells. However, we shed some light on the current and future challenges of these all-oxide solar cells.
2.1. Quality control of p-n heterojunction

The low efficiency of Cu$_2$O-based solar cells reported until now can be attributed to a non-optimized heterojunction in terms of electronic properties, need for nanostructuring and purity [704]. Further research on n-type semiconductors is needed to reduce energy discontinuities at the interface. Furthermore, pure cuprous oxide phase is required for obtaining higher efficiencies. Gunnæs et al. [705] investigated the formation of CuO at Cu$_2$O/ZnO interfaces experimentally and by DFT calculations. They found that even in stable thermodynamic conditions for the Cu$_2$O, a 5 nm-textured CuO is obtained onto the c-axis-oriented ZnO. The CuO buffer layer is favored in terms of strain and interface energy and adversely affecting the efficiency. This result is an important contributor to the difference of theoretical and experimental efficiency values of these Cu$_2$O-based solar cells.

2.2. Tandem solar cells

A tandem solar cell is a multi-layered material which is equivalent to the junction of two or more single solar cells in series. According to Kirchhoff’s law, for a tandem solar cell connected in series, the $V_{oc}$ across the device corresponds to the summation of each $V_{oc}$ sub-solar cell while $J_{sc}$ depends on the fill factor of each sub-device and remains constant. Since each semiconductor absorbs light in a different part of the solar spectrum (because of different band gaps), the integration of two different semiconductors enables a broader absorption range. In consequence, a tandem solar cell should present a better performance than a single solar cell by itself. Since Cu$_2$O has a maximum absorbance in the blue-UV spectra and silicon absorbs poorly in this wavelength range, a tandem solar cell could be applied to integrate both semiconductors (Fig. 48). Since the manufacturing of silicon-based tandem heterojunction solar cells (STHSC) with a high efficiency at low cost has still not been accomplished, a flexible and low-cost process will be needed to satisfy the future solar cell demand.

2.3. Inkjet printing technology: towards low cost all-oxide solar cells

Chemical Solution Deposition (CSD) is a well-known methodology widely used to obtain functional ceramic oxides such as thin films, nanoparticles, mesoporous solids films or bulk ceramics [706–708]. The main advantage of this methodology is the low cost associated to the process because the use of vacuum systems is not required as in vapour deposition techniques. In addition, the precursor solution can be modified in terms of molarity and solvent to achieve different desirable thickness. Another advantage is the possibility to dope in-situ by adding a doping agent to the solution. Furthermore, CSD enables for scaling-up a process. In particular, inkjet printing is presented as a cost-effective alternative for depositing functional metal oxides [709]. It allows an accurate control of the volume solution deposition and it does not require high vacuum conditions. Moreover, it is a flexible technology which permits the easy switching of the deposited material and the pattern during the process. For this reason, inkjet printing technology can be suitable for tandem solar cells. The idea of manufacturing flexible Cu$_2$O-based solar cells [710] with a low-cost process by ink jet printing stands as a good opportunity for many applications in which long lengths and flexibility are required. In Fig. 49 we present our

![Fig. 48. Scheme of a Cu$_2$O-Si-based tandem solar cell.](image)

**Fig. 48.** Scheme of a Cu$_2$O-Si-based tandem solar cell.

![Fig. 49.](image)

**Fig. 49.** (a) SEM image of the surface of a Cu$_2$O layer grown by CSD on metallic tape MgO/Hastelloy. (b) Electrical properties by Hall effect measurement: resistivity dependence on temperature and carrier density dependence on magnetic field of a Cu$_2$O layer deposited by CSD on LaAlO$_3$ (1 0 0).
The results of non-doped Cu$_2$O obtained by CSD at high temperature. The layer has been deposited from a metalorganic solution compatible with inkjet printing technology onto a flexible metallic tape MgO/Hastelloy. The SEM image shows a compact surface with a good crystallinity. Some electrical properties have been studied by Hall effect measurement in a temperature and applied magnetic field range of 240–400 K and 0–9.5 T, respectively. At 298 K the measured resistivity and hole concentration are 130 $\Omega$cm and $10^{16}$ cm$^{-3}$, respectively, which are reasonable values for non-doped Cu$_2$O.

We also present some photovoltaic characterization results by photoconductive AFM (pc-AFM) of a CSDCu$_2$O/SPUTTERINGZnO/FTO/Quartz solar cell (Fig. 50). In this solar cell the substrate is FTO/Quartz in which a ZnO layer of 50 nm has been deposited by sputtering at room temperature. The subsequent Cu$_2$O layer which has been deposited from a metalorganic solution at high temperature has a thickness of 100 nm. Since the final morphology of the Cu$_2$O layer is not dense. To deposit a back metal for further efficiency measurement has not been possible because it would has caused the device to short-circuit. For this reason, pc-AFM is a good alternative to characterize the photovoltaic response by applying a bias voltage between the Cu$_2$O and FTO layers and measuring the current in dark and light conditions. In Fig. 50(a) a current-topographic map obtained at 0.3 V of bias voltage is presented. An increase of current in light conditions can be observed due to the photovoltaic response of the device. Due to the local capacity of this technique, a current-voltage curve has been measured in a single Cu$_2$O grain in a bias voltage range from-2 to 2 V (Fig. 50(b)). Both in dark and light conditions the device shows an electrical rectifying behaviour, as in a p-n heterojunction is expected. The approximately three orders of difference in current between the two curves shows a good grain electrical conductivity. However, a back-metal contact is required to collect the photoelectrons generated at the heterojunction for an operative solar cell.

2.4. Internet of things

Cu$_2$O-based solar cells can be part of the new technological revolution so-called ‘Internet of Things (IoT)’, in which a variety of daily objects -such as food packages, furniture and books- interact with each other. The main idea of this revolution is to make smart cities with objects (or things) that are able to collect information of its environment and send it by wireless connection. In this paradigm objects are composed of small sensors, microchips and antennas which permit to sense, process and send data (Fig. 51). In many applications, objects could use available ultra-low power chips in sensors with voltage supply of 1–2 V and current consumption of 100–500 pA [711]. The intermittent work mode of many devices enables the use of batteries which could be charged continuously by solar energy. Since trillions of IoT devices are expected in the next decade, low cost and clean power supply is required.

3. Concluding remarks

Low cost, abundant and non-toxic p-type Cu$_2$O continues to be the favourite choice for all-oxide heterojunction solar cells, offering a high theoretical efficiency of 20%. Great progress has been made in the last years by theoretical works, properly understanding doping in this oxide
to mainly reduce the resistivity, by careful selection and optimization of n-type oxides and TCOs as well as nanoengineering of the p-n interface [712]. Maximum efficiency so far is 8.1%. Furthermore, reaching the 10% mark in efficiency in the near future seems likely, which would strongly accelerate the research and development in this type of photovoltaic solar cells architectures.

Nonetheless, to demonstrate cost-efficiency and continuous reel-to-reel processing for applications like IoT might be a good opportunity towards real market for these materials. In this way, deposition of Cu2O from chemical solutions by inkjet printing holds promise towards that goal, as it has been shown in our experimental results. Future work must be focused on obtaining a compact and continuous Cu2O layer in the solar cell integration while doping the Cu2O to decrease the electrical resistivity.

Acknowledgements

The authors would like to acknowledge the financial support from ACCIO-Generalitat de Catalunya and the European Union for the 3D-PHOTOXIDES ERA-NET MANUNET II and from IV Fondo de Emprendedores of Fundación Repsol for the SOLAR OXIDES project. The authors would also thank Mr. Andrés Gómez from Institut de Ciencia de Materials de Barcelona (ICMAB) for the pc-AFM measurements and Mr. Alex Hernández from Institut de Recerca en Energia de Catalunya (IREC) for the sputtering and valuable guidance.

Photoferroelectrics

I. Fina a,*, C. Paillard b, B. Dkhil c

aInstitut de Ciencia de Materials de Barcelona (ICMAB-CSIC), Campus UAB, Bellaterra 08193, Barcelona, Spain
bPhysics Department, University of Arkansas, Fayetteville AR 72701, USA
cLaboratoire Structures, Propriétés et Modélisation des Solides, CentraleSupélec, CNRS-UMR 8580, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
*Corresponding author.
E-mail: ifina@icmab.es (I. Fina).

Abstract

The giant photovoltaic effect due to the bulk photovoltaic effect observed in multiferroic BiFeO3 thin films has triggered renewed interest in photoferroelectric materials for photovoltaic applications. Tremendous advances have been achieved in photoferroelectrics yielding improved power conversion efficiencies (by up to 8.1%) by increasing absorption using narrow bandgap ferroelectrics. Other strategies, such as the more efficient use of the internal electric fields in ferroelectrics, are ongoing. Moreover, as a byproduct, several significant progress has been made into the characterization of the photostriction phenomenon. Here, we review ongoing and promising routes to improve the photoresponse of ferroelectric materials.

Keywords: Oxides; Ferroelectrics; Photovoltaics; Photostriction; Photoferroelectrics; Films

1. Introduction

Ferroelectric materials demonstrate switchable spontaneous electric polarization when subject to an external electric field. In the case of photoferroelectric materials, most of which are oxides, both their electrical and polar properties are affected by visible light. Photoferroelectric materials in the form of thin films can present giant open-circuit voltages under illumination [713], or a switchable photoelectric response [714–716], which are exclusive characteristics of interest for use in applications. Photovoltaic (among others) applications of photoferroelectrics materials [717–721] is a research area receiving an in crescendo interest due to the potential for high photovoltaic efficiencies. Bulk photoferroelectric materials have recently revealed another interesting property; high levels of photostriction, or the ability to change shape under illumination [722]. Here, we will present the latest advances in the study of photoferroelectric oxide thin films, concentrating on results that attempt to improve photovoltaic properties as well as photostrictive efficiency.

2. State of the art

The main factors that determine the efficiency of photoferro-electrics are the same as those which apply to any photovoltaic device: absorption, charge collection, and exciton dissociation (Fig. 52). The main endeavors of the scientific community to increase absorption principally revolve around research into narrow bandgap ferroelectrics; namely solid state solutions ([KNbO3]1-x[BaNi1/2Nb1/2O3-δ]x [723]), novel ferroelectric materials (Bi4FeCrO6 with record efficiencies of 8.1% [724]), or others, particularly Bi-based compounds [725–728]. Hexagonal manganites have also been proposed as potentially

Fig. 52. Schematic summary of the factors that determine photovoltaic efficiency and possible strategies to improve them using ferroelectric oxide materials. In black typeface are the strategies that have delivered success, while potentially interesting new strategies are shown in blue.
interesting materials [729] due to their small bandgap (approximately 1.5 eV); experimental results are still scarce but they show improved efficiency compared to typical ferroelectrics [730].

Charge collection efficiency is related to the presence of internal electric fields (IEFs), although it also depends on carrier mobility and lifetime. In photoferroelectrics, charge collection can also result from the bulk photovoltaic effect (BPV). In presence of BPV, the asymmetric momentum distribution of the photogenerated carriers, due to the non-centrosymmetric crystal environment, generates a photocurrent [731]. Although, it can be also explained in terms of a virtual shift in the real space [732,733]. Seminal demonstrations of the presence of BPV were those observed in doped LiNbO3 [734], and BaTiO3 [735]. More recently, BPV has been reported in thin film forms of BiFeO3 [736–738], and BaTiO3 [739–741]. While the BPV effect can produce giant open-circuit voltages under illumination, the photocurrent generated is low because of the short distances (l0 = 10–100 nm) [731,742] that photo-carriers can travel. In presence of BPV, efficient collection has been achieved by reducing the size of the device: by growing the studied material in the form of thin films [739], or through the use of micro- or nanometer-sized tips acting directly on the ferroelectric surface (tip-enhanced technique) [736], thus reaching high local efficiencies [740] although the device’s overall efficiency remains low [743]. Ferroelectric domain wall engineering has also been suggested to improve efficiency [713], but with contrasting results regarding domain walls relevant role [737,738]. The IEFs in a ferroelectric material are the imprint electric field (Eimp) and the depolarization electric field (Edep), and therefore Edep = 1/ε Edep + Eimp. Eimp can be generated by strain gradients, among other effects, which is of particular interest because they can generate very strong electric fields (up to 2 MV/cm [744]). Edep results from the unscreened ferroelectric surface charge. In a standard metal-ferroelectric-metal capacitor, the amount of unscreened surface charge is closely related to the metal’s electronic properties and the metal-ferroelectric interface [745]. The upper limit of Edep is equal to the electric field generated by a fully unscreened polarization (that is Edep < 5 MV/cm in BaTiO3). However, Edep also induces suppression of ferroelectricity (consequently leading to Edep suppression) and so we must find an equilibrium between ferroelectricity and Edep. IEFs in ferroelectric oxide thin films have been studied widely and the different contributions of Edep and Eimp to the photocurrent analyzed, leading to conflicting results regarding Edep and relative relevance [715,746–751]. Metal-ferroelectric-metal junctions also present an electric field associated with the presence of a Schottky barrier at the metal-ferroelectric interface, usually called the built-in electric field (Eb). The presence of an Eb and its dependence on polarization produces interesting features, such as switchable rectifying behavior [714,752,753], but in this case the efficiency is limited by the ferroelectric bandgap [721] and one cannot expect greater efficiencies than those observed for semiconductor-based devices. Note that the artificial design of IEFs, for example, via nanolayering, which is the combination of a pure ferroelectric material with a high-absorption oxygen-deficient one [754] or the combination of different ferroelectric materials [755], can significantly enhance the photovoltaic output.

Photostriction is a further functional property of ferroelectric materials that can be combined with the photovoltaic effect. Photostriction can, in principle, appear in all materials, through light-induced thermal expansion, potential deformation (modification of the electron density distribution and thus the internal electronic pressure), or electrostriction [756]. Ferroelectric materials provide an additional property with the natural occurrence of piezoelectricity, which creates a strain field under illumination when coupled with a photo-induced electric field. To date, photostriction in ferroelectrics has mainly been reported in traditional ferroelectric materials, such as LiNbO3 [754] and (Pb1-xLa)x(Zr1-xTi)xO3 [757–760], yet more recent research has focused on multiferroic BiFeO3 [722,761] because it has a relatively low bandgap (2.2–2.7 eV) compared to other typical ferroelectric materials. The current use of photostrictive materials as optically-driven mechanical actuators, such as in micro-walking robots and photophones, was proposed long ago based on a PLZT bimorph structure [759]. More novel applications have also been investigated by depositing a magnetic nickel film on a single BiFeO3 crystal. It was demonstrated that the coercive field of the nickel magnetic hysteresis loop could be increased by 50% when the bismuth ferrite crystal was illuminated [762], thereby opening up routes to optically-controlled magnetic memories.

Recent studies have also focused on the ultra-short dynamics (below the picosecond or femtosecond) of the photostriction effect in thin films. In particular, strains lasting less than 100 ps were photo-induced in bismuth ferrite thin films and monitored by means of Bragg peak shifts in pump-probe time-resolved X-ray diffraction experiments [763–765]. Similar results were observed in lead titanate thin films [766]. Pump-probe transient reflectivity experiments have also highlighted the possibility of exciting GHz-frequency acoustic waves, plus shear acoustic waves, with a greater intensity than longitudinal ones, in bismuth ferrite ceramics and sin-gle crystals [767,768].

3. Challenges and future prospects

The discovery of new materials with a bandgap approaching the ideal value of 1.4 eV (lower limit to avoid thermalization [769,770]) is a very important milestone in photoferroelectrics. However, the search for new oxide materials with a focus on carrier mobility optimization is another important avenue for research. Cationic substitution is a potentially interesting strategy along this line. For instance, taking the archetypical example of BaTiO3, where the conduction band is mainly dominated by carriers moving through orbitals, appropriate dopants can help manipulate the conduction band to allow transfer through orbitals, thus enhancing the material’s properties, resulting in greater charge collection and enhanced efficiency.

Other options to improve efficiency concern the intrinsic differential properties inherent to ferroelectric materials. These can be summarized as: IEFs and high dielectric permittivity. Although there are several studies exploring the role of IEFs in ferroelectrics on photovoltaic efficiency, there is a lack of research addressing their manipulation to improve performance [771]. As mentioned previously, large internal electric fields can be obtained by engineering the electrode-ferroelectric interface to improve Edep or by generating large strain gradients to improve Eimp. Otherwise, the large dielectric permittivity observed near the Curie point in ferroelectric materials, which is easily tailored by manipulating the epitaxial strain [772] or composition [773], can enhance exciton dissociation, whose binding energy is inversely proportional to the dielectric permittivity. Furthermore, selecting appropriate electrodes might help improve the IEFs and also inhibit recombination [774] or increase the fill factor [775] using similar strategies as those applied in semiconductor technology. This potential improvement in photovoltaic efficiency would also result in greater surface photocatalytic activity, as ferroelectric materials have already been reported to possess more enhanced levels of this property than their non-ferroelectric counterparts [776]. This will open new paths to light energy conversion beyond isolated photovoltaic effects.

Regarding IEF-driven photoelectric effects and their relationship with photostriction, the converse piezoelectric effect generated by a photo-induced electric field is believed to be the main source of deformation in ferroelectric materials due to light. Recent first-principle calculations have produced further evidence of this scenario [777–779]. However, Paillard et al. [779] have shown that some subtleties exist; the piezoelectric effect appears to be the dominant force in directions with large piezoelectric constants and large photo-induced changes in polarization (often along or close to the polar axis), meanwhile photo-induced potential deformation can become sizable in directions perpendicular to the polar axis in lead titanate, for example. Moreover, the potential deformation effect has been demonstrated to either enhance or contrastingly impede the photostriction effect.
depending on the transition involved. This means the photostriction effect can be improved by tailoring the band structure properly and selecting a specific transition (e.g., through meticulous selection of the light polarization).

It is also important to realize that coupling light to the lattice dynamics in ferroelectrics offers an exciting perspective. For instance, coupling light to IR phonons was predicted and subsequently proven to offer the possibility of reverse polarization, provided that there is strong anharmonic coupling between the IR phonon and soft-mode phonons [780,781]. Similarly, Li et al. [763] have observed the presence of large photo-induced strain gradients at the picosecond scale, which could lead to polarization control through the flexoelectric effect. Finally, Lejman et al. [768] reported that the generation of GHz coherent acoustic phonons through strong laser pulses allowed for strong, fast, acousto-optic, sub-bandgap light propagation control, which is promising for the design of the next generation of ultrafast acousto-optic devices.

4. Concluding Remarks

Perovskite halide-based materials [782] are grabbing the attention of the scientific community working on photovoltaics due to their high conversion efficiencies, but they present the significant disadvantages of containing toxic materials and low stability [783]. Ferroelectric oxides might improve an interesting alternative, since they are stable and some of the examples studied to date do not contain any toxic elements. The tremendous advances during recent years have achieved high efficiencies by increasing absorption (8.1% [724]) and even locally exceeded the Shockley–Queisser limit [740] in ferroelectric oxide films, which could give rise to promising applications. Future lines of research aimed at increasing efficiency include the optimization of the properties (either “positive” or “negative”) that distinguish photoferroelectric oxide materials from the materials currently used in photovoltaics, namely internal electric fields, high dielectric permittivity, and low mobility. There is a lot of room for improvement by exploiting the positive properties of photoferroelectrics and reducing the impact of the negative ones.

Acknowledgements


Progress of indium-free transparent conducting oxides

Shrabani Panigrahi *, Rodrigo Martins, Elvira Fortunato *

CENIMAT/ISTN, Departamento de Ciência dos Materiais, Faculdade de Ciências e Tecnologia (FCT), Universidade NOVA de Lisboa (UNL), and CEMOP/UNINOVA, 2829-516 Caparica, Portugal.

E-mail: emf@fct.unl.pt (E. Fortunato).

Abstract

The current status and future prospects for further development of indium-free transparent conductive oxides (TCOs) have been discussed in this report. The greater part of the industrially accessible TCOs are n-type such as Sn doped In2O3, Al doped ZnO, F doped SnO2 due to their much superior electrical performances when compared to transparent p-type oxides. In general, low mobility is observed for p-type oxides due to large hole mass or hopping transport mechanisms. The preparation of these and other TCOs semiconductors are essential for their increasing demand in optoelectronic device applications. The best indium-free transparent electrodes alternative to ITO are Al- and Ga-doped ZnO (AZO and GZO). Improvement in the depositions techniques is needed for the development of new TCOs over large area substrates.

Keywords: n-type and p-type TCOs; AZO; GZO; ATO; FTO

1. State of the art

TCOs have led the field of transparent electrodes since the 1950s. Today, the TCOs adopted by industry are typically based on In2O3, SnO2 and ZnO and their subsequent combinations being used in several optoelectronic applications such as solar cells [784,785], flat panel displays [786], touch screens, light emitting diodes [787,788] and transparent electronics [789,790]. Another reason for TCO developments is their applications in flexible displays and electronics [791,792]. The ideal TCO should have simultaneously two fundamental properties: high transparency (average transmittance above 80% in the visible range) and low resistivity (~10-4 Ω cm). The most important TCO utilized from the last 50–60 years is tin doped indium oxide, Sn:In2O3, commonly called indium-tin-oxide or ITO due to its good electrical and optical properties associated with an easy and well-established thin-film technology. At present, this material shows the best available performance in terms of conductivity, transmissivity, and also its reproducibility. Nevertheless, ITO has some drawback for its high price as well as the price fluctuation due to market demand.

Notwithstanding the enormous commercial success, extensive research is still being going-on in order to find alternatives to replace indium. There is a vast literature available on TCOs since the first production of a TCO thin film by Badeker in 1907, but more than 90% are related to n-type and many potential applications of TCOs are still limited by the availability of p-type TCOs. On the other hand, it is quite difficult to find out shallow acceptors in wide band gap oxides to make a significant hole population. In 1993, a TCO thin film consisting of p-type semiconductor was first reported [793]. A p-type TCO, CuAlO2, with useful conductivity ~1 S cm-1 was reported later in 1997 [794]. After the successful preparation of CuAlO2, motivation to investigate similar kind of p-type TCO materials increased. There are many reports on the preparation of p-type TCO thin films using new TCO semiconductors [795–798]. However, there is no report which provides details about the p-type TCO thin film applied as a transparent electrode. So, suitable p-type TCOs are still being researched. In addition, future research should concentrate on some issues like the development of new deposition techniques, the role of impurities and defects in order to design the next generation TCO films for practical applications.

Fig. 53 shows the evolution over the last 16 years of indium-free TCOs in terms of the optical and electrical properties. Due to the high number of available papers, only the best values are reported. We would like to comment that over the years we have observed stabilization in terms of the optical and electrical properties, while for the recent years remarkable advances have been obtained for TCOs produced by solution process reaching values of the ones produced by vacuum processes.

2. Future prospects/current and future challenges

The improvement of the physical and chemical properties of the TCO films is required due to their huge demand as thin-film transparent...
The discovery of TiO$_2$Nb shows another kind of TCO, rather than mixing of the known binary TCOs. The research on new TCO materials tries to upgrade the efficiency of the optoelectronic gadgets like solar cells with maintaining maximum optical transparency and charge carrier mobility. For this the band alignment of TCO with other existing layers is highly necessary. Finally, further development in the deposition techniques is also required for the preparation of TCOs over large area substrates needed for industrial applications.

3. Concluding remarks

We are facing an increasing demand of TCOs due to their huge applications as transparent electrodes in flat panel displays, solar cells, LEDs and other opto-electronic devices. In future, the development of transparent thin film electrodes using polycrystalline or amorphous transparent conducting oxide is important because a stable supply of ITO will not be sure since indium is an exceptionally costly and rare material. A number of researchers are trying from both theoretical and experimental point of view for further development of TCO thin films. The potential outcomes for the development of TCOs performance in the following decade are very inspiring.

Looking into the future, newer requirements from completely disruptive applications will lead to a segmentation of the transparent conductive materials industry in order to follow the industry demands. The need for lower resistivity and simultaneously high transparency especially in the infra-red region will continue to drive researchers and industries. Disruptive solution-based thin film technology will reduce the use of vacuum processes. Developments in lighting and display industries. Disruptive solution-based thin film technology will reduce the use of vacuum processes. Developments in lighting and display applications will require minimal optical losses, maximum light extraction and low cost materials. For shielding and sensor packaging, lightweight, 3D and conformal solutions will be still more important. All these areas will impose new and sustainable materials followed by advanced engineering solutions.

Acknowledgements

The authors would like to acknowledge L. Pereira and P. Bar-quinha for fruitful discussions during the manuscript preparation. We

![Diagram of TCOs](image-url)
Electrochromic and thermochromic oxide materials

Gunnar A. Niklasson*, Claes G. Granqvist

Department of Engineering Sciences, The Ångström Laboratory, Uppsala University, P.O. Box 534, SE 75121 Uppsala, Sweden

*Corresponding author.
E-mail: gunnar.niklasson@angstrom.se (G.A. Niklasson).

Abstract

Electrochromic (EC) and thermochromic (TC) oxides have important applications in devices for modulating optical properties. EC thin films change their optical transmittance of visible and solar radiation under electrically stimulated insertion/extraction of ions. A typical device sustains ion transport between two EC oxide films via an electrolyte; one film (typically W-oxide-based) darkens under ion insertion and the other film (typically Ni-oxide-based) darkens under ion extraction. The film/electrolyte/film three-layer construction is positioned between transparent electrical conductors which normally are oxide-based films. The main application of these devices is in “smart” windows and glass facades for energy efficient buildings. Future developments include multi-component oxides with optimized optical properties and superior electrochemical durability. TC oxides have qualitatively different optical properties when their temperature is below/above a “critical” value \( T_c \). The most widely used TC materials are based on VO\(_2\); thin films are infra-red transmitting/reflecting at low/high temperature while the luminous transmittance remains moderately high. Corresponding nanoparticle composites are transparent/near-infrared absorbing at low/high temperature. TC thin films and nanoparticles can be used in windows which admit solar energy preferentially when there is a heating demand in the building. Future developments include multicomponent VO\(_2\)-based materials with additions for adjusting \( T_c \) to near room temperature and to decrease the luminous absorptance, integration of these materials in multilayer constructions and nanoparticle composites, and implementation of oxidation-protecting coatings. Multi-functional devices—e.g., with EC and TC properties—yield interesting future scenarios.

Keywords: Electrochromism; Thermochromism; Optical transmittance modulation; Smart windows

1. State of the art

Electrochromic (EC) and thermochromic (TC) oxides have optical properties that can be modulated electrically and thermally, respectively. They have numerous applications in current and forthcoming technology; the largest one—at least in terms of areas—is in “smart” windows and glass facades (“smart” glazing) for energy efficient buildings with good indoor comfort. Other uses include windows for vehicles, switchable motorcycle helmets and eyewear, non-emissive information displays and surfaces with tunable thermal emittance.

There exist EC oxides of two types: those that darken under ion insertion are called “cathodic” and others that darken under ion extraction are referred to as “anodic”\(^{[813]}\). Oxides based on W, Mo, Ti and Nb are cathodic while oxides based on Ni and Ir are anodic. An intermediate situation prevails for V\(_2\)O\(_5\) which exhibits anodic and cathodic features in different wavelength ranges. The right-hand panel of Fig. 54 introduces a typical oxide-based EC device design\(^{[813,814]}\). The central part is an ion conductor, which can be a polymer electrolyte containing H\(^+\) or Li\(^+\) ions, or consist of a thin film, which may be an oxide, for example Ta\(_2\)O\(_5\) or ZrO\(_2\). This part joins an EC thin film and a thin-film counter electrode, and the entire three-layer stack is positioned between transparent electrical conductors both of which must be transparent if the device is for see-through applications. The EC film is almost always W-oxide-based, because this oxide appears to be more stable than the alternatives, whereas the counter electrode can be of various kinds but is typically based on Ni oxide or V oxide. Film thicknesses are usually 200–300 nm. The transparent conductors normally consist of Sn-based or In-based doped wide band gap oxide semiconductors. Optical modulation takes place when a few volts is applied between the transparent conductors so that charge is shuttled...
between the W-oxide-based film (darkening under ion insertion) and the Ni-oxide-based film (darkening under ion extraction). This charge exchange involves small ions, typically H$^+$ and Li$^+$, in the two films and the corresponding insertion/extraction of charge-balancing electrons is responsible for the modulation of optical absorption. Clearly, the operation of an EC device is akin to that of a thin-film electrical battery, and both kinds of devices share characteristic response times for property changes, open-circuit memory, and degradation under long-time cycling.

Two principally different device configurations are used for EC glazing: one is a monolithic design with five superimposed thin films on a single glass substrate and the other—the principle of which is given in Fig. 54—embodies a polymer electrolyte laminate between two substrates each with a two-layer coating backed by glass or polyester. The polyester-based construction permits flexible EC devices which can be employed for glass lamination as indicated in the left-hand panel of Fig. 54. The optical modulation extends over the solar energy wavelength range (300–3000 nm), which includes luminous radiation (i.e., visible light at 400–700 nm).

TC devices are in principle simpler than EC devices and can comprise a single layer [815,816]. Only VO$_2$-based oxides stand out as viable alternatives for most applications. VO$_2$ has a reversible structural transformation between a monoclinic and semiconductor phase and a triclinic and metallic phase at a “critical” temperature $T_c$ of 68°C. Thin films of VO$_2$ transmit/reflect infrared solar radiation (700–3000 nm) moderately well below/above $T_c$ whereas VO$_2$ nanoparticles in transparent hosts are transparent/absorbing in the same wavelength range below/above $T_c$. Both thin films and nanoparticle composites have a significant and largely temperature-independent luminous transmittance ($T_{lum}$).

Pure VO$_2$ is of limited interest for glazing and most other applications, but recent advances in materials development have changed this situation [815]. Fig. 55 shows the state of the art for VO$_2$-based thin films and nanoparticle composites as compiled from experimental and computational work related to glazing [817]. The desired property is a large magnitude of $T_{lum}$ especially in the low-temperature state and large difference between solar transmittance in the low-temperature and high-temperature states ($\Delta T_{sol}$). It is apparent that dramatic improvements have ensued from the use of antireflection (AR) coating, Mg-doping to widen the semiconducting band gap and enhance $T_{lum}$, and nanoparticle formation.

### 2. Current and future challenges

EC technology has been under development since the 1970s but significant practical applications have emerged only in the 2010s [818]. A number of challenges have impeded the technology—in particular with regard to large-area devices—including development of EC thin films with adequate coloration and open-enough nanostructure to allow facile ion transport, availability of high-performance transparent electrical conductors, transparent ion conductors (electrolytes) of sufficient quality, strategies for electrical powering that assure long-term durability, and low-cost manufacturing technology.

Thin films based on W oxide can be improved by intermixing with other oxides. Recent work on W–Mo–Ti oxide shows that both color and electrochemical durability can be significantly improved [819] and this result points towards further refinement by use of multi-component oxides. Ni-oxide-based films in EC devices are even more challenging, but recent work [820,821] once again shows the advantages of multi-component oxides. Ni oxide is different from W oxide in that the EC effect is confined to the film’s surface, at least for Li$^+$-conducting electrolytes [822], which indicates that surface engineering can be deployed in future work to enhance the properties.

The ion conductor, at least in the laminated EC design, can be improved and functionalized by addition of doped oxide semiconductor nanoparticles, which are able to limit the throughput of solar radiation while $T_{lum}$ remains large in the high-transmittance state of the device [823]. Initial work needs to be extended to establish performance limits for this approach.

Long-term durability is as important as it is challenging for EC devices. Recent explorative work on life-time assessment of individual EC films [824,825] needs further extension to other EC oxides as well as to full EC devices. Other recent research has shown that degraded EC films can be rejuvenated by straight-forward electro-chemical treatments [826,827], which may open new avenues towards superior EC devices, but the initial studies again need expansion to more types of EC films and, especially, to full EC devices. Furthermore, the physical and chemical processes underlying life-time prediction and film rejuvenation are virtually unknown and require a sustained research effort with many challenges, experimental as well as theoretical.

Regarding manufacturing, further work is desired for large-area deposition of EC films and for continuous EC films. Such technologies have been implemented [828] but further developments are needed to cut costs.

Looking now at TC technology, significant advances have been reported recently, but much further work is needed to unleash the full potential of this technology. VO$_2$ is likely to remain in focus, but different admixtures to this oxide are needed to displace $r_c$ to the vicinity of room temperature while $T_{lum}$ is concurrently boosted by band gap widening. Recent investigations [815,816,829] have shown that such properties are possible but challenges remain to optimize the materials.

A large magnitude of $\Delta T_{sol}$ is needed for practical TC-based glazing, and the key seems to be structuring at the nanoscale (i.e., “nanothrombromism”). Efficient and reliable preparation of VO$_2$-based nanoparticles is challenging and needs further development, which is feasible by following several paths encompassing chemical preparation, gas-phase synthesis, physical vapor deposition of alternate discontinuous VO$_2$ films and continuous dielectric films, and others.

Furthermore, it should be realized that VO$_2$ films and nanoparticles are not thermodynamically stable but revert slowly to non-TC V$_2$O$_5$ over time. Protective coatings of Al$_2$O$_3$ seem to stabilize VO$_2$ films [830], but it is urgent that analogous core–shell structures be studied in more detail for nanoparticles.

### 3. Concluding remarks

EC and TC technologies have come a long way since their inceptions in the 1980s. Many challenges have been met but others remain, as outlined above. One fundamental impediment is that some of the most crucial materials—such as NiO and VO$_2$—are rather poorly understood from a basic physics perspective. For example, the nature of the reversible phase transformation at $T_c$ has been debated for decades—and
Multi-functionality is a notion that recurs in recent work on EC and TC devices and should be further exploited in the future. There are many possibilities, e.g., electrochromism combined with energy generation [831] and energy storage [832], “dual-band” EC devices with separate tuning of $T_{\text{lim}}$ and transmittance in the 700–3000 nm range [833], and combined EC-TC devices with TC nanoparticles embedded in the electrolyte of the EC construction.

Another, more technical, area ripe for research and development deals with large-area flexible EC foil structures, probably based on transparent or translucent ethylene tetrafluoroethylene (ETFE) with proven durability under long-term solar exposure. Such foils would enable light-weight membrane architecture with tunable optical properties and conceivably lead to a new paradigm for the built environment.

Acknowledgements

Several of the recent advances in electrochromic and thermo-chromic materials, reported above, were reached with financial support received from the European Research Council under the European Community’s Seventh Framework Program (FP2007–2013)/ERC Grant Agreement No. 267234 (“GRINDOOR”).

Ionotronics and nanoionics in energy devices: Current status and future of micron-solid state fuel cells (μ-SOFC)

Simone Sanna, Nini Pryds*

Department of Energy Storage and Conversion, Technical University of Denmark, DK-4000 Roskilde, Denmark

*Corresponding author.

E-mail: nipr@dtu.dk (N. Pryds).

Abstract

The progress on interfacial engineering of oxide electronic where the electrons are the primary charge carriers offers ample opportunities to discover new physics targeting a large number of applications. Equally significant, is the area involving migration of ions in solid matters which is known as nanoionics. By combining the electronic and ionic migration of charges, a new –area emerged: ionotronics. In this part of the roadmap we will discuss the possible use of ionotronics in energy related technologies with an emphasis on micro Solid Oxide Fuel Cells (μ-SOFC). We will discuss the importance of confining the ionic and electronic currents in a predefined path designed using, e.g., heterostructures and/or interfaces. The opportunity to combine electronic and electronic charges in a single device holds great potential in many fields, such as optoelectronics, information technology and energy storage/conversion.

Keywords: Ionotronics; Nanoionics; Oxide heterostructure; Ionic and electronic conductivity; μ-SOFC

1. Where are we today?

Ion and electron transport in a solid is one of the most fundamental processes in solid state science. The interest in ionic and electronic oxide conductors has rapidly increased in recent years generating key applications in solid-state electrochemical devices, including sensors, solid oxide fuel cells and oxygen separation membranes [7,834]. When two oxides “meet each other” at the interface, an unprecedented access to new physics and chemistry emerges due to the breaking of the symmetry of the system, thus promoting unique new properties which otherwise do not exist in the parent compounds, such as magnetism, superconductivity and ionic conductivity [4,835–838]. We are now facing the opportunity to achieve further advances, this time by combining the ionic and electronic motion at a confined system, i.e. ionotronics [4,7,834–838]. Ionotronics rely on coupled transport of charges and are useful in a wide variety of devices, e.g. electrochemical transistors [839], resistive switching [840], magnetoresistive devices [4], liquid electrolyte gating in electronic and ionotronic devices [841], batteries [842], energy storage devices [843] and fuel cells [7,836,838,844,845]. Common to all of these applications are that the performance of the device relies heavily on interface engineering [4,7,835,836,838,844]. To further nurture the development in this area, a better understanding and control of the interfacial phenomena is needed. Utilizing, e.g., polar discontinuities [4], epitaxial strain [838], size effect [844] and charge carrier doping [839] has proven successful. The palette of strategies to tune the properties at the interfaces provide a “special tool box” useful for optimizing the transport of ions and electrons. A high concentration of charge carries and a fast-ionic mobility at low temperature are the keys parameters to achieve an improve performance of solid state based energy devices such as SOFCs [845]. However, while defects, e.g. oxygen vacancies and protons, can be controlled in metal oxides by both doping and nano-structuring [843,844], enhancement of the ionic charge mobility at low temperature is still challenging. This is due to the intrinsic limitations of the ionic bonding in the solid compounds, restricting the achievement of fast ionic mobility only in liquid electrolyte materials. Material engineering at the nano- and even at the atomic scale has opened a new avenue for increasing the ionic mobility. Fig. 56 shows two different examples of interface engineered devices: (1) A heterostructure with interfaces parallel to the substrate [836] and (2) a nano-scaffold film with vertical heterointerfaces, also called vertical aligned nanostructures (VANs) [846]. By confining the charges in vertical or lateral channels one can achieve a control over the device performance with high uniformity and reproducibility.

Solid Oxide Fuel Cells (SOFCs) is a good example illustrating the use of ionotronics and nanoionics in energy related devices In μ-SOFCs, ionotronics and nanoionics are used as a tool to improve the building blocks of the device, i.e. electrolytes, anodes and cathodes. As examples, combined fast mixed ionic and electronic conductivity can be observed in VANs, where fast oxygen redox cycles, e.g. for cathodic applications, can be achieved in oxide thin films [847]. Moreover, the reduced size of these components allows achieving a high energy density in small power sources applicable for portable electronic devices. Conventional SOFCs operate currently in the range 650–1000°C [848]. However, for small portable devices, this temperature is too high. Reduction of the operation temperature can be achieved by: (1) reducing the electrolyte thickness to few nanometers leading to a decrease in the diffusion path length of the oxygen ions, thus reducing the ohmic resistance considerably, and (2) choosing electrolyte material with high ionic conductivity at low temperature [849,850]. In μ-SOFC, all the components are fabricated as thin films of a few hundreds of nanometres or less, by using for example thin freestanding ionic yttria-stabilized zirconia (YSZ) membranes as electrolyte, coated e.g. with porous platinum on each side serving as electrodes [851–853]. To make a useful system and give functionality to the thin film based μ-SOFC, this freestanding membrane is supported by a silicon-based microfabricated platform [851]. Such a design allows an easy integration of the functional components of the fuel cell on Si utilizing microfabrication technologies. Other designs of thin-film-based SOFC proposed in the last years used alternative substrates, ranging from pure metals to technical glasses [845]. More recently, the use of electro ceramic, metals and/or oxide electrodes, have been also shown to push down the limits of the operating temperature [849,850,854–858].
Although progress has been made in promoting μ-SOFC-system in portable electronics, for example ceramic supported μ-SOFCs showing remarkable performances in the range of ca. 588 mW/cm² at 500°C [854], the research is still at laboratory scale. Table 7 shows a comparison of a number of μ-SOFC devices, their substrates, electrodes, electrolyte materials as well as the performance of the cells.

### Future challenges

Despite significant advancements in the μ-SOFCs technology, this power sources are still plagued by problems that inhibit their viability for many commercial uses. The development of nanoionics for μ-SOFCs has been extremely challenging due to the following obstacles: (1) large surface to volume ratio resulting in a structural stability of the nanos-structured materials which is highly thermally sensitive [849,850,854–858], (2) degradation and instability due to the mismatch in the thermal expansion between the different materials of the cell [845,849,850,854–858], (3) the nanoionic electrode layer must be mesoporous for gas penetration and subsequent electrochemical reaction at the triple phase boundaries (TPBs), but mesoporous metal oxide thin films often result in morphology changes and loss of mesoporous structure at high temperature operation of the cell [858], and (4) the few nanometer thick electrolyte layers require a high uniformity on large areas (20–30 cm²) without pinholes and cracks. Indeed dense layers having pinholes and cracks in porous substrates lead to cross mixing of gases through the electrolyte and thereby lower the power density for the μ-SOFC.

Recent advances in engineering ionic-electronic transport, i.e. ionotronics using a nanoionic approach have shown a large potential of improving the ionic conductivity when the material is in the form of thin films and nanocrystalline structures. Mixed ionic and electronic conducting thin films, e.g. in VANs, shows very promising performances for ultra-fast oxygen reduction, even for low temperatures, where usually electronic polarization control the fuel cell performances. The ionic conductivity at the interface can be increased by several orders of magnitude or even change its nature, e.g. from a pure ionic to mixed ionic and electronic conductor, owing to the symmetry breakdown near the surface leading for example to space charge effects, vacancy concentration and/or local strain fields [4,7,835,836,838]. Furthermore, single electrolyte layer can be extended by periodic heterostuctures (multilayers) which offer the possibility to exploit and carefully investigate interfacial effects on transport, charge, and reactivity. The potential impact of heterostructures is already manifested by showing that an increase in ionic conductivity was observed with increasing number of interfaces [837]. There are, however, still interesting thermodynamic questions concerning the limits of heterostructures and the transition towards a newly formed artificial layered phase that need to be taken into consideration. The reduction of the electrolyte size to the nanometer range enhances the ionic conductivity but also, in some cases, leads to undesired effects such as instability due to fast mass diffusion and increase of electronic defects, such as the small polarons [835]. These polarons can at high concentrations hamper the SOFCs performances, e.g. by introducing parasite electronic leakage at the electrolyte side [835].

Ionic conduction enhancement combined with strategies for controlling the ratio of ionic-to-electronic contributions could be therefore very useful.

### Table 7

Comparison of fabrication, materials, OCV and power density of various micro-SOFC.

<table>
<thead>
<tr>
<th>Group</th>
<th>Substrate</th>
<th>Anode</th>
<th>Electrolyte (process/thickness)</th>
<th>Cathode</th>
<th>OCV [V]</th>
<th>Power density [mWcm⁻²]</th>
<th>Temp [°C]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsuchiya et al. (2011) [855]</td>
<td>Si wafer</td>
<td>Pt</td>
<td>YSZ (SP/54nm)</td>
<td>LSCF</td>
<td>0.75</td>
<td>155</td>
<td>510</td>
</tr>
<tr>
<td>J. An et al. (2013) [849]</td>
<td>Si wafer</td>
<td>Pt</td>
<td>YDC/YSZ (ALD/60nm)</td>
<td>Pt</td>
<td></td>
<td>1300</td>
<td>450</td>
</tr>
<tr>
<td>Noh et al. (2014) [854]</td>
<td>Ni-YSZ</td>
<td>Ni-YSZ</td>
<td>YSZ/GDC (PLD/600nm)</td>
<td>LSC-GDC/LSC</td>
<td>1.1</td>
<td>588</td>
<td>500</td>
</tr>
<tr>
<td>K. Kerman et al. (2015) [856]</td>
<td>Si wafer</td>
<td>Pt</td>
<td>YSZ (SP/100nm)</td>
<td>Pt</td>
<td>0.8–1</td>
<td>1037</td>
<td>500</td>
</tr>
<tr>
<td>Jong Dae Baek et al. (2016) [850]</td>
<td>Si wafer</td>
<td>Pt</td>
<td>YSZ (PLD/80nm)</td>
<td>Pt</td>
<td>1.04</td>
<td>317</td>
<td>400</td>
</tr>
<tr>
<td>K. J. Kim et al. (2016) [857]</td>
<td>Si wafer</td>
<td>Pt</td>
<td>YSZ (PLD/2lm)</td>
<td>LSC</td>
<td>1.0</td>
<td>110</td>
<td>570</td>
</tr>
<tr>
<td>Jong Dae Baek et al. (2017) [858]</td>
<td>Si wafer</td>
<td>Pt</td>
<td>YSZ (ALD/10nm)</td>
<td>Pt</td>
<td>1.05</td>
<td>70.6</td>
<td>400</td>
</tr>
</tbody>
</table>

of great interest for developing future nanoionic structures for cathodes, anodes and electrolytes which are the building blocks of μ-SOFC. Accordingly, the future research and application of nanoionics for μ-SOFC of these artificial structures will be primarily driven by a number of factors, namely: (a) developing mesoporous nanoionics, (b) improving the cell performance by tuning the dopant distribution and the thickness of the nanoionic heterostructures, (c) controlling space charge layers, (d) controlling local strain by lattice mismatch or sequence of heterostructures layers and (e) integrating catalysts in the cell for reforming the fuel which can be adjusted by species, size, and the distribution of the catalysts.

From a practical point of view, the technical challenges which need to be resolved are: (a) high crystalline quality of epitaxial thin films on a large area, (b) synergy between deposition techniques together with silicon technologies and (c) high efficient patterning techniques. The potential of the pulsed laser deposition (PLD) technique as emerging deposition technique for such applications should not be underestimated. Nowadays there are PLD systems that can produce films up to 200 mm in diameters with high reproducibility and uniformity.

Finally, in addition to the μ-SOFCs discussed in this paper, the nanoionic network and nanocatalyst open further research opportunities for their potential applications in catalysts, photocatalyst, sensors and electrode materials, especially for those subjected to high/medium temperatures in harsh environments. There is plenty of room for using nanoionics and ionotronics in oxide materials and the future challenges once solved, provide a road towards improving existing and emerging applications. So the future looks bright!

Acknowledgment

The authors would like to thank Prof. Vincenzo Esposito and Dr. Dennis Valbjørn Christensen for their valuable comments and suggestions as well as for reading the manuscript. This work is part of the BioWings project which has received funding from the European Union’s Horizon 2020 under the Future and Emerging Technologies (FET) programme with a grant agreement No. 801267.

Piezo-MEMS for vibration-based energy harvesting

Minh D. Nguyen, Evert Houwman, Gertjan Koster, Guus Rijnders*

MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands

*Corresponding author.
E-mail: a.j.h.m.rijnders@utwente.nl (G. Rijnders).

Abstract

Vibration-based piezoelectric energy harvesting, which converts the kinetic energy from mechanical vibration into electrical energy, is a promising technique to supply unending sources for low power electronics as wireless sensor nodes. The output powers of harvesters mainly depend on the piezoelectric coefficients which can be tuned by the piezoelectric film quality and film growth. Up to now, various deposition techniques and deposition conditions have been investigated to improve piezoelectric coefficients. In this paper, we give a brief overview on the optimization of orientation and microstructure of lead zirconate titanate (PZT) thin films using pulsed laser deposition in order to tune the longitudinal piezoelectric $d_{33}$ coefficient and transverse piezoelectric $d_{31}$ coefficient for the $d_{33}$- and $d_{31}$-mode piezoelectric energy harvesters. We also make a brief comparison with lead-free thin films as future replacement materials of PZT.

Keywords: Energy harvesters; Vibration modes; Piezoelectric films; Buffer layers; Deposition conditions

1. Introduction

Piezoelectric thin films have been used in a variety of applications in the microelectronics industry for example in sensors and MEMS piezoelectric energy harvesting (PEH) for a micro electricity generator. Vibration-based piezoelectric energy harvesters, that convert the kinetic energy from vibration into electrical energy, have received increasing attention as a potential power source for microelectronics because of its simplicity of design, fabrication and operation. Energy harvesters have been considered as one of the most important technologies in green and sustainable energy science today. Moreover, energy harvesters provide endless sources of energy for low-power electronics such as wireless sensor nodes where replacement of batteries is not practically possible [859].

In a conventional structure, a PEH consists of a thin-layer piezoelectric material sandwiched between top and bottom electrodes on the passive elastic layer, as shown in Fig. 57(a). This is called the $d_{31}$-mode structure. Another possible type of PEH was introduced recently with interdigitated (IDT) electrodes, as indicated in Fig. 57(b). This type is called a $d_{33}$-mode structure. Good reviews of the IDT or the $d_{31}$-mode structure were given by Hagood et al. [860] and Zhang et al. [861]. More recent works has been covered by Hong [862], and others [863,864]. It is well known that the value of the $d_{33}$ piezoelectric coefficient (e.g. lead
zirconate titanate, PZT) is about two times higher than that of the \( d_{31} \) coefficient, therefore the output voltage is expected to be double for the devices based on the \( d_{33} \)-mode in polycrystalline films.

In most previous studies on piezoelectric energy harvesters based on the \( d_{33} \) and \( d_{31} \)-modes, the PZT films, for which the magnitude of \( d_{33} \) is about twice that of \( d_{31} \), were deposited by chemical spin-coating sol-gel techniques, on Pt/Ti/SiO\(_2\)/Si (Pt/Si, for \( d_{33} \)-mode devices) and ZrO\(_2\)/Si (for \( d_{31} \)-mode devices) [861,864,865]. However, the ratio of \( d_{33}/d_{31} \) in PZT films can be tuned in the range from 3.5 to 0.9 by changing the deposition technique (such as pulsed laser deposition), type of film growth (epitaxial and textured) and the deposition conditions [866,867]. In this paper, we present a brief overview on the microstructure and crystallization of Pb(Zr\(_{0.52}\)Ti\(_{0.48}\))O\(_3\) (PZT) thin films, as a function of deposition conditions and buffer layers, used in energy harvesting devices.

2. Current state and future prospects

Currently, most reported piezoelectric energy harvesters utilize the vibration of a cantilever beam structure at its resonance frequency. These novel piezoelectric cantilever beams consist of a piezoelectric film (such as PZT), which are fabricated by chemical solution deposition (such as sol-gel spin-coating) on bottom electrodes (such as Pt for \( d_{33} \)-mode devices) or template layers (such as ZrO\(_2\) and TiO\(_2\) for \( d_{31} \)-mode devices) on a supporting elastic beam (such as Si) [861,864,865,868]. However, the ability to convert energy in these devices is still low due to the low piezoelectric coef-ficients typical for the sol-gel polycrystalline PZT films [869].

Recently, improved piezoelectric coefficients have been obtained in epitaxial and highly textured PZT films deposited by pulsed laser deposition (PLD). Moreover, the ratio of \( d_{33}/d_{31} \) in these PZT films can be tuned by changing the crystalline orientation and microstructure [866]. By using optimized PLD conditions, the 1-μm-thick polycrystalline PZT films grown on Pt/Si have a \( d_{33}/d_{31} \) ratio of 1.55 (\( d_{33} = 124 \) pm/V, \( d_{31} = -80 \) pm/V). An even more reduced \( d_{33}/d_{31} \) ratio is observed in the (110)-oriented PZT films grown on SrRuO\(_3\)/Ti\(_{0.87}\)O\(_{2}\)-nanosheet/Si with a columnar structure (\( d_{33} = 122 \) pm/V, \( d_{31} = -86 \) pm/V, \( d_{33}/d_{31} = 1.42 \)), and in the highly textured (001)-oriented PZT films grown on SrRuO\(_3\)/CaNb\(_2\)O\(_{10}\)-nanosheet/Si with a dense film structure (\( d_{33} = 110 \) pm/V, \( d_{31} = -120 \) pm/V, \( d_{33}/d_{31} = 0.92 \)) [866]. Operating piezoelectric elements in the \( d_{33} \)-mode is more advantageous than those in the \( d_{31} \)-mode for MEMS-scale polycrystalline PZT harvesters [870,871], but the \( d_{31} \)-mode energy harvesters based on the recently developed epitaxial/ textured PZT films seem more useful.

The most important advantage of using nanosheets as a template layer for the growth of PZT films on Si, which are deposited using the Langmuir-Blodgett method, is the reduction in growth temperature. The growth of common buffer layers on Si generally requires a very high thermal budget (750–800°C for YSZ and CoO\(_2\)), whereas the process for the growth of SrTiO\(_3\) on Si is performed by molecular beam epitaxy and is very delicate and slow. Highly textured growth of PZT films on Si is currently achieved by using oxide nanosheets as a template layer. The growth of nanosheets with full in-plane orien-tation may be a further route for the improvement of energy harvesters based on epitaxial PZT films.

In practical commercial applications, the manufacturing process of energy harvesters based on piezoelectric cantilevers is very important. In terms of the fabrication process, the \( d_{33} \)-mode harvester is much simpler than the \( d_{31} \)-mode harvester because the \( d_{33} \)-mode harvester eliminates the requirement of a bottom electrode, thus reducing the number of photomasks needed and providing the possibility to generate high strain at low voltages [872]. Moreover, the wire-bonding process of contact-pads is also easier for a \( d_{33} \)-mode device. In order to enhance the \( d_{33} \) coefficient in PZT films, we have modified the microstructure in the PZT films by changing the pulse rate during the PLD process [867]. The results indicated that less densely packed columnar grains in the film deposited at high pulse rate of 50 Hz give rise to a significantly higher \( d_{33} \) value (\( d_{33} = 305 \) pm/V, \( d_{31} = -86 \) pm/V, \( d_{33}/d_{31} = 3.54 \)) than that of film deposited at low pulse rate of 10 kHz with a denser columnar structure (\( d_{33} = 192 \) pm/V, \( d_{31} = -120 \) pm/V, \( d_{33}/d_{31} = 1.60 \)) for 2-μm-thick PZT films grown on Pt/Si. For the \( d_{33} \)-mode harvester, the highly textured (001)-oriented PZT film can be directly grown on Ca\(_3\)Nb\(_2\)O\(_{10}\)-nanosheet/Si (\( d_{33} = 284 \) pm/V for 2-μm-thick PZT film deposited at 50 Hz) [873]. The high \( d_{33} \) value obtained in these films, even higher those of the respective bulk PZT ceramics (\( d_{33} = 223 \) pm/V), is remarkable and good news for \( d_{31} \)-mode harvester. This is a particularly important feature leading to experimentally observed enhancement of the performance of \( d_{33} \)-mode harvesters with high out-put voltage/power due to the large amplitude vibration of cantilever beams in comparison with those based on polycrystalline PZT films grown on ZrO\(_2\)/Si [874]. Of course, in a working device, besides the piezoelectric coefficients, the coupling coefficients and mechanical quality factor are of importance, in particular when operating the device at resonance. Depending on the configuration, these parameters need additional optimization which might adversely affect the piezoelectric displacement of devices.

During the last decades, lead-free piezoelectric materials have been investigated for powering medical implants and the application of energy harvesters in the human body that should offer comparable piezolectric properties to that of PZT materials [874–876]. Among all the lead-free piezoelectric materials, much attention has been paid to the potassium sodium niobate (KNN) materials, since Saito et al. made the breakthrough in Li-, Sb- and Ta- co-doped KNN [877]. The \( d_{33} \) values can reached up to 400–416 pm/V in the doped KNN ceramics, while pure KNN ceramics suffering a relative low \( d_{33} \) around 160 pm/V [877–879]. The experimental results indicated that the pure KNN thin films have strongly reduced \( d_{31} \) and \( d_{33} \) values (\( d_{33} = 58 \) pm/V, \( d_{31} = -42 \) pm/V) compared with PZT thin films in the same device configuration (\( d_{33} = 116 \) pm/V, \( d_{31} = -97 \) pm/V, both 750-nm-thick films were grown on Pt/Si using PLD) [880]. To compare the output power of PZT films and KNN films based harvesters, Kanno fabricated \( d_{31} \)-mode energy harvesters with the same dimensions and tip mass in a cantilever structure. The experimental results showed that the output power in the PZT harvesters (6.7 μW) was much higher than that in the KNN har-vesters (1.6 μW) [881].

Future developments should be focussed on the fabrication process of high quality KNN lead-free piezoelectric materials in thin film form with high \( d_{33} \) or high \( d_{31} \) values, as a potentially attractive alternatives to PZT. This is ongoing research. For resonant devices the focus should also be on relation between the damping and the thin film properties used in the devices. It is expected that these films will be utilized in MEMS energy harvesters in the near future. For example the KNN film quality and enhanced piezoelectric properties can probably be achieved by changing the deposition conditions to prevent the volatility of the alkali components and/or by using doped KNN films [882,883].

3. Concluding remarks

Recent advances in the fabrication of piezoelectric thin films and their integration into the harvesting structures will make battery-less autonomous sensors systems and networks more realistic. For the practical application, the piezoelectric coefficients (\( d_{33} \) and \( d_{31} \)) have been tuned in cantilever-type piezoelectric \( d_{33} \) and \( d_{31} \)-mode energy harvesters by controlling the growth, orientation and microstructure of thin films. Large \( d_{31} \) coefficients can be obtained in the films with highly textured/epitaxial growth and a dense structure, whereas giant \( d_{33} \) coefficients found in the films with less densely packed columnar grains are recent developments made toward improvement of electro-mechanical energy conversion of \( d_{33} \)-mode piezoelectric harvesters. Lead-free materials need further optimization to compete with PZT materials but are expected to become real future solutions.
Gallium oxide for power electronics
Shizuo Fujita*
Kyoto University, Katsura, Kyoto 615-8520, Japan
*Corresponding author.
E-mail: fujitasz@kuee.kyoto-u.ac.jp

Abstract
Gallium oxide (Ga2O3) is one of the most promising examples of wide-bandgap oxide materials for use in power electronics. Sup-ported by its ultra-wide bandgap, low-cost epitaxy, and bandgap engineering, in addition to a device process technology that is compatible with conventional device materials, Ga2O3 devices can compete with or complement silicon carbide (SiC) and gallium nitride (GaN) devices, especially in home appliances and in mid- to high-voltage application areas. The various crystal phases of Ga2O3 also offer opportunities for unique devices. The market of Ga2O3 devices is predicted to be 38% of SiC and GaN power devices in 2025.

Keywords: Gallium oxide; Polymorphs; Meta-stable materials; SiC and GaN; Epitaxial growth

1. Introduction

The performance characteristics of power devices, such as the breakdown voltage and on-resistance, tends to be enhanced by increases in the bandgap of the semiconductor materials used. Low cost is obviously an important key for widespread commercial use of such devices. From these points of view, gallium oxide (Ga2O3) has become recognized as one of the most promising semiconductor materials following silicon carbide (SiC) and gallium nitride (GaN), whose bandgaps are 3.3–3.4 eV, owing to its wider bandgap (about 5 eV) and low-cost growth process made possible by its use of safe sources of oxygen.

Ga2O3 has at least five different phases, which are named α, β, γ, δ and ε [884–886], as shown in Table 8. Of these, β-Ga2O3 is the most stable. It is noteworthy that β-Ga2O3 bulk substrates have been grown by conventional solution methods [887] in contrast to SiC and GaN that need dedicated technology for bulk growths. This provides a marked advantage because device research can be started with homoepitaxy from the earliest stage. On the other hand, the crystal structure of β-Ga2O3, i.e., b-gallia structure, is characteristic to β-Ga2O3 and is rarely seen in other materials. This is against the formation of a variety of alloy materials with different bandgaps, as well as heterostructures. α-Ga2O3 exists in the corundum structure, which is common to many materials including α-Al2O3 and α-In2O3, thus allowing a wide range of bandgap engineering. However, issues lie in how to grow high-quality metastable α-Ga2O3 by heteroepitaxy.

Current device-oriented research is focused primarily on β- and α-Ga2O3. In this paper, the author reports on the up-to-date evolution and prospects of Ga2O3 electronics, focusing on β- and α-Ga2O3. It should be noted that Ga2O3 films of other polymorphs are also attracted increasing interest and may be targets of device-oriented research soon. ε-Ga2O3 has a hexagonal structure that possesses ferroelectric properties [888–890], which is promising for forming highlyconcentrated two-dimensional electron gas (2DEG) at the heterointerface like ZnMgO/ZnO and AlGaN/GaN. It is suggested that γ- Ga2O3 is attractive for spintronic applications [891]. Unique properties and applications will be brought forth by Ga2O3 depending on the characteristic phases.

Table 8
Crystal structure of Ga2O3.

<table>
<thead>
<tr>
<th>Phases</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>δ</th>
<th>ε</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal stability</td>
<td>Metastable</td>
<td>Stable</td>
<td>Metastable</td>
<td>Metastable</td>
<td>Metastable</td>
</tr>
<tr>
<td>Crystal structure</td>
<td>Rhombohedral corundum</td>
<td>Monoclinic β-gallia</td>
<td>Cubic defective spinel</td>
<td>Cubic bixbyte</td>
<td>Hexagonal</td>
</tr>
<tr>
<td>Examples of compounds of the same structure</td>
<td>α-Al2O3</td>
<td>MgAl2O4</td>
<td>ε-In2O3</td>
<td>GaN</td>
<td></td>
</tr>
</tbody>
</table>
Fundamental properties of β- and α-Ga2O3.

Table 9 compares the fundamental properties and the up-to-date achievements for β-Ga2O3 and α-Ga2O3 thin films. Devices of β-Ga2O3 are made by ideal homoepitaxy, with which high-quality β-Ga2O3 can be used as active regions. α-Ga2O3 on the other hand, is grown on sapphire substrates by heteroepitaxy. Therefore, we cannot escape from dislocation defects in α-Ga2O3 layers. In order to guarantee low leakage current, high endurance, and high reliability, issues lie in how we can reduce and/or inactivate the defects.

Owing to the expectable high-quality β-Ga2O3 layers grown by homoepitaxy and their high breakdown strength, β-Ga2O3 will be suitable for devices operating at ultra-high voltages that are even higher than that of SiC devices used in power grids and railways. For SiC, transitions among polytypes during high-current operations might not be neglected, but β-Ga2O3 is quite stable and may not be free from the instability of crystal structure. Automotive applications are among the major targets for power devices. In such cases, cost is an important factor in ensuring that power devices can widely be used in market. SiC devices have been becoming inexpensive year-by-year, but might not become markedly cheaper in the future. The cost of devices seems to be bottle-necked by the cost of the substrates, which may be saturated as far as substrates that are fabricated by the sublimation method. β-Ga2O3 substrates are currently far more expensive than SiC substrates but may become cheaper judging from their solution-based growth technology. If so, automotive applications can be a potential target for β-Ga2O3 devices.

For further reducing the device cost, α-Ga2O3 devices will be candidates. On January 4, 2018, Denso Co., an automotive supplier, and FLOSFIA, Inc. [909], a developer of α-Ga2O3 devices, issued a press-release stating that they agreed to co-develop α-Ga2O3 power devices for next generation automobiles [910]. The devices will be developed while taking into consideration the balance among cost, performance, and reliability.

It is noteworthy that FLOSFIA, Inc. claims that α-Ga2O3 devices may become even cheaper than Si devices because of advances in inexpensive sapphire substrates and growth technology, as well as small device sizes. This encourages the idea that α-Ga2O3 devices are promising candidates for use in home appliances, which would contribute markedly to energy-saving.

Defects in α-Ga2O3, generated by heteroepitaxy is a severe issue for devices that require high levels of endurance and reliability. We have already confirmed that such defects can be reduced by introducing multilayer buffer layers and by epitaxial layer overgrowth (ELO). Fig. 59 shows a cross-sectional transmission electron microscope (TEM) image of α-Ga2O3 on sapphire with quasi-alloy buffer layers [911]. The dislocation density was reduced from the order of $10^{10} \text{cm}^{-2}$ to $10^{8} \text{cm}^{-2}$ by the introduction of such the buffer layers, and the use of ELO was also found to be markedly effective for reducing defect levels. By using these sophisticated technology, we can expect that the defect density is further reduced, thereby contributing to the evolution of α-Ga2O3 devices.

β-Ga2O3 RF devices are progressing remarkable. The competitors are GaN devices, but β-Ga2O3 devices possess advantages of high-quality crystal production via homoepitaxy compared to those on Si.

4. Prospective power devices in the future

Table 9 compares the fundamental properties and the up-to-date achievements for β-Ga2O3 and α-Ga2O3 thin films. Devices of β-Ga2O3 are made by ideal homoepitaxy, with which high-quality β-Ga2O3 can be used as active regions. α-Ga2O3 on the other hand, is grown on sapphire substrates by heteroepitaxy. Therefore, we cannot escape from dislocation defects in α-Ga2O3 layers. In order to guarantee low leakage current, high endurance, and high reliability, issues lie in how we can reduce and/or inactivate the defects.

Owing to the expectable high-quality β-Ga2O3 layers grown by homoepitaxy and their high breakdown strength, β-Ga2O3 will be suitable for devices operating at ultra-high voltages that are even higher than that of SiC devices used in power grids and railways. For SiC, transitions among polytypes during high-current operations might not be neglected, but β-Ga2O3 is quite stable and may not be free from the instability of crystal structure. Automotive applications are among the major targets for power devices. In such cases, cost is an important factor in ensuring that power devices can widely be used in market. SiC devices have been becoming inexpensive year-by-year, but might not become markedly cheaper in the future. The cost of devices seems to be bottle-necked by the cost of the substrates, which may be saturated as far as substrates that are fabricated by the sublimation method. β-Ga2O3 substrates are currently far more expensive than SiC substrates but may become cheaper judging from their solution-based growth technology. If so, automotive applications can be a potential target for β-Ga2O3 devices. For further reducing the device cost, α-Ga2O3 devices will be candidates. On January 4, 2018, Denso Co., an automotive supplier, and FLOSFIA, Inc. [909], a developer of α-Ga2O3 devices, issued a press-release stating that they agreed to co-develop α-Ga2O3 power devices for next generation automobiles [910]. The devices will be developed while taking into consideration the balance among cost, performance, and reliability.

It is noteworthy that FLOSFIA, Inc. claims that α-Ga2O3 devices may become even cheaper than Si devices because of advances in inexpensive sapphire substrates and growth technology, as well as small device sizes. This encourages the idea that α-Ga2O3 devices are promising candidates for use in home appliances, which would contribute markedly to energy-saving.

Defects in α-Ga2O3, generated by heteroepitaxy is a severe issue for devices that require high levels of endurance and reliability. We have already confirmed that such defects can be reduced by introducing multilayer buffer layers and by epitaxial layer overgrowth (ELO). Fig. 59 shows a cross-sectional transmission electron microscope (TEM) image of α-Ga2O3 on sapphire with quasi-alloy buffer layers [911]. The dislocation density was reduced from the order of $10^{10} \text{cm}^{-2}$ to $10^{8} \text{cm}^{-2}$ by the introduction of such the buffer layers, and the use of ELO was also found to be markedly effective for reducing defect levels. By using these sophisticated technology, we can expect that the defect density is further reduced, thereby contributing to the evolution of α-Ga2O3 devices.

β-Ga2O3 RF devices are progressing remarkable. The competitors are GaN devices, but β-Ga2O3 devices possess advantages of high-quality crystal production via homoepitaxy compared to those on Si.

### Table 9

<table>
<thead>
<tr>
<th>Phases</th>
<th>β-Ga2O3</th>
<th>α-Ga2O3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal structure</td>
<td>Monoclinic</td>
<td>Rhombohedral</td>
</tr>
<tr>
<td>Band gap</td>
<td>$4.4 - 4.9 \text{eV}$</td>
<td>$5.2 - 5.3 \text{eV}$</td>
</tr>
<tr>
<td>Substrate</td>
<td>β-Ga2O3 (n⁺, insulating)</td>
<td>α-(Al,Ga,In)2O3</td>
</tr>
<tr>
<td>Growth</td>
<td>Homoepitaxy</td>
<td>Heteroepitaxy</td>
</tr>
<tr>
<td>Growth temperature</td>
<td>&gt; 700°C</td>
<td>&lt; 600°C</td>
</tr>
<tr>
<td>Thermal stability</td>
<td>Stable</td>
<td>Metastable transition to β-Ga2O3 (^{b})</td>
</tr>
<tr>
<td>P-type material</td>
<td>Under development Mg, Zn, N, etc. may act as acceptors in β-Ga2O3</td>
<td>α-(Ir,Ga)2O3, α-(Ir,Ga)2O3, and other corundum-structured p-type oxides doping under development</td>
</tr>
<tr>
<td>Device structure</td>
<td>Lateral, vertical</td>
<td>Basically lateral vertical reported</td>
</tr>
<tr>
<td>Initial marketing strategy</td>
<td>Ultra-high voltage high temperature high radiation</td>
<td>Inexpensive devices low voltage (home appliances)</td>
</tr>
</tbody>
</table>

\(^{a}\) Recently we achieved the growth of α-Ga2O3 at 800°C by modifying precursors [907].

\(^{b}\) The transition temperature can be highered by introducing slight Al [908].
substrates, and of low cost compared to those on GaN substrates. The advancement of the Internet of things (IoT) society requires inexpensive, highly efficient, and low-power RF devices; qualities that are characteristic to Ga2O3 RF devices. High endurance against high temperature and radiation is also an advantage that will open applications of β-Ga2O3, and will open applications of β-Ga2O3 devices in extreme environment as well as power devices.

Besides Ga2O3, other wide bandgap oxides can be candidates for power device applications. One such example is α-In2O3, which shows world-wide market of Ga2O3 devices will reach one billion JPY in 2020. Marketing report published by Fuji Keizai Co, Japan [913] that shows the advantage of Ga2O3 devices over SiC power devices is predicted to be approximately 186 billion JPY in 2025, while that of Ga2O3 power devices can be expected to reach as much as 70 billion JPY (38% of SiC + GaN devices).

5. Concluding remarks

In conclusion, the authors would like to introduce a recent marketing report published by Fuji Keiza i Co, Japan [913] that shows the worldwide market of Ga2O3 devices will reach one billion JPY in 2020. It also predicts that the advantages of Ga2O3 devices over SiC power devices will become more enhanced from 2023, especially in mid- and high-voltage application areas above 600 V, particularly train infrastructures and power networks. It is interesting to note that the market of SiC and GaN power devices is predicted to be approximately 186 billion JPY in 2025, while that of Ga2O3 power devices can be expected to reach as much as 70 billion JPY (38% of SiC + GaN devices).

References

[27] More info about the Nomad Project can be found at the link: [https://www.nomad-coe.eu/].


S.J. May, C.R. Smith, J.W. Kim, E. Karapetrova, a Bhattacharya, P.J. Ryan, Control of octahedral rotations in (LaNiO3)1-x(SrMnO3)2 superlattices, Physical Review B 83 (2011) 081108(R).


S.J. May, C.R. Smith, J.W. Kim, E. Karapetrova, a Bhattacharya, P.J. Ryan, Control of octahedral rotations in (LaNiO3)1-x(SrMnO3)2 superlattices, Physical Review B 83 (2011) 081108(R).


D. Wesenberg, T. Liu, D. Balzar, M. Blü, J. Shi, Observation of magnon-mediated current drag in Pt/yttrium iron garnet/Pt(Ta) trilayers, Nat. Commun. 7 (2016) 10858, https://doi.org/10.1038/ncomms10858.


Y. Sun, F. Guo, S. Chen, S. Zhao, Improved ferroelectric and photovoltaic properties of Bi2O3-0.7Bi2O3-0.3MnO2, solid-solution films, Appl. Phys. Lett. 111 (2017) 253901.


