

Oxygen vacancy chain and conductive filament formation in hafnia

Kan-Hao Xue^{1,2,a)} and Xiang-Shui Miao^{1,2,b)}

¹*School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, China*

²*Wuhan National Laboratory for Optoelectronics, Wuhan 430074, China*

(Received 10 June 2017; accepted 11 August 2017; published online 6 October 2017)

The stability and aggregation mechanisms of oxygen vacancy chains are studied for hafnia using self-energy corrected density functional theory. While oxygen vacancies tend not to align along the *c*-axis of monoclinic HfO₂, oxygen vacancy chains along *a*-axis and *b*-axis are energetically favorable, with cohesive energies of 0.05 eV and 0.03 eV per vacancy, respectively. Nevertheless, with an increase of the cross section area, intensive oxygen vacancy chains become much more stable in hafnia, which yields phase separation into Hf-clusters and HfO₂. Compared with disperse single vacancy chains, intensive oxygen vacancy chains made of 4, 6, and 8 single vacancy chains are energetically more favorable by 0.17, 0.20, and 0.30 eV per oxygen vacancy, respectively. On the other hand, while a single oxygen vacancy chain exhibits a tiny electronic energy gap of around 0.5 eV, metallic conduction emerges for the intensive vacancy chain made of 8 single vacancy chains, which possesses a filament cross section area of $\sim 0.4 \text{ nm}^2$. This sets a lower area limit for Hf-cluster filaments from metallic conduction point of view, but in real hafnia resistive RAM devices the cross section area of the filaments can generally be much larger ($>5 \text{ nm}^2$) for the sake of energy minimization. Our work sets up a bridge between oxygen vacancy ordering and phase separation in hafnia, and shows a clear trend of filament stabilization with larger dimensions. The results could explain the threshold switching phenomenon in hafnia when a small AFM tip was used as the top electrode, as well as the undesired multimode operation in resistive RAM cells with 3 nm-thick hafnia. *Published by AIP Publishing.* [<http://dx.doi.org/10.1063/1.4989621>]

I. INTRODUCTION

Reproducible resistive switching phenomena have been widely utilized to achieve resistive random access memory¹ (RRAM), one of the leading technologies in advanced non-volatile memories. Binary oxides such as HfO_x and TaO_x are the most promising candidates for RRAMs due to their supreme technical compatibility to standard CMOS and outstanding performances.^{2,3} In particular, hafnia is the key ingredient of the state-of-the-art CMOS gate oxide^{4,5} and extremely high quality hafnia thin films can be obtained straightforwardly in CMOS flow-line. Moreover, HfO_x-based RRAM with over 10¹²-cycle write endurance has been demonstrated recently.⁶ Notwithstanding its technical advantages, the exact nature of the conductive filaments (CFs) in hafnia is still under debate. While it is widely accepted that oxygen vacancies play a significant role in the CF formation in hafnia,⁷ it is still uncertain whether the ultimate CF consists of ordered oxygen vacancies⁸ [oxygen vacancy chains], conductive sub-oxides,⁹ metal Hf (either in the *h.c.p.* form or in other metal cluster forms),^{10,11} or metal Hf with oxygen interstitials such as Hf₆O.¹² Identifying the exact compositions of the CFs as well as the CF formation mechanism in hafnia remains a highly challenging task.

Celano *et al.* utilized conductive atomic force microscopy (C-AFM) to investigate the nature of CF in bipolar Hf/HfO₂/TiN RRAM and directly observed a CF of a conical shape,¹³

with a cross section area of 38.9 nm² near the Hf (active) electrode but merely 7.8 nm² near the TiN (inert) electrode. Such conical shape can be attributed to the formation of high density oxygen vacancies near the active Hf electrode, and subsequent migration of these vacancies towards the inert electrode during the electroforming step. The CF was identified as a local region rich of oxygen vacancies. Unfortunately, it was not possible to investigate the structure of the CF in that experiment. In a later work¹⁴ they also observed an interesting phenomenon when small area (typically $3 \times 3 \text{ nm}^2$) C-AFM tips were used as the top electrode to switch the resistive state of hafnia. Under such situation, threshold switching commonly occurs, i.e., the CF becomes unstable when the number of oxygen vacancy defects is limited. Recently Li *et al.* has investigated the CF in hafnia through *in situ* electron holography together with electron energy loss spectrum, and also concluded that a continuous channel filler with oxygen vacancies accounts for the low resistance state of their hafnia-RRAM.¹⁵ Their work provides a direct evidence for the evolution of CFs in hafnia due to the motion of oxygen vacancies. Unfortunately, there is still a lack of understanding towards the composition and stoichiometry of the CFs.

On account of the difficulty in the experimental methods, *ab initio* calculation affords an effective way towards the composition and the formation mechanism of CFs in hafnia. Oxygen vacancy ordering has long been proposed as the trigger mechanism for CF formation in binary oxides such as TiO₂ and NiO.¹⁶ Kamiya *et al.* proposed a generalized model for the resistive switching in binary oxides based on the oxygen vacancy cohesion-isolation transition.¹⁷ They regard an

^{a)}Email: xkh@hust.edu.cn

^{b)}Email: miaoxs@hust.edu.cn

oxygen vacancy chain (VC) as the CF and proposed that charge injection/removal is a strong driving force for such transitions. While this model is highly universal for binary oxides, it did not investigate the stability of the VC and therefore the energetics for the CF formation. The experimentally observed CF, however, typically possesses a substantial cross-section over 5 nm^2 even at its narrowest end.¹³ In addition, the electronic structure of a single oxygen VC shows some defect bands that are either fully occupied or empty at zero temperature, though the energy gap is vanishingly small. This does not capture the metallic conduction and the positive temperature coefficient for the resistance of CFs. It is necessary to study VCs with a wide range of cross section areas for deeper understanding into the CF composition.

McKenna has pointed out,¹⁸ through first-principles and thermal dynamic calculations, that sub-stoichiometric HfO_x ($0.2 < x < 2$) would decompose into $\text{HfO}_{0.2}$ and HfO_2 , and thus proposed an $\text{HfO}_{0.2}$ filament model. He calculated the electronic structures of an $\text{HfO}_{0.2}$ filament embedded in HfO_2 dielectric, and confirmed the filamentary conduction picture. Furthermore, he identified the optimal stoichiometry of the initial HfO_x dielectric to be $1.5 < x < 1.75$, such that efficient nucleation and growth of stable Hf-rich clusters can be realized. Zhang *et al.* studied the phase diagram of the Hf-O system under pressures,¹² and confirmed the only two stable HfO_x ($x > 0$) phases at atmospheric pressure and room temperature are $P2_1/c$ HfO_2 and $R\bar{3}$ Hf_6O ($\sim\text{HfO}_{0.17}$). These two works list the approximate composition $\text{HfO}_{0.2}$ as the strong candidate for the CF in hafnia. Hence, there are at least two models for the CFs in hafnia: VCs inside HfO_2 and Hf-rich clusters close to the Hf_6O phase. *Ab initio* calculations show that oxygen di-vacancy pairs and multiple vacancies are energetically more favorable than isolated vacancies in HfO_2 , Si-doped HfO_2 and Ti-doped HfO_2 .^{9,19} It is reasonable to infer that the formation of VC is favorable at the beginning stage of electroforming. Yet, it is still unclear whether and how VCs can be transformed into Hf-rich clusters with approximately Hf_6O stoichiometry.

In this work, we extend the VC model of Kamiya *et al.* and study the VC aggregation effects through *ab initio* calculations, in addition to their electronic structures. We aim at answering the following questions:

- Why the stable CFs in hafnia are large in cross sections (typically 7.8 nm^2 even in their narrow ends)?
- Why small electrodes, and thus small possible filament cross section areas, can lead to threshold switching?
- What is the influence of local and global stoichiometry on the formation of CFs in hafnia?

To resolve these issues we carry out two series of studies successively. First, we set up 108-atom HfO_2 supercells (9 unit cells each containing 12 atoms) and study the cohesive effects of oxygen VCs, comparing the energies of either dispersed or aggregated VCs with various densities. For the aggregated case we shall identify the minimum size of the VC that leads to metallic conduction, i.e., VC becomes CF. Next, for the CF model we increase the size of the dielectric to study the influence of stoichiometry on electronic structures. We shall show that oxygen VCs tend to aggregate to

render phase separation, and metal Hf (possibly with a few O interstitials) is likely to be the composition of the CF.

II. COMPUTATIONAL METHOD

For our *ab initio* calculations we used density functional theory^{20,21} (DFT), as implemented in the plane-wave based Vienna *Ab initio* Simulation Package^{22,23} (VASP). The electrons considered as valence were: $5p$, $5d$ and $6s$ for Hf; $2s$ and $2p$ for O. Core electrons were approximated by projector augmented-wave pseudopotentials.^{24,25} The generalized gradient approximation (GGA) was used for the exchange-correlation energy, within the PBEsol functional,²⁶ which has been shown to result in very accurate structural parameters for solids. The plane wave kinetic energy cutoff was 500 eV for all cases, and sufficiently dense Monkhorst-Pack k -meshes²⁷ were utilized to sample the Brillouin zone. All hafnia and defective hafnia supercells were fully relaxed until the Hellmann-Feynman forces were below $0.02 \text{ eV}/\text{\AA}$. Unless in isolated oxygen vacancy and single oxygen VC studies, the lattice parameters were also optimized until the stress in each direction of the supercell is less than 500 MPa.

It is well known that DFT-GGA systematically underestimates the band gaps for semiconductors and insulators. As our supercell typically consists of more than 100 atoms, hybrid functionals²⁸ and the GW approximation²⁹ become impractical in their computational complexity. Hence, we shall implement a self-energy corrected GGA-1/2 method, which was proposed by Ferreira *et al.* in 2008 (together with local density approximation (LDA)-1/2)^{30,31} and is effective in resolving the band gap problem for semiconductors.³²⁻³⁹ The method corrects the spurious electron self-interaction term in local density approximation (LDA) or GGA using the Slater half-occupation technique, while keeping the computational cost at the same level of plain LDA/GGA. According to LDA-1/2 and GGA-1/2, the well-known problem of band gap underestimation in LDA and GGA stems from their neglecting the self-energy of the hole in the valence band. Instead of calculating the self-energy of the hole, which is a formidable task, GGA-1/2 attaches the trimmed self-energy potentials to the anions in real space. Specifically, in our work electronic structure calculations were carried out as a separate step after structural relaxation and static energy calculation. The self-energy potentials were derived from atomic calculations using a modified ATOM code (supplied with the Siesta simulation package⁴⁰) where we included the PBEsol functional. The trimmed self-energy potential for oxygen, i.e., $V[\text{O}: 2s^2 2p^4] - V[\text{O}: 2s^2 2p^{3.5}]$ with a specific cutoff radius, was added to the oxygen atoms as an external potential. A series of self-consistent runs were carried out to fix the optimal cutoff radius for the self-energy potential, which ought to maximize the band gap of HfO_2 according to the principles of GGA-1/2. Electronic structures were obtained with this optimal cutoff radius only, and no empirical parameters were involved.

III. RESULTS AND DISCUSSIONS

A. Bulk HfO_2 and isolated oxygen vacancy

We studied two phases of bulk HfO_2 : monoclinic HfO_2 (m - HfO_2) and cubic HfO_2 (c - HfO_2). The former is the ground state phase of HfO_2 at room temperature and atmospheric

TABLE I. Calculated structural parameters and band gaps for *c*-HfO₂ and *m*-HfO₂, with comparison to experiments.

	Cubic			Monoclinic		
	GGA	GGA-1/2	Experimental	GGA	GGA-1/2	Experimental
a (Å)	5.022	Same as GGA	5.09 (Ref. 43)	5.082	Same as GGA	5.117 (Ref. 44)
b (Å)				5.156		5.175
c (Å)				5.252		5.292
β angle (°)	90		90	99.68		99.22 (Ref. 44)
Band gap (eV)	3.76	6.05	N. A. ^a	3.97	5.98	5.68 (Ref. 41) 5.9 eV (Ref. 42)

^aExperimental band gap value for cubic HfO₂ was rarely reported. A Y-stabilized hafnia study reveals a band gap of 5.8 eV.⁴⁵

pressure, while the latter is the simplest phase with high symmetry. Their optimized lattice parameters are compared with experimental values in Table I. The GGA-1/2 band diagrams in Fig. 1 show band gaps of 5.96 eV for *m*-HfO₂ and 6.05 eV for *c*-HfO₂. In both phases the optimal oxygen self-energy potential cutoff radii are 2.7 bohr. The experimental band gap for *m*-HfO₂ is typically 5.68 to 5.9 eV,^{41,42} in good agreement with our GGA-1/2 band gap.

There is only one oxygen site in *c*-HfO₂ but two distinct oxygen sites in *m*-HfO₂. To study the formation energy of isolated oxygen vacancy point defects, we set up $2 \times 2 \times 2$ supercells (96 atoms in total) for these two phases and define the vacancy formation energy as

$$E_{\text{form}} = E_{\text{defective}} - E_{\text{perfect}} + \mu_{\text{O}}, \quad (1)$$

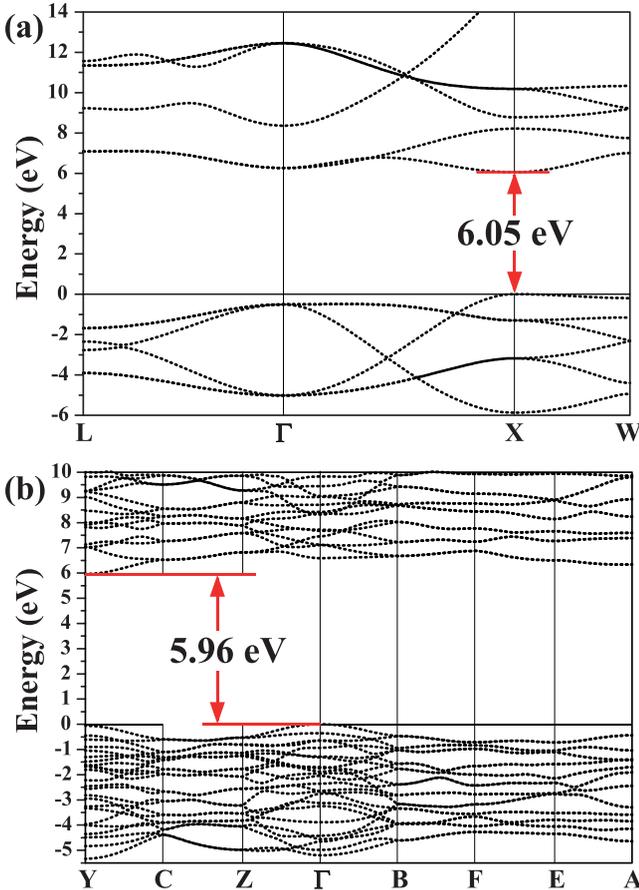


FIG. 1. The band diagrams for (a) cubic HfO₂ and (b) monoclinic HfO₂, calculated using GGA-1/2.

where $E_{\text{defective}}$ and E_{perfect} are the energies of the supercells with and without a single oxygen vacancy, respectively; μ_{O} is the chemical potential of oxygen. There are multiple choices for the value of μ_{O} , depending on the O-rich or O-poor environments. As we only care about the relative stability of defects, we choose μ_{O} as that in the O₂ gas form. However, it is well known that DFT-GGA severely overestimates the binding energy for O₂, therefore an energy correction term is usually applied to the O₂ molecule energy as well. In this case we have

$$\mu_{\text{O}} = \frac{1}{2}(E_{\text{O}_2} + E_{\text{corr}}), \quad (2)$$

where a proper E_{corr} ought to yield a correct binding energy for O₂. Using GGA-PBESol we find that $E_{\text{corr}} = 1.37$ eV rectifies the binding energy of O₂, from -6.60 eV to the experimental value -5.23 eV.⁴⁶ Compared with the GGA-PBE results by Wang *et al.*⁴⁷ (-6.02 eV binding energy for O₂ before correction), our PBESol binding energy of O₂ is even more overestimated, which is reasonable since the bond lengths predicted with PBESol are generally smaller than with PBE. It is therefore even more necessary to correct the DFT-energy of O₂ with the PBESol functional.

With the above definition, we calculated the formation energy of a neutral oxygen vacancy as 6.95 eV in *c*-HfO₂. In *m*-HfO₂, the III-coordinate oxygen vacancy (V_{O-III}) has $E_{\text{form}} = 7.37$ eV, while the IV-coordinate oxygen vacancy (V_{O-IV}) has $E_{\text{form}} = 7.27$ eV. The fact that V_{O-IV} is slightly easier to form is consistent with previous studies.^{48,49} In isolated oxygen vacancy studies we did not alter the lattice parameters of HfO₂, but only allowed the relaxation of atomic coordinates within the supercell, which is a common way to keep the nature of isolated point defects.

B. Single oxygen vacancy chains in *m*-HfO₂

We define the single oxygen VC as a hollow path enclosed by Hf atoms where no oxygen atoms can be found along the path. In *c*-HfO₂ a single VC is shown in Fig. 2(a). In *m*-HfO₂, the single oxygen VCs along *a*-, *b*-, and *c*-axes are schematically shown in Figs. 2(b)–2(d). Starting from a 12-atom HfO₂ unit cell, all these model supercells are enlarged to 3×3 size over the plane that is perpendicular to the aligning direction of the VCs. For example, in the *b*-axis VC model we used a $3 \times 1 \times 3$ supercell, where the VCs are separated by around 1.5 nm from their nearest periodic images.

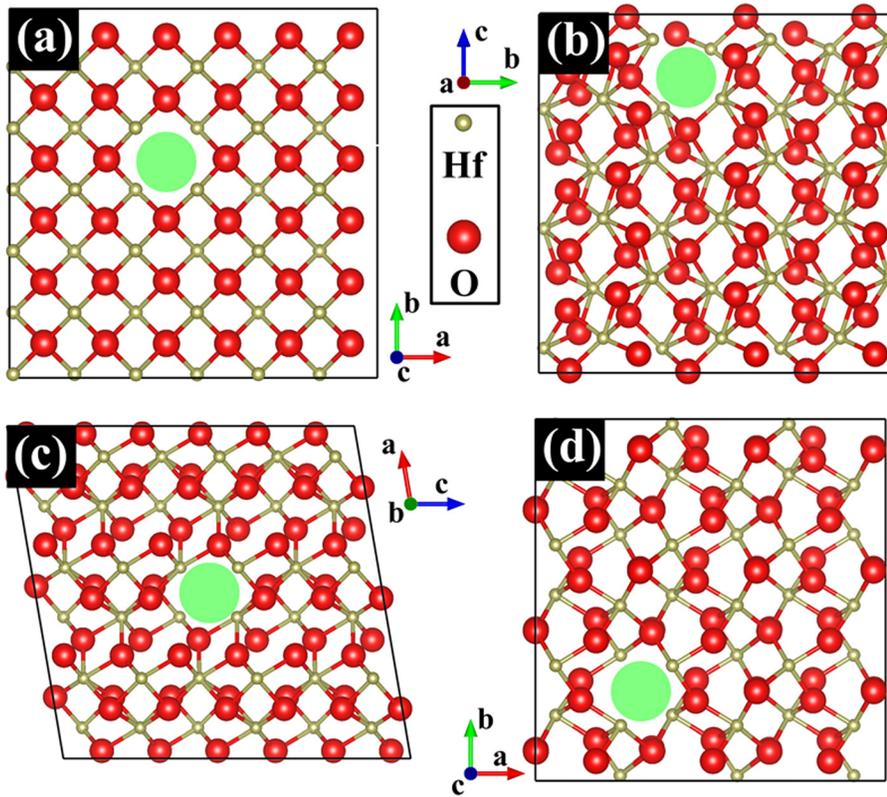


FIG. 2. Atomic structures of the single oxygen VCs in hafnia. (a) A VC in c -HfO₂; (b) A VC along a -axis in m -HfO₂; (c) A VC along b -axis in m -HfO₂; (d) A VC along c -axis in m -HfO₂. The big red balls represent O while the small brown balls represent Hf. The same convention for atomic illustration is used in all the following structural figures.

The formation energy of an oxygen VC can be defined as follows:

$$E_{\text{form}} = \frac{1}{n}(E_{\text{defective}} - E_{\text{perfect}}) + \mu_{\text{O}} \quad (3)$$

with the unit of eV per oxygen vacancy (denoted by eV/O for short), where n is the number of oxygen vacancies in one defective supercell. For single VCs, $n = 2$ always holds.

In the single VC study, we always set up a 108-atom supercell before introducing the VCs. During the structural relaxation we only optimized the internal atomic coordinates of the defective supercells. Nevertheless, upon allowing the supercell lattice parameters to be fully relaxed, we discovered almost no difference in the formation energies of single VCs. Hence, we fix the supercell lattice parameters as in the isolated oxygen vacancy study. In c -HfO₂ the supercell is $3 \times 3 \times 1$ and there is only one kind of single oxygen VC, as shown in Fig. 2(a). Under the definition above, its formation energy is 6.73 eV/O, lower than the 6.95 eV energy for oxygen vacancy point defect formation. Hence, oxygen vacancies tend to accumulate in c -HfO₂. In m -HfO₂ we find three most energetically favorable oxygen VCs, along a -, b - and c -axes, with formation energies of 7.22 eV/O, 7.24 eV/O and 7.30 eV/O, respectively. These values were obtained using $1 \times 3 \times 3$, $3 \times 1 \times 3$ and $3 \times 3 \times 1$ supercells where the corresponding oxygen atoms were stripped. The most probable chain is the one along a -axis, while the b -axis oxygen VC is also likely to emerge because its formation energy per O-vacancy is still slightly lower than the single V_O-IV case. However, the VC along c -axis is not energetically favorable. In m -HfO₂ we choose the b -axis VC as our focus, because even a larger number

of VCs in this configuration will not cause structural instability, while the VC along a -axis may lead to phase transition in the high density regime.

Figures 3(a) and 3(b) show the electron localization function (ELF) of the VC in c -HfO₂ and the b -axis VC of m -HfO₂. The red region indicates localized electron density (ELF~1), the green region represents delocalized electrons (ELF~0.5), and the blue region means the absence of electron (ELF~0). In both cases there are highly localized electrons observed at the vacancy sites, but they are connected through certain delocalized electrons in the green region, which implies that at least hopping conduction is possible through these VCs. In the c -HfO₂ case the continuous region of delocalized electrons is even more remarkable. From the band diagrams [Figs. 3(c) and 3(d)] one further finds that strong dispersion exists only along the vacancy chain direction, which is c -axis and b -axis for c -HfO₂ and m -HfO₂, respectively. Such phenomenon clearly originates from the interaction between neighboring oxygen vacancies along the VC, and is consistent with previous researches.⁵⁰ The interaction of oxygen vacancies along the a - b plane is, however, absent since their minimum distance is above 1.5 nm in our single VC models.

In principle, the working mechanism of HfO₂ RRAM is not always due to the formation of CFs. In certain highly nonlinear devices such as those designed for self-selected 3D integration,⁵¹ the current compliance is so low that CFs cannot be created. Hence, in certain HfO₂ RRAMs the hopping conduction and trap-assisted tunneling processes may dominate the ON-state. It is also possible that metallic conduction and hopping conduction coexist due to both extended and localized electron states.^{52,53} Nevertheless, our present work

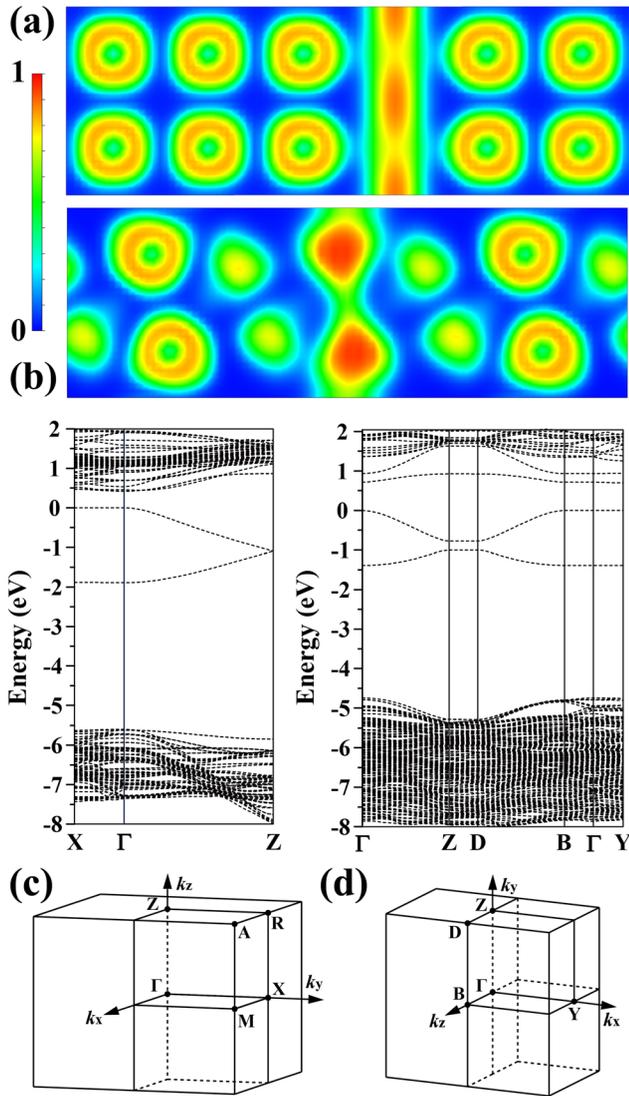


FIG. 3. (a) Electron localization function (ELF) for the single oxygen VC in *c*-HfO₂; (b) ELF for the single oxygen VC in *m*-HfO₂ along *b*-axis; (c) Electronic band diagram for the model in Fig. 2(a); (d) Electronic band diagram for the model in Fig. 2(c). Red, green and blue colors represent ELF = 1, ELF = 0.5 and ELF = 0, respectively.

mainly focuses on the composition and formation mechanism of metallic CFs in HfO₂, and in this sense we find that the electronic density of states analysis for all these VCs in Fig. 2 does not support a metallic conduction picture. As shown in Fig. 4(a), the defect bands induced by these VCs are fully occupied in *c*-HfO₂, though the energy gap to the conduction band is quite small (~ 0.5 eV). Note that even at room temperature, the presence of such a small gap still yields a semiconductor-like temperature coefficient for resistivity, due to the thermal excitation. In reality, however, such thermal excitation is usually absent in the ON-state of hafnia RRAMs.⁵⁴ In other words, our results do not support single VCs as the CF composition in hafnia RRAMs.

C. Multiple and intensive oxygen vacancy chains

To achieve metallic conduction, it is reasonable to require that the density of VCs should exceed a threshold, or the cross section area of the VC must be larger than a critical value. A

study on a series of multiple VCs in the same supercell-setting may give some hints on the criteria for metallic conduction. We choose oxygen VCs along *b*-axis in *m*-HfO₂ as our model system. The *m*-HfO₂ phase is preferred because *c*-HfO₂ is not the ground state, where the introduction of multiple CFs may cause a global phase transition, and the formation energies of VCs can lose their physical meaning accordingly. The $3 \times 1 \times 3$ *m*-HfO₂ supercell is sufficiently large such that we can study two family of VCs: one with dispersed single VCs (referred to as the “disperse model”) and the other with only one intensive VC per supercell (referred to as the “intensive model”), both with the same stoichiometry. Figures 5(a) and 5(b) demonstrates the fully relaxed structures of the disperse model and the intensive model both with the stoichiometry Hf₃₆O₆₄ (\sim HfO_{1.78}). The formation energies are 7.24 eV/O and 7.07 eV/O for the disperse model and the intensive model, respectively. The intensive model is energetically superior by 0.17 eV/O, revealing that oxygen vacancy cohesion is favorable in hafnia. To further verify this point, we constructed 4 more disperse and intensive VC models with Hf₃₆O₆₀ (\sim HfO_{1.67}) and Hf₃₆O₅₆ (\sim HfO_{1.56}) global stoichiometries, whose atomic structures are demonstrated in Figs. 5(c)–5(f). Their formation energies show an interesting monotonic trend. In Hf₃₆O₆₀, the intensive VC model shows even superior stability, whose formation energy (7.02 eV/O) is 0.20 eV/O lower than that of the disperse model (7.22 eV/O). Furthermore, in Hf₃₆O₅₆ such difference is enlarged to 0.30 eV/O (formation energies: 6.87 eV/O and 7.17 eV/O for the intensive and the disperse model, respectively). In other words, oxygen vacancy accumulation becomes even more energetically favorable when the reduced region (potential CF region) is larger.

Next we examine the electronic structures of all these multiple VC models. The density of states calculated using GGA-1/2 are plotted in Figs. 6(a)–6(f), corresponding to the models in Figs. 5(a)–5(f), respectively. The most prominent feature is that all the disperse models, even in the Hf₃₆O₅₆ stoichiometry, do not show metallic conduction, though their gaps can be as small as a few tenths of eV at zero temperature. For intensive models, however, we find that metallic conduction does emerge in Fig. 6(f) as the CF becomes sufficiently large (~ 0.4 nm² in cross section area). The insets in Figs. 6(d) and 6(f) clearly indicate a transition from a semiconductor to a metallic state. Hence, the model in Fig. 5(f) is both energetically very favorable and metallic, revealing that phase separation into Hf clusters (possibly with some O interstitials) and HfO₂ is a highly probable process during the electroforming in hafnia RRAMs. Note that our Hf filament in Fig. 5(f) is not in the *h.c.p.* structure, but its presence is still energetically favorable.

To verify that phase separation is the key towards metallic conduction, we further checked the valence of Hf atoms in these models through Bader charge analysis. Table II shows the charge value of several selected Hf atoms. In Fig. 5(a), the net charge values of Hf #1 (far from the vacancy sites) is +2.70, which can be regarded as a standard value for HfO₂. On the other hand, for the four Hf atoms surrounding the VC in Fig. 5(a), i.e., Hf #2 to #5, their average net charge is +2.05, lower than the standard Hf case. When it comes to the intensive model as in Fig. 5(b), however, the Hf atoms within the VC region show much reduced charge values than

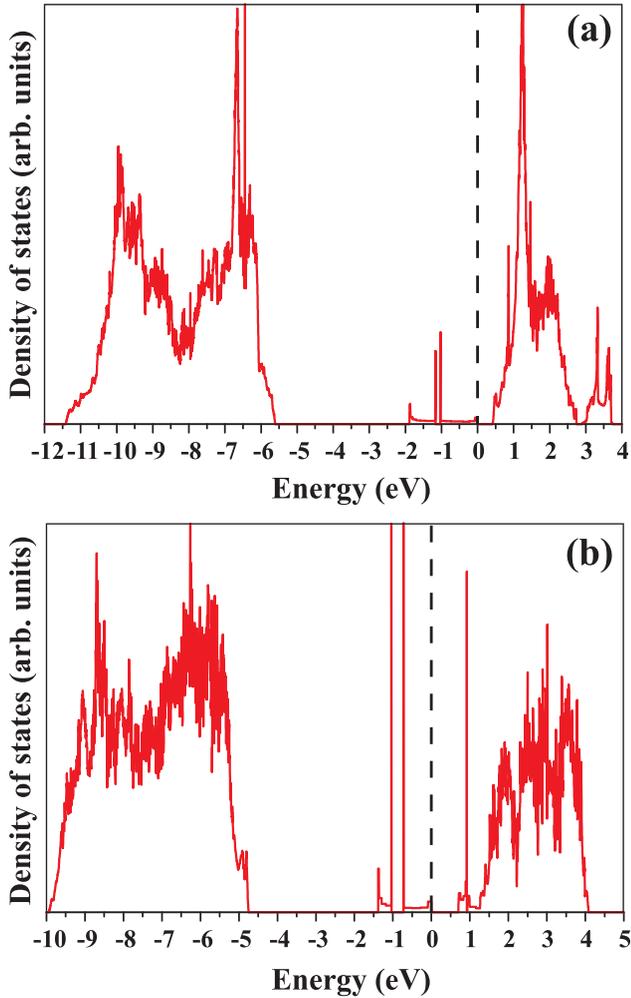


FIG. 4. Density of states of the single oxygen VC models: (a) VC in c -HfO₂; (b) VC in m -HfO₂ along b -axis. The vertical dashed lines at zero energy stand for the Fermi levels.

the disperse model. For example, Hf #6 is charged by +1.42, while Hf #7 residing at the center of the CF has a charge of merely +0.07, very close to metal Hf. For Hf #8 and #9 in Fig. 5(f), which are at the core of an even larger CF, their charge values according to Bader analysis are +0.28 and -0.17 . As a slightly negatively charged Hf atom is found, the Hf atoms in the core CF region are thus extremely different from the case of ionic Hf cation in HfO₂. With stronger CFs, there are more Hf atoms showing the behavior of charge neutrality. This reflects the phase separation phenomenon and the formation of metal Hf cluster.

We point out here that the disperse models, by definition, do not contain large CFs that stem from merged VCs, but consist of individual single VCs dispersed in the dielectric matrix. However, once the individual VCs meet each other, such as under thermal activation or high applied electric field, they are prone to merge into large CFs. This is exactly a possible route towards the ultimate phase-separation scenario.

D. Influence of global stoichiometry

The intensive VC model shown in Fig. 5(f) and its electronic structure in Fig. 6(f) demonstrate our typical

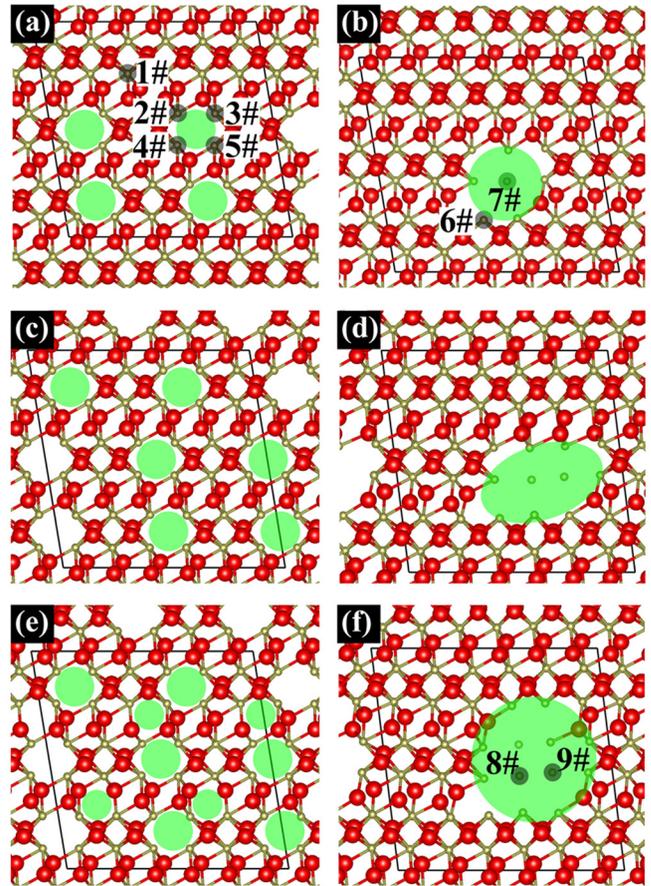


FIG. 5. Multiple VCs along b -axis with various global stoichiometries in m -HfO₂. (a) Disperse model in Hf₃₆O₆₄; (b) intensive model in Hf₃₆O₆₄; (c) disperse model in Hf₃₆O₆₀; (d) intensive model in Hf₃₆O₆₀; (e) disperse model in Hf₃₆O₅₆; (f) intensive model in Hf₃₆O₅₆. The parallelogram indicates a supercell region, while the green circles stand for the VCs. Some particular Hf atoms are marked by grey circles with number indices.

filamentary conduction scenario. It is subsequently a relevant question whether such conduction can be maintained when the supercell becomes larger, while keeping the size of the CF fixed. For this purpose, we constructed a $4 \times 1 \times 4$ m -HfO₂ supercell and introduced the same CF. After full structural relaxation (optimized structure shown in Fig. 7), we obtained its electronic structure using GGA-1/2. The density of states shown in the inset of Fig. 7 depicts a continuous band across the Fermi level, similar to that in Fig. 6(f). This confirms that the local filamentary conduction does not depend on the global stoichiometry of the whole dielectric layer, as long as the filament is stable by itself and also being highly metallic, such as the metal Hf clusters.

On the other hand, in our disperse VC models there exist energy gaps for the carriers, thus it is desirable to investigate whether the global stoichiometry should influence the band gap. For this purpose we chose the model in Fig. 5(c) and enlarged the HfO₂ supercell to $4 \times 1 \times 4$. The fully optimized VC model is shown in Fig. 8(a), where Fig. 8(b) demonstrates its electronic density of states near the Fermi level calculated with GGA-1/2. When the supercell is enlarged from $3 \times 1 \times 3$ to $4 \times 1 \times 4$, the band gap increases from ~ 0.2 eV to ~ 0.3 eV. This phenomenon can be explained by the Coulomb repulsion among the trapped electrons on the

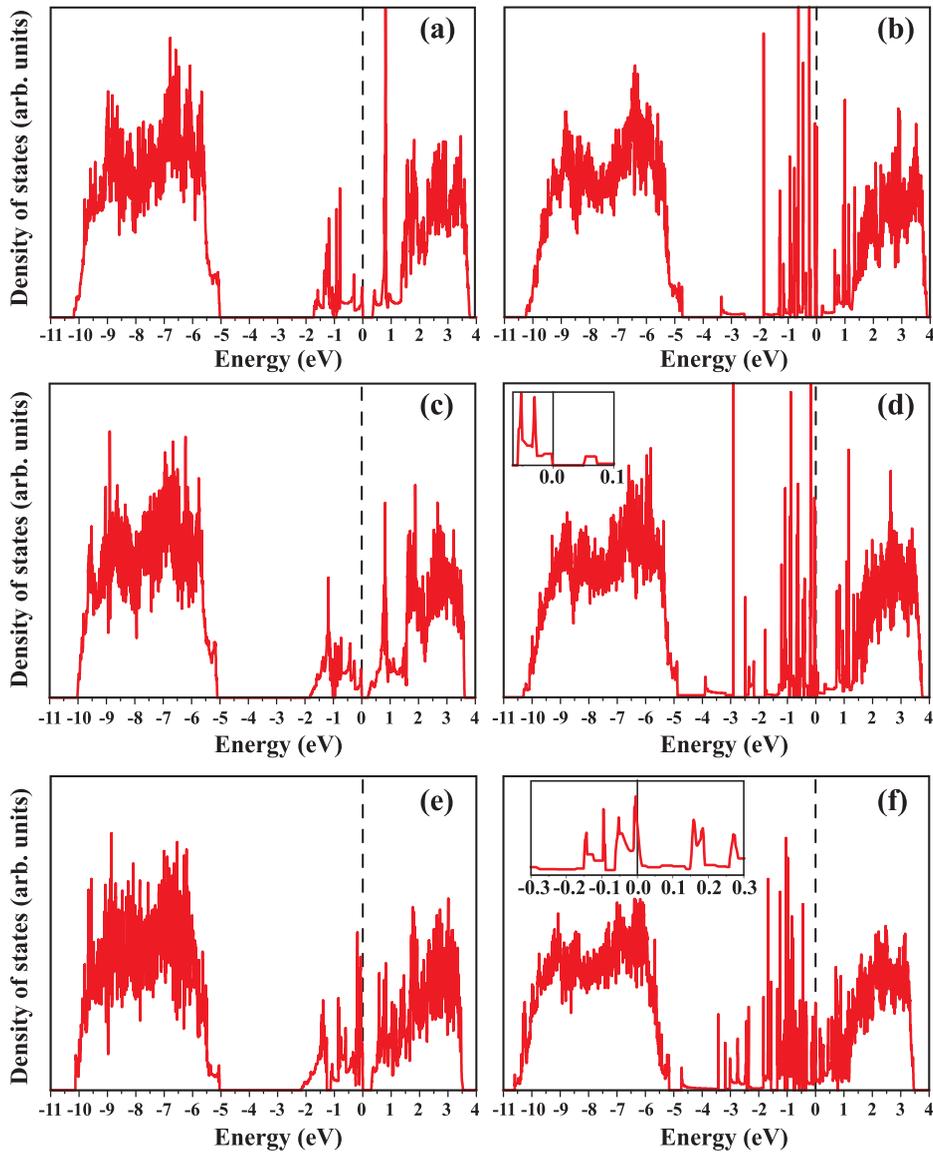


FIG. 6. Electronic density of states for the models in Fig. 5. (a) Disperse model of $\text{Hf}_{36}\text{O}_{64}$; (b) intensive model of $\text{Hf}_{36}\text{O}_{64}$; (c) disperse model of $\text{Hf}_{36}\text{O}_{60}$; (d) intensive model of $\text{Hf}_{36}\text{O}_{60}$; (e) disperse model of $\text{Hf}_{36}\text{O}_{56}$; (f) intensive model of $\text{Hf}_{36}\text{O}_{56}$. The Fermi levels are set to zero energy. Insets in (d) and (f) show the density of states close to the Fermi levels with higher resolution.

neutral V_o sites. As the global stoichiometry becomes more oxygen-deficient, the density of trapped electrons increases, which yields stronger repulsion and higher energy levels for the localized states stemming from these trapped electrons. On the other hand, the conduction band electrons are close to delocalized Bloch states. Hence, in case of an existing band gap, such band gap in the disperse VC model depends on the global stoichiometry, where more VC leads to smaller gap.

E. Energetic comparison for various phase separations

The models shown in Figs. 5(f) and 7 both contain a sufficiently large Hf-cluster CF per each supercell, and they both show local metallic conduction inside their CF regions.

TABLE II. Charges on certain Hf atoms marked in Fig. 5 using Bader analysis.

Atom	1#	2#	3#	4#	5#	6#	7#	8#	9#
Charge (e)	+2.70	+2.05	+2.08	+2.05	+2.03	+1.42	+0.07	+0.28	-0.17

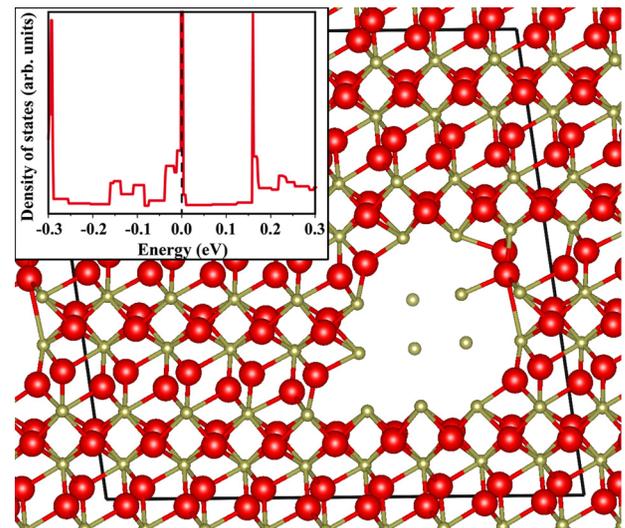


FIG. 7. Optimized atomic structures of an intensive filament in the $4 \times 4 \times 4 m\text{-HfO}_2$ supercell. The inset shows the GGA-1/2 electronic density of states near the Fermi level for this model.

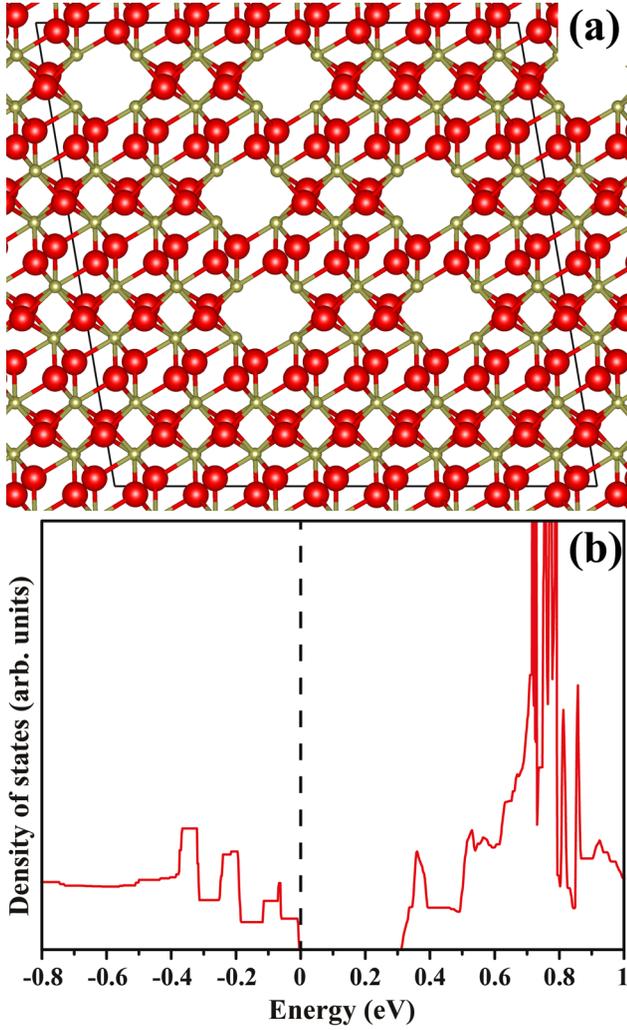


FIG. 8. (a) Optimized atomic structures of a disperse VC model in the $4 \times 1 \times 4$ m -HfO₂ supercell; (b) its GGA-1/2 electronic density of states near the Fermi level.

Their global stoichiometries are Hf₃₆O₅₆ (\sim HfO_{1.56}) and Hf₆₄O₁₁₂ (\sim HfO_{1.75}). Of course these models cannot be the lowest energy configurations for the two stoichiometries, but in an RRAM cell the geometrical constraints due to top and bottom electrodes may forbid arbitrary phase separations in the dielectric layer, thanks to the influence of various interfaces. Yet, it is still desirable to estimate the relative stability of these models by comparing with bulk phases. At zero temperature and atmospheric pressure, the thermodynamically stable phases in the Hf-O systems are *h.c.p.* Hf, hexagonal Hf₆O (*h*-Hf₆O) and *m*-HfO₂. Two metastable phases, tetragonal Hf₂O₃ (*t*-Hf₂O₃) and hexagonal HfO (*h*-HfO), have also been predicted in recently years^{9,54,55} and may also be taken into account. Table III lists the comprehensive results of such energy comparison. Since both Hf₃₆O₅₆ and Hf₆₄O₁₁₂ are between the stable phases Hf₆O and HfO₂ in stoichiometry, ideally they should decompose into *h*-Hf₆O and *m*-HfO₂ in vacuum. We thus define the proper *h*-Hf₆O and *m*-HfO₂ mixture states as the ground state

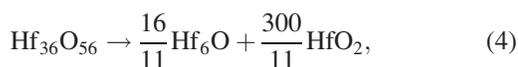
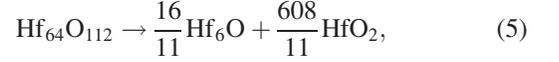


TABLE III. Energy comparison of our filamentary Hf₃₆O₅₆ and Hf₆₄O₁₁₂ supercell models with respect to various mixtures of pure bulk phases.

Configuration	Energy cost per oxygen vacancy (eV/O)	
	Hf ₃₆ O ₅₆	Hf ₆₄ O ₁₁₂
Hf-in- <i>m</i> -HfO ₂ model	0.69	0.71
<i>h</i> -Hf ₆ O + <i>m</i> -HfO ₂	0 (defined)	0 (defined)
<i>h.c.p.</i> Hf + <i>m</i> -HfO ₂	0.04	0.04
<i>t</i> -Hf ₂ O ₃ + <i>m</i> -HfO ₂	0.25	0.25
<i>h</i> -HfO + <i>m</i> -HfO ₂	1.84	1.84



and their total energies are used for other compositions as the zero point. For all other configurations, their energies for an Hf₃₆O₅₆ or Hf₆₄O₁₁₂ supercell are calculated with respect to these standard states. The energy cost of a certain configuration is therefore its energy divided by the effective number of oxygen vacancies in one supercell, where adding the same number of oxygen atoms back to the supercell shall yield an HfO₂ stoichiometry.

Under both stoichiometries, our Hf-in-*m*-HfO₂ filamentary model costs about 0.7 eV per oxygen vacancy, which is superior to the formation of *h*-HfO with *m*-HfO₂, but less favorable than bulk Hf (or Hf₆O) with *m*-HfO₂, as well as *t*-Hf₂O₃ with *m*-HfO₂ (Table III). Nevertheless, we emphasize that the filamentary model already includes the interfacial energy of Hf/HfO₂, thus whether the Hf CF is stable also depends on the interfacial energies of other configurations, such as Hf₆O/HfO₂ and Hf₂O₃/HfO₂. Under normal conditions the 0.7 eV/O energy cost is already sufficiently low to be achieved in real RRAM devices, and it is more favorable than the HfO filament formation even neglecting the interfacial energy between HfO and HfO₂.

There is almost no difference in the relative energy costs between the two stoichiometries, while for Hf-in-*m*-HfO₂ models we find a tiny discrepancy (0.71 eV/O versus 0.69 eV/O). The 0.02 eV/O difference is within the computational error range of DFT-GGA. Hence, if Hf metal clusters constitute the CFs, their formation kinetics and characteristics are insensitive to the regions about 1 nm far away within the dielectric, which further confirms the filamentary nature in the low resistance state of hafnia-RRAM.

F. Discussions

According to the results above, oxygen VCs show a clear cohesive tendency in HfO₂. During electroforming, local Hf clusters may emerge in HfO₂ due to the combination of oxygen VCs, where the Hf atoms at the center of the cluster usually exhibits zero-valence characteristic. Phase decomposition into Hf (possibly with some O interstitials) and HfO₂ is energetically favorable when the Hf cluster is large in cross section. The larger the Hf CF, the more stable it is. On the other hand, keeping fixed the same size of CF, their inter-distance does not play an important role in the

stability of individual CFs (Sec. III E), as long as they are not too close to each other.

An oxygen VC with a rather small cross section cannot lead to metallic conduction, but we predict that an intensive VC region with 12 Hf atoms and around 0.4 nm^2 area within the cross section is sufficient to yield metallic conduction. The experimental finding that the CF is usually much larger than this critical size should be understood from an energetic point of view. During the phase decomposition into Hf and HfO_2 , larger Hf CFs cost less energy than dispersed small Hf CFs, thus small CFs tend to merge into large CFs to minimize the energy.

Such argument can also explain the phenomenon discovered by Celano *et al.*¹⁴ that using a small AFM tip as an electrode one usually discovers only threshold switching. Suppose metal Hf cluster is the composition of the CF, then its stability strongly depends on its cross section area. Under the small electrode situations, the metal Hf cluster can become metastable and may dissolve automatically as long as the electrical stress is removed. Moreover, our calculations can explain the variability of operation modes of hafnia RRAM when the dielectric thickness is merely 3 nm, recently reported by Hou *et al.*⁵⁶ As the vertical dimension of hafnia shrinks to around 3 nm, the volume of a CF is limited by the vertical dimension, therefore the stability of the CFs degrades due to a low cohesive energy.

It is usually argued that grain boundaries are preferential leakage paths and the seed for electroforming in hafnia.^{57,58} If the hafnia thin film was polycrystalline before electroforming, such argument makes sense because grain boundaries sites are preferred sites for defect segregation, including oxygen vacancies and metal interstitials.⁵⁹ Our previous work on hafnia grain boundary indicates that the incorporation of metal interstitials (either Hf or foreign metal elements) is the most probable route towards metallic conduction along the grain boundaries. This seems easier compared with the merging of VCs, since there are usually pre-existing hollow regions at the grain boundary, accepting metal interstitials. Nevertheless, most hafnia-based RRAM devices utilize amorphous hafnia grown by atomic layer deposition (ALD), even with Si or Al doping to stabilize the amorphous phase. Our models of Hf-rich cluster CF embedded in an $m\text{-HfO}_2$ dielectric, stemming from the aggregation of VCs, is a probable scenario for the vicinity of the CF in the electroformed device.

On the technical side, the optimal stoichiometry of the initial hafnia was estimated as $\text{HfO}_{1.50}$ to $\text{HfO}_{1.75}$ according to McKenna.¹⁸ Our Hf-in- HfO_2 filamentary models as shown in Figs. 5(f) and 7 fall exactly within this range. As in real devices there are plenty of choices for electrodes, and also many possible dopants in HfO_2 , one cannot fully eliminate the possibility of metastable sub-oxides as the CF in hafnia-based RRAMs. However, our calculation suggests that with this optimal initial stoichiometry, phase decomposition into $h\text{-HfO}$ and HfO_2 is much less probable than into $t\text{-Hf}_2\text{O}_3$ and HfO_2 , though both cases are less favorable than decomposing into Hf-rich cluster HfO_x ($0 < x < 0.17$) and HfO_2 .

IV. CONCLUSIONS

We show through *ab initio* calculations that in hafnia-based RRAMs the aligned oxygen vacancy chains have the tendency to aggregate into an intensive Hf-cluster filament region during the electroforming process. The cohesion effect of oxygen vacancy chains is even more prominent when the Hf-cluster is larger, thus the conductive filaments in hafnia tend to expand in lateral directions. In addition, dispersed single oxygen vacancy chains in hafnia do not show metallic conduction behavior according to their electronic structure analysis, but an intensive Hf-rich cluster exhibits metallic behavior when the cross section of the filament consists of 12 Hf atoms or more. Hence, the usually observed large cross section area for the filaments in hafnia ($> 5 \text{ nm}^2$) stems from an energetic cohesion effect for oxygen vacancy defects, though an Hf-cluster filament with even smaller cross section ($\sim 1 \text{ nm}^2$), if existing, may still exhibit metallic conduction. Our results can also explain the threshold switching observed in hafnia when the small size AFM tip was used as an electrode, because the small volume of the filament incurs energetic instability. Our work excludes the possibility of individual oxygen vacancy chains with small cross section dimensions as the filaments in hafnia, but supports the scenario of phase decomposition into HfO_2 and Hf-rich clusters during electroforming.

ACKNOWLEDGMENTS

This work was financially supported by the MOST of China under Grant Nos. 2016YFA0203800 and 2017YFB0701701, the Natural Science Foundation of Hubei Province under Grant No. 2016CFB223, and the Fundamental Research Funds for the Central Universities of China under Grant No. HUST:2016YXMS212.

¹R. Waser and M. Aono, *Nat. Mater.* **6**, 833 (2007).

²K. M. Kim, D. S. Jeong, and C. S. Hwang, *Nanotechnology* **22**, 254002 (2011).

³F. Miao, J. P. Strachan, J. J. Yang, M.-X. Zhang, I. Goldfarb, A. C. Torrezan, P. Eschbach, R. D. Kelley, G. Medeiros-Ribeiro, and R. S. Williams, *Adv. Mater.* **23**, 5633 (2011).

⁴Z. Zeng, F. Triozon, and Y.-M. Niquet, *J. Appl. Phys.* **121**, 114503 (2017).

⁵A. Kumar, S. Mondal, and K. S. R. Koteswara Rao, *J. Appl. Phys.* **121**, 085301 (2017).

⁶T. M. Tsai, C. H. Wu, K. C. Chang, C. H. Pan, P. H. Chen, N. K. Lin, J. C. Lin, Y. S. Lin, W. C. Chen, H. Wu, N. Deng, and H. Qian, *IEEE Electron Device Lett.* **38**, 445 (2017).

⁷P. Huang, X. Y. Liu, B. Chen, H. T. Li, Y. J. Wang, Y. X. Deng, K. L. Wei, L. Zeng, B. Gao, G. Du, X. Zhang, and J. F. Kang, *IEEE Trans. Electron Devices* **60**, 4090 (2013).

⁸D. Duncan, B. Magyari-Kope, and Y. Nishi, *IEEE Electron Device Lett.* **37**, 400 (2016).

⁹K.-H. Xue, P. Blaise, L. R. C. Fonseca, and Y. Nishi, *Phys. Rev. Lett.* **110**, 065502 (2013).

¹⁰Y.-E. Syu, T.-C. Chang, J.-H. Lou, T.-M. Tsai, K.-C. Chang, M.-J. Tsai, Y.-L. Wang, M. Liu, and S. M. Sze, *Appl. Phys. Lett.* **102**, 172903 (2013).

¹¹J. Shang, W. Xue, Z. Ji, G. Liu, X. Niu, X. Yi, L. Pan, Q. Zhan, X.-H. Xu, and R.-W. Li, *Nanoscale* **9**, 7037 (2017).

¹²J. Zhang, A. R. Oganov, X. Li, K.-H. Xue, Z. Wang, and H. Dong, *Phys. Rev. B* **92**, 184104 (2015).

¹³U. Celano, L. Goux, R. Degraeve, A. Fantini, O. Richard, H. Bender, M. Jurczak, and W. Vandervorst, *Nano Lett.* **15**, 7970 (2015).

¹⁴U. Celano, A. Fantini, R. Degraeve, M. Jurczak, L. Goux, and W. Vandervorst, *AIP Adv.* **6**, 085009 (2016).

- ¹⁵C. Li, B. Gao, Y. Yao, X. Guan, X. Shen, Y. Wang, P. Huang, L. Liu, X. Liu, J. Li, C. Gu, J. Kang, and R. Yu, *Adv. Mater.* **29**, 1602976 (2017).
- ¹⁶B. Magyari-Köpe, S. G. Park, H.-D. Lee, and Y. Nishi, *J. Mater. Sci.* **47**, 7498 (2012).
- ¹⁷K. Kamiya, M. Young Yang, S.-G. Park, B. Magyari-Köpe, Y. Nishi, M. Niwa, and K. Shiraishi, *Appl. Phys. Lett.* **100**, 073502 (2012).
- ¹⁸K. P. McKenna, *Model. Simul. Mater. Sci. Eng.* **22**, 025001 (2014).
- ¹⁹L. Zhao, S. Clima, B. Magyari-Köpe, M. Jurczak, and Y. Nishi, *Appl. Phys. Lett.* **107**, 013504 (2015).
- ²⁰P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- ²¹W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- ²²G. Kresse and J. Furthmüller, *Comput. Mater. Sci.* **6**, 15 (1996).
- ²³G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- ²⁴P. E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994).
- ²⁵G. Kresse and D. Joubert, *Phys. Rev. B* **59**, 1758 (1999).
- ²⁶J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke, *Phys. Rev. Lett.* **100**, 136406 (2008).
- ²⁷H. J. Monkhorst and J. D. Pack, *Phys. Rev. B* **13**, 5188 (1976).
- ²⁸J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003).
- ²⁹L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- ³⁰L. G. Ferreira, M. Marques, and L. K. Teles, *Phys. Rev. B* **78**, 125116 (2008).
- ³¹L. G. Ferreira, M. Marques, and L. K. Teles, *AIP Adv.* **1**, 032119 (2011).
- ³²R. R. Pelá, M. Marques, L. G. Ferreira, J. Furthmüller, and L. K. Teles, *Appl. Phys. Lett.* **100**, 202408 (2012).
- ³³J. P. T. Santos, M. Marques, L. G. Ferreira, R. R. Pelá, and L. K. Teles, *Appl. Phys. Lett.* **101**, 112403 (2012).
- ³⁴J. Furthmüller, F. Hachenberg, A. Schleife, D. Rogers, F. Hosseini Teherani, and F. Bechstedt, *Appl. Phys. Lett.* **100**, 022107 (2012).
- ³⁵F. Matusalem, R. R. Pelá, M. Marques, and L. K. Teles, *Phys. Rev. B* **90**, 224102 (2014).
- ³⁶M. Ribeiro, *Can. J. Phys.* **93**, 261 (2015).
- ³⁷F. Matusalem, M. Marques, L. K. Teles, and F. Bechstedt, *Phys. Rev. B* **92**, 045436 (2015).
- ³⁸X.-B. Yin, R. Yang, K.-H. Xue, Z.-H. Tan, X.-D. Zhang, X.-S. Miao, and X. Guo, *Phys. Chem. Chem. Phys.* **18**, 31796 (2016).
- ³⁹K.-H. Xue, L. R. C. Fonseca, and X.-S. Miao, *RSC Adv.* **7**, 21856 (2017).
- ⁴⁰J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, *J. Phys. Condens. Matter* **14**, 2745 (2002).
- ⁴¹M. Balog, M. Schieber, M. Michman, and S. Patai, *Thin Solid Films* **41**, 247 (1977).
- ⁴²V. V. Afanas'ev, A. Stesmans, F. Chen, X. Shi, and S. A. Campbell, *Appl. Phys. Lett.* **81**, 1053 (2002).
- ⁴³E. Rauwel, A. Galeckas, and P. Rauwel, *Mater. Res. Express* **1**, 015035 (2014).
- ⁴⁴R. E. Hann, P. R. Suitch, and J. L. Pentecost, *J. Am. Ceram. Soc.* **68**, C-285 (1985).
- ⁴⁵S.-G. Lim, S. Kriventsov, T. N. Jackson, J. H. Haeni, D. G. Schlom, A. M. Balbashov, R. Uecker, P. Reiche, J. L. Freeouf, and G. Lucovsky, *J. Appl. Phys.* **91**, 4500 (2002).
- ⁴⁶J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari, and L. A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989).
- ⁴⁷L. Wang, T. Maxisch, and G. Ceder, *Phys. Rev. B* **73**, 195107 (2006).
- ⁴⁸A. S. Foster, F. Lopez Gejo, A. L. Shluger, and R. M. Nieminen, *Phys. Rev. B* **65**, 174117 (2002).
- ⁴⁹J. X. Zheng, G. Ceder, T. Maxisch, W. K. Chim, and W. K. Choi, *Phys. Rev. B* **75**, 104112 (2007).
- ⁵⁰L. Zhao, S.-G. Park, B. Magyari-Köpe, and Y. Nishi, *Appl. Phys. Lett.* **102**, 083506 (2013).
- ⁵¹Q. Luo, X. Xu, H. Liu, H. Lv, T. Gong, S. Long, Q. Liu, H. Sun, W. Banerjee, L. Li, J. Gao, N. Lu, and M. Liu, *Nanoscale* **8**, 15629 (2016).
- ⁵²B. Gao, J. F. Kang, Y. S. Chen, F. F. Zhang, B. Chen, P. Huang, L. F. Liu, X. Y. Liu, Y. Y. Wang, X. A. Tran, Z. R. Wang, H. Y. Yu, and A. Chin, in *2011 International Electron Devices Meeting* (2011), p. 17.4.1.
- ⁵³J. F. Kang, B. Gao, P. Huang, H. T. Li, Y. D. Zhao, Z. Chen, C. Liu, L. F. Liu, and X. Y. Liu, in *IEEE International Electron Devices Meeting IEDM* (2015), p. 5.4.1.
- ⁵⁴K.-H. Xue, B. Traore, P. Blaise, L. R. C. Fonseca, E. Vianello, G. Molas, B. D. Salvo, G. Ghibaudo, B. Magyari-Köpe, and Y. Nishi, *IEEE Trans. Electron Devices* **61**, 1394 (2014).
- ⁵⁵B. Puchala and A. Van der Ven, *Phys. Rev. B* **88**, 094108 (2013).
- ⁵⁶Y. Hou, U. Celano, L. Goux, L. Liu, R. Degraeve, Y. Cheng, J. Kang, M. Jurczak, and W. Vandervorst, *Appl. Phys. Lett.* **109**, 023508 (2016).
- ⁵⁷G. Bersuker, D. C. Gilmer, D. Veksler, P. Kirsch, L. Vandelli, A. Padovani, L. Larcher, K. McKenna, A. Shluger, V. Iglesias, M. Porti, and M. Nafria, *J. Appl. Phys.* **110**, 124518 (2011).
- ⁵⁸K. Shubhakar, S. Mei, M. Bosman, N. Raghavan, A. Ranjan, S. J. O'Shea, and K. L. Pey, *Microelectron. Reliab.* **64**, 204 (2016).
- ⁵⁹K.-H. Xue, P. Blaise, L. R. C. Fonseca, G. Molas, E. Vianello, B. Traoré, B. D. Salvo, G. Ghibaudo, and Y. Nishi, *Appl. Phys. Lett.* **102**, 201908 (2013).