CHIN. PHYS. LETT. Vol. 38, No. 8 (2021) 087103 **Express Letter**

Universal Theory and Basic Rules of Strain-Dependent Doping Behaviors in Semiconductors

Xiaolan Yan(晏晓岚)[1], Pei Li(李培)[1], Su-Huai Wei(魏苏淮)[1,2*], and Bing Huang(黄兵)[1,2*]

[1]*Beijing Computational Science Research Center, Beijing 100193, China*
[2]*Department of Physics, Beijing Normal University, Beijing 100875, China*

(Received 10 June 2021; accepted 15 July 2021; published online 23 July 2021)

Enhancing the dopability of semiconductors via strain engineering is critical to improving their functionalities, which is, however, largely hindered by the lack of basic rules. In this study, for the first time, we develop a universal theory to understand the total energy changes of point defects (or dopants) with different charge states under strains, which can exhibit either parabolic or superlinear behaviors, determined by the size of defect-induced local volume change (ΔV). In general, ΔV increases (decreases) when an electron is added (removed) to (from) the defect site. Consequently, in terms of this universal theory, three basic rules can be obtained to further understand or predict the diverse strain-dependent doping behaviors, i.e., defect formation energies, charge-state transition levels, and Fermi pinning levels, in semiconductors. These three basic rules could be generally applied to improve the doping performance or overcome the doping bottlenecks in various semiconductors.

DOI: 10.1088/0256-307X/38/8/087103

The application of semiconductors in electronic and optoelectronic devices critically depends on their dopability. Generally, there are three important factors that can basically limit the dopability in semiconductors: (i) The desirable defects or dopants (generally denoted as defects hereafter) have limited solubility, i.e., their formation energies ($H_{\rm f}^{D,q}$) are too high.[1,2] (ii) The desirable defects have sufficient solubility, but they are too difficult to be ionized at room temperature, i.e., their charge-state transition levels ($\varepsilon^{q/q'}$) are too deep inside the bandgap.[1,2] (iii) The desirable defects have low $H_{\rm f}^{D,q}$ and shallow $\varepsilon^{q/q'}$, unfortunately, the intrinsic compensating defects can easily form and pin the Fermi-level position ($E_{\rm pin}$) deep inside the bandgap, preventing the further increase of desired free carriers.[1,2] Comparing (i) and (ii), (iii) is the most difficult one to be overcome as it belongs to the intrinsic property of semiconductors.

In the past decades, strain engineering is widely adopted to enhance the performances of semiconductors, e.g., optimize the electronic structures,[3–9] improve phase stabilities,[10,11] generate spin currents,[12] control carrier excitations or transports,[13–15] and modulate ion diffusion paths.[16,17] It is unsurprising that strain engineering has also been used to tune the doping performances in semiconductors.[18–30] Meanwhile, the strain effects are unavoidable during the growth or operation of various semiconductors.[23,29,31] However, it is rather puzzled that the strain-induced changes of doping behaviors for different point defects in different semiconductors are dramatically different.[18,20–26,28,30] Unfortunately, a universal theory that can intuitively understand all these diverse doping behaviors in different systems is still lacking, which prevents us to establish the basic rules to overcome the doping bottlenecks in semiconductors.

In this Letter, we develop a simple but *universal* theory for understanding the strain-dependent total energy changes of isolated point defects ($\Delta E_{\rm t}^{D,q}$) under different charge states, which is critically determined by the defect-induced local volume change (ΔV). Depending on the size of ΔV, the $\Delta E_{\rm t}^{D,q}$ of a defect can exhibit either parabolic ($\Delta V \sim 0$) or monotonic ($dE_{\rm t}^{D,q}/dV \sim -\Delta V$) dependences. Noticeably, the ΔV is charge-state q-dependent, which increases (decreases) for more negatively (positively) charged defects. Based on this universal theory of $\Delta E_{\rm t}^{D,q}$, we can establish *three basic rules* on understanding the strain-dependent $H_{\rm f}^{D,q}$, $\varepsilon^{q/q'}$, and $E_{\rm pin}$, which may consequently be applied to overcome the above (i)–(iii) doping problems in various semiconductor systems.

A Universal Theory on $\Delta E_{\rm t}^{D,q}$. For a system without a defect, its total energy $E_{\rm t}^{\rm host}(V)$ as a function of volume V, to the lowest order, follows $E_{\rm t}^{\rm host}(V) = \alpha_0(V - V_0)^2$, where V_0 is the equilibrium volume of host lattice and $\alpha_0 = \frac{1}{2}B_0/V_0$, with B_0 being the bulk modulus. Similarly, for a system with a point defect in the q charge state, its total energy $E_{\rm t}^{D,q}(V)$ follows $E_{\rm t}^{D,q}(V) = (\alpha_0 + \Delta\alpha)[V - (V_0 + \Delta V)]^2$, where $\Delta\alpha$ and ΔV are the changes of α_0 and V_0 induced by the defect, respectively.[22] Ignoring the high order terms,

Supported by the National Natural Science Foundation of China (Grant Nos. 11634003, 11991060, and 12088101), and NSAF (Grant No. U1930402).
*Corresponding authors. Email: suhuaiwei@csrc.ac.cn; bing.huang@csrc.ac.cn

the total energy changes $\Delta E_{\rm t}^{D,q}$ induced by the defect in the charge-state q as a function of V can be derived as

$$\begin{aligned}\Delta E_{\rm t}^{D,q}(V) &= E_{\rm t}^{D,q}(V) - E_{\rm t}^{\rm host}(V) \\ &= -2\alpha_0 \Delta V(V - V_0) + \Delta\alpha(V - V_0)^2. \quad (1)\end{aligned}$$

Obviously, $\Delta E_{\rm t}^{D,q}(V)$ is determined by the two terms that are mainly associated with $\Delta\alpha$ and ΔV in Eq. (1). The $\Delta\alpha$ is usually negligible, especially for substitutional defects where the chemical and size differences are small.[32] The ΔV depends on the size difference between the dopant and the host element, therefore, is noticeable in most cases.[21,22,26] In these cases, $\Delta E_{\rm t}^{D,q}(V)$ is largely determined by the first term of Eq. (1), giving rise to a linear dependence on V [Fig. 1(a)]. However, if the defect-induced ΔV is not significant, the high-order second term in Eq. (1) could become dominant, giving rise to a parabolic change of $\Delta E_{\rm t}^{D,q}(V)$ under strain [Fig. 1(b)].[23–25,30] Importantly, the ΔV can be rather sensitive to the different charge states of a defect, i.e., the ΔV of negatively (positively) charged state is usually larger (smaller) than that of the corresponding neutral one for a defect, due to the increased (decreased) electron occupation. Therefore, it is expected that dramatically different q-dependent $\Delta E_{\rm t}^{D,q}(V)$ under strains could exist even for the same defect in a semiconductor.

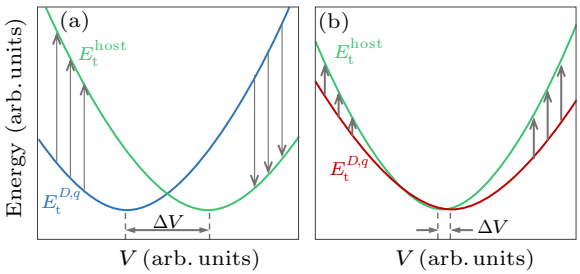

Fig. 1. Schematic plotting of the universal theory of ΔV-dependent $\Delta E_{\rm t}^{D,q}(V)$. $E_{\rm t}^{D,q}(V)$ and $E_{\rm t}^{\rm host}(V)$ are the total energies as a function of volume V for a host with and without a defect, respectively. $\Delta E_{\rm t}^{D,q}(V)$ is indicated by the black arrows, which are mostly determined by the (a) first and (b) second terms of Eq. (1) for the large and small ΔV cases, respectively.

The defect formation energy $H_{\rm f}^{D,q}(V, E_{\rm F})$ as a function of V can be written as[2,33–35]

$$H_{\rm f}^{D,q}(V, E_{\rm F}) = \Delta E_{\rm t}^{D,q}(V) + \sum n_i\mu_i + qE_{\rm F}, \quad (2)$$

where $E_{\rm F}$ is the absolute Fermi energy in the bandgap, μ_i is the chemical potential referenced to the total energy of elements in its lowest energy bulk form. Therefore, $H_{\rm f}^{D,q}(V)$ follows the same trend as $\Delta E_{\rm t}^{D,q}(V)$ as a function of V at a given $E_{\rm F}$. The $\varepsilon^{q/q'}(V)$ referenced to the valence band maximum (VBM) energy $E_{\rm VBM}(V)$ is the Fermi energy at which the same defect D with different charge states q and q' have the same formation energy, therefore, it can be described as

$$\begin{aligned}\varepsilon^{q/q'}(V) = &\left[\frac{\Delta E_{\rm t}^{D,q}(V) - \Delta E_{\rm t}^{D,q'}(V)}{q' - q} - E_{\rm VBM}(V_0)\right] \\ &- \Delta E_{\rm VBM}(V). \quad (3)\end{aligned}$$

As shown in Eq. (3), the $\varepsilon^{q/q'}(V)$ includes two terms: the first term represents the absolute value of $\varepsilon^{q/q'}(V)$ [denoted as a-$\varepsilon^{q/q'}(V)$] referenced to the VBM energy of the unstrained host, and the second term is the V-dependent shift of the VBM energy, $\Delta E_{\rm VBM}(V)$. It should be noticed that in a semiconductor, the $\Delta E_{\rm VBM}(V)$ depends on its absolute volume deformation potentials (AVDP). For most common semiconductors with a dominant bonding VBM state, the AVDP is positive,[36] i.e., the $\Delta E_{\rm VBM}(V)$ will increase (decrease) under tensile (compressive) strains. Therefore, the understanding of V-dependent a-$\varepsilon^{q/q'}(V)$ is the key for understanding the trend of $\varepsilon^{q/q'}(V)$. Based on the above discussions, in the following, we will derive three basic rules of strain-dependent doping behaviors in semiconductors with numerical verifications.

Rule I: Strain-Dependent $H_{\rm f}^{D,q}$. Taking GaN as a typical example, we have employed the first-principles calculations (see the computational methods in the Supplementary Material) to study the uniform strain η on the change of $H_{\rm f}^{D,q}$ [$\Delta H_{\rm f}^{D,q}(\eta)$] for different point defects in this system, in order to verify the universal theory we developed. Firstly, we consider N vacancy ($\rm V_N$), the dominant intrinsic defect in GaN.[37,38] As shown in Fig. 2(a), our calculation shows that a rather small value of ΔV ($\Delta V \sim -1.87\,\text{Å}^3/\rm V_N^0$) exists for $\rm V_N$ under its neutral charge state ($\rm V_N^0$). According to Fig. 1(b), the $\Delta H_{\rm f}^{D,q}(\eta)$ of $\rm V_N^0$ will be mostly determined by the second term of Eq. (1) and exhibit a parabolic dependence of η. Indeed, as shown in Fig. 2(b), the calculated $H_{\rm f}^{D,q}(\eta)$ of $\rm V_N^0$ tends to decrease under both compressive and tensile η. The negative sign of $\Delta\alpha \sim -0.06$ suggests that the local bulk modulus is reduced with the formation of $\rm V_N$, as is expected. Interestingly, when $\rm V_N^0$ is converted to its +1 state ($\rm V_N^{+1}$), its local volume is significantly reduced to $\Delta V \sim -7.3\,\text{Å}^3/\rm V_N^{+1}$ due to reduced charge occupation, giving rise to a large left-shift of its energy curve in Fig. 2(a). Consequently, differing from $\rm V_N^0$, it is expected that the $\Delta H_{\rm f}^{D,q}(\eta)$ of $\rm V_N^{+1}$ (at a given $E_{\rm F}$) is now mainly determined by the first term of Eq. (1) [Fig. 1(a)] and exhibits a linear dependence of η. Again, as confirmed in Fig. 2(b), the $H_{\rm f}^{D,q}(\eta)$ of $\rm V_N^{+1}$ linearly decreases (increases) under the compressive (tensile) η. In addition, compared to $\rm V_N^{+1}$, the linear slope of $\Delta H_{\rm f}^{D,q}(\eta)$ for $\rm V_N^{+3}$ is only slightly larger than that for $\rm V_N^{+1}$ due to the similar ΔV in both charge states, a reflection of Coulomb interaction between the defects and its local environment.

We emphasize that the similar q-dependent $\Delta H_{\rm f}^{D,q}(\eta)$ have also been observed for the intrinsic defects in other semiconductors, e.g., V_C in SiC (Fig. S1a), V_{Ga} in GaN (Fig. S1b), and V_{Zn} in ZnTe (Fig. S2a).

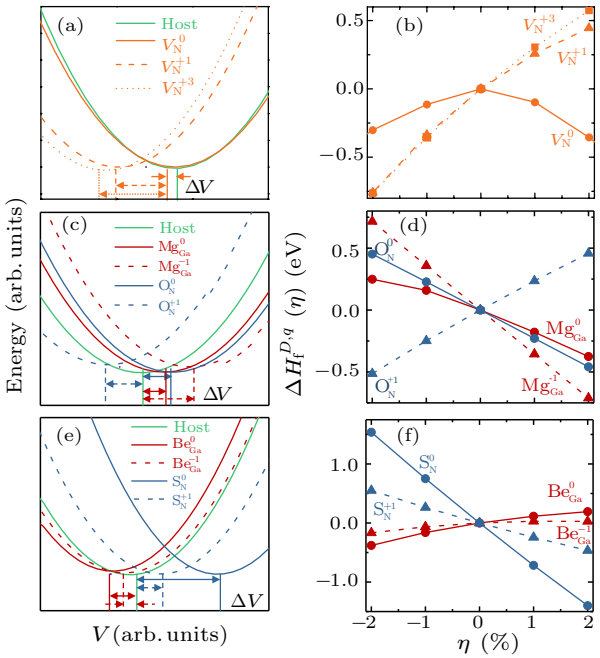

Fig. 2. Rule I on strain-dependent $H_{\rm f}^{D,q}$. Left panels: schematic plotting of total energies $E_{\rm t}^{D,q}(V)$ as a function of V for (a) N vacancy (V_N), (c) n-type (O_N) and p-type (Mg_{Ga}), and (e) n-type (S_N) and p-type (Be_{Ga}) in different charge states in GaN, respectively. $E_{\rm t}^{\rm host}(V)$ for GaN host (green lines) are also shown for comparison. Right panels: (b) Change of formation energies $\Delta H_{\rm f}^{D,q}(\eta)$ for V_N in different charge states in GaN as a function of strain η. (d) Similar to (b) but for O_N and Mg_{Ga} dopants in different charge states in GaN. (f) Similar to (b) but for S_N and Be_{Ga} dopants in different charge states in GaN.

Secondly, we consider the substitutional doping in GaN. In order to understand the size effects of dopants, O and S are selected as n-type dopants, while Mg and Be are selected as p-type ones. For the case of O_N^0, the electronic environment around the anion site induces a positive $\Delta V \sim +5.59\,\text{Å}^3/O_N^0$ [Fig. 2(c)]. Therefore, it is expected that the $\Delta H_{\rm f}^{D,q}(\eta)$ of O_N^0 will be dominated by the first term of Eq. (1) and exhibit a linear dependence of η. Indeed, as shown in Fig. 2(d), the calculated $\Delta H_{\rm f}^{D,q}(\eta)$ of O_N^0 linearly increases (decreases) as a function of compressive (tensile) η. When O_N^0 is ionized to O_N^{+1}, its ΔV largely shrinks to a negative value of $\Delta V \sim -5.75\,\text{Å}^3/O_N^{+1}$, which can make its energy curve in Fig. 2(c) largely left-shift. Consequently, the linear slope of $\Delta H_{\rm f}^{D,q}(\eta)$ for O_N^{+1} will be inverted in comparison to that of O_N^0, as also confirmed by our calculations shown in Fig. 2(d). Differing from O_N, as shown in Fig. 2(e), the larger ionic size of S than N induces a much larger ΔV after the S_N doping ($\Delta V \sim +17.72\,\text{Å}^3/S_N^0$), which can give rise to a much larger linear slope of $\Delta H_{\rm f}^{D,q}(\eta)$, as confirmed in Fig. 2(f). When S_N^0 is ionized to S_N^{+1}, its ΔV largely shrinks to a (still positive) value of $\sim+6.13\,\text{Å}^3/S_N^{+1}$, with a significant left-shift of its energy curve in Fig. 2(e). As a result, it is expected that the linear slope of $\Delta H_{\rm f}^{D,q}(\eta)$ for S_N^{+1} will be reduced in comparison to that of S_N^0, as confirmed by the calculations in Fig. 2(f). Therefore, depending on the initial different ΔV at neutral charge states, the n-type O_N^{+1} and S_N^{+1} can surprisingly have an *opposite* linear dependence of $\Delta H_{\rm f}^{D,q}(\eta)$. However, the ΔV always decreases when electron is removed from the dopant site (more positively charged), and $H_{\rm f}^{D,q}(\eta)$ follows $dE_{\rm t}^{D,q}(\eta)/dV \sim -\Delta V$.

Thirdly, we consider the p-type doping in GaN. As shown in Fig. 2(c), the ΔV of Mg_{Ga}^0 is positive ($\Delta V \sim +4.06\,\text{Å}^3/Mg_{Ga}^0$) due to the larger ionic size of Mg than Ga, which can further expand when it is ionized to Mg_{Ga}^{-1} ($\Delta V \sim +8.73\,\text{Å}^3/Mg_{Ga}^{-1}$). Therefore, it is expected that the Mg_{Ga}^{-1} can have a similar linearity but enlarged slope effect after it is ionized, as indeed confirmed by our calculations in Fig. 2(d). On the other hand, as shown in Fig. 2(e), the ΔV for Be_{Ga}^0 ($\Delta V \sim -3.26\,\text{Å}^3/Be_{Ga}^0$) is negative due to the small size of Be and after it is negatively charged, forming Be_{Ga}^{-1}, again, the ΔV increases to $\sim -1.07\,\text{Å}^3/Be_{Ga}^{-1}$. Because of the rather small (absolute) value of ΔV, it is expected that the $\Delta H_{\rm f}^{D,q}(\eta)$ of Be_{Ga} will be dramatically converted from a linear dependence of η to a parabolic dependence of η after it is ionized. Indeed, this surprising phenomenon is confirmed by our calculations in Fig. 2(f). Therefore, depending on the initial different ΔV at the neutral charge states, the p-type Mg_{Ga}^{-1} and Be_{Ga}^{-1} can have either linear or parabolic dependence of $\Delta H_{\rm f}^{D,q}(\eta)$. In general, we emphasize that the diverse trends of q-dependent $\Delta H_{\rm f}^{D,q}(\eta)$ for different dopants in different semiconductors, e.g., Cl_{Te} in ZnTe (Fig. S2b), Al_{Si} and N_C in SiC (Fig. S3), C_N and Ge_{Ga} in GaN (Figs. S4a and S4b), Zn_{Ga}, Si_P, Ge_{Ga}, and S_P in GaP (Figs. S4c and S4d) and As_{Si} in Si,[21] could be understood in a similar way via tracking the evolution of the q-dependent ΔV. Importantly, some related experimental observations, e.g., the solubility of B in Si can be greatly enhanced by a compressive strain,[39,40] can be well understood in a similar way by our above analysis.

From the above analysis, we can safely reach the Rule I on η-dependent $\Delta H_{\rm f}^{D,q}(\eta)$ in semiconductors: $\Delta H_{\rm f}^{D,q}(\eta)$ of point defects under different charge states can have either parabolic ($\Delta V \sim 0$) or linear ($\Delta V \neq 0$) dependence, i.e., $dE_{\rm t}^{D,q}(V)/dV \sim -\Delta V$. Here, the defect-induced ΔV is strongly q-dependent, it increases (decreases) when electron is added (removed) to (from) the dopant site. We would emphasize that this rule is independent of the sizes of supercell calculations (Fig. S5) or the forms of strains (Fig. S6).

Rule Ⅱ: Strain-Dependent $\varepsilon^{q/q'}$. According to Eq. (3), the η-dependent a-$\varepsilon^{q/q'}(\eta)$, i.e., without the consideration of $\Delta E_{\rm VBM}(V)$, is solely determined by the η-dependent $\Delta E_{\rm t}^{D,q}(\eta)$. In general, for a p-type acceptor in a semiconductor [Fig. 3(a)], in order to increase (decrease) the value of a-$\varepsilon^{0/-1}$, one needs to increase (decrease) its $H_{\rm f}^{D,-1}$ more than $H_{\rm f}^{D,0}$, which can be achieved under a negative (positive) η in terms of Rule I. Similarly, for an n-type donor [Fig. 3(a)], in order to increase (decrease) the value of a-$\varepsilon^{0/+1}$, one needs to increase (decrease) its $H_{\rm f}^{D,0}$ more than $H_{\rm f}^{D,+1}$, which can also be achieved under a negative (positive) η. We emphasize that those trends shown in Fig. 3(a) are universal for different point donors or acceptors in different semiconductors, due to the robustness of Rule I on q-dependent $\Delta H_{\rm f}^{D,q}(\eta)$ [$\Delta E_{\rm t}^{D,q}(\eta)$].

As shown in Fig. 3(b), taking the p-type Mg_{Ga} and n-type O_N in GaN as typical examples, it is unsurprising that the calculated strain-dependent a-$\varepsilon^{0/-1}(\eta)$ of Mg_{Ga} and a-$\varepsilon^{0/+1}(\eta)$ of O_N in GaN fully confirm our analysis in Fig. 3(a). The a-$\varepsilon^{0/-1}$ (a-$\varepsilon^{0/+1}$) moves towards VBM (CBM) under a tensile (compressive) η. Again, this trend shown in Fig. 3(b) is general for all the isolated point defects due to the generality of Rule I. It is also independent of the ΔV of defects at their neutral charge states, e.g., see the cases of Be_{Ga} and S_N in GaN (Fig. S7), Al_{Si} and N_C in SiC (Fig. S8) and Mg_{Al} in AlN (Fig. S9). The above discussions can drive us to Rule Ⅱ that *the negative (positive)* η *is always beneficial for the realization of shallower a-*$\varepsilon^{q/q'}(\eta)$ *for the donor (acceptor) in semiconductors.*

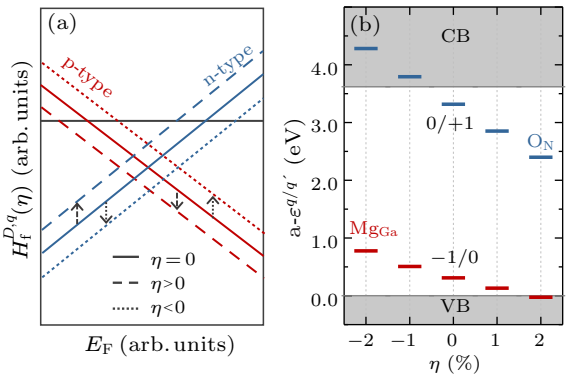

Fig. 3. Rule Ⅱ on strain-dependent a-$\varepsilon^{q/q'}$. (a) Schematic plotting of the change of formation energy $H_{\rm f}^{D,q}(\eta)$ for positively charged donor (blue) and negatively charged acceptor (red) in a semiconductor, with respect to their neutral states (gray), under different η as a function of $E_{\rm F}$. Arrows indicate the directions of those changes from without η to with η. (b) The a-$\varepsilon^{0/-1}(\eta)$ and a-$\varepsilon^{0/+1}(\eta)$ for Mg_{Ga} and O_N in GaN as a function of strain η, respectively. Here the band edge positions are fixed at the values of $E_{\rm VBM}(V_0)$ and $E_{\rm CBM}(V_0)$.

Rule Ⅱ can explain some related experimental observations, e.g., the ionization energy of B in Ge_xSi_{1-x} alloy gradually decreases with the increase of Ge composition,[41] attributed to the Ge-induced tensible strain effects. As discussed later, it should be noticed that the shallowness of $\varepsilon^{q/q'}(\eta)$ relative to the VBM of the strained system also depends on the shift $\Delta E_{\rm VBM}(V)$ as shown in Eq. (3).

Rule Ⅲ: Strain-Dependent $E_{\rm pin}$. Rules I and Ⅱ demonstrate that the strain can always induce opposite changes of $H_{\rm f}^{D,q}(\eta)$ for donors and acceptors, which can be utilized to tuning the Fermi level pinning $E_{\rm pin}$ positions. The $E_{\rm pin}$ level is determined by the Fermi energy at which the compensating donor and acceptor defects holding opposite charge states have the same energy. It is known that n-type $E_{\rm pin-n}$ and p-type $E_{\rm pin-p}$ positions set up the doping limits of n-type and p-type doping, respectively, in a semiconductor, which is an intrinsic problem of semiconductors that are difficult to overcome.[35]

Basically, based on the doping limit rules, a semiconductor with low VBM, e.g. GaN,[35] is difficult to be doped p-type, while a semiconductor with high CBM, e.g. ZnTe,[42] is difficult to be doped n-type. For GaN, Mg (Mg_{Ga}) has been widely selected as a p-type dopant,[2,37,43] which has an $\varepsilon^{0/-1}$ at VBM +0.3 eV [Fig. 3(b)]. However, when Fermi level $E_{\rm F}$ is shifted towards VBM after Mg_{Ga} doping, the spontaneous formation of V_N^{+3} can compensate the p-type doping induced by Mg_{Ga}^{-1}, giving rise to a deep $E_{\rm pin-p}$ position locating at VBM +0.7 eV, agreeing with previous calculations.[38] As shown in Fig. 4(a), although the $\varepsilon^{0/-1}$ of Mg_{Ga} is relatively shallow, its p-type doping performance is strongly downgraded by the formation of V_N^{+3}. To reduce $E_{\rm pin-p}$ position in GaN, one needs to increase the $H_{\rm f}^{D,q}$ of V_N^{+3} but decrease the $H_{\rm f}^{D,q}$ of Mg_{Ga}^{-1}, which can be achieved under a tensile-η-induced synergistic effect. Indeed, as shown in Fig. 4(c), our calculations confirm that the $H_{\rm f}^{q}(\eta)$ of V_N^{+3} (Mg_{Ga}^{-1}) can gradually increase (decrease) as a function of tensile η, linearly shifting the absolute $E_{\rm pin-p}(\eta)$ [a-$E_{\rm pin-p}(\eta)$, without consideration of $\Delta E_{\rm VBM}(V)$] towards lower energy positions.

For the case of ZnTe, Cl (Cl_{Te}) is commonly used as an n-type dopant[44,45] with a calculated $\varepsilon^{0/+1}$ at CBM −0.43 eV, agreeing with the previous calculations.[46] Unfortunately, as shown in Fig. 4(b), the spontaneous formation of V_{Zn}^{-2} can largely pin the $E_{\rm pin-n}$ position deeply inside the bandgap, i.e., $E_{\rm pin-n}$ = CBM −0.84 eV, preventing the ionization of Cl_{Te}. Similarly, as shown in Fig. 4(d), a compressive-η-induced synergistic effect may decrease the $H_{\rm f}^{D,q}$ of Cl_{Te}^{+1} but increase the $H_{\rm f}^{D,q}$ of V_{Zn}^{-2}, giving rise to the shift of a-$E_{\rm pin-n}$ towards higher energy positions.

Based on the above understandings, we can arrive at Rule Ⅲ on the η-dependent a-$E_{\rm pin}(\eta)$ in semiconductors: the tensile (compressive) strain is always beneficial for the realization of shallower a-$E_{\rm pin-p}(\eta)$ [a-$E_{\rm pin-n}(\eta)$] in semiconductors. Again, as discussed later, the shallowness of $E_{\rm pin-p}(\eta)$ [$E_{\rm pin-n}(\eta)$] rela-

tive to the VBM (CBM) of the strained system also depends on the shift $\Delta E_{\rm VBM}(V)$ [$\Delta E_{\rm CBM}(V)$]. Again, we emphasize that Rule Ⅲ is robust for different semiconductors due to the robustness of Rules I and Ⅱ. In addition, Rules I–Ⅲ are generally valid for different semiconductor systems under the biaxial strain (Fig. S6), because Rule I depends solely on the defect-induced ΔV.

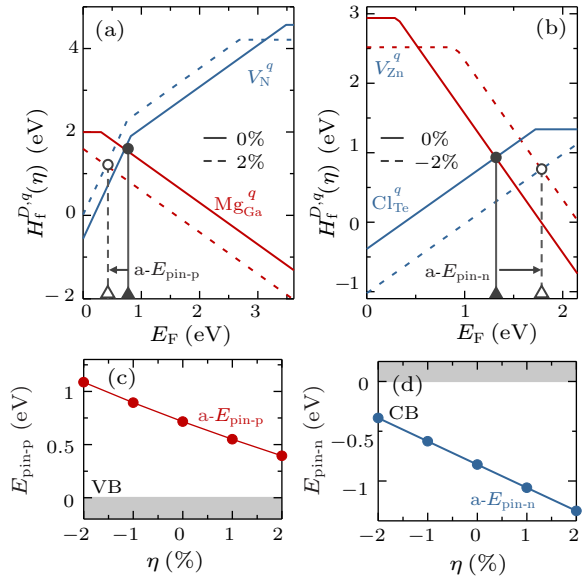

Fig. 4. Rule Ⅲ on strain-dependent a-$E_{\rm pin}$. (a) Formation energy $H_{\rm f}^{D,q}$ of external $\rm Mg_{Ga}$ and intrinsic compensating $\rm V_N$ in GaN without and with a +2% strain. (b) Formation energy $H_{\rm f}^{D,q}$ of external $\rm Cl_{Te}$ and intrinsic compensating $\rm V_{Zn}$ in ZnTe without and with a −2% strain. The a-$E_{\rm pin}$ positions are marked by the vertical lines in (a)–(b). (c) a-$E_{\rm pin-p}(\eta)$ and (d) a-$E_{\rm pin-n}(\eta)$ positions as a function of η in GaN and ZnTe, respectively. Here the band edge positions are fixed at the values of $E_{\rm VBM}(V_0)$ and $E_{\rm CBM}(V_0)$.

Role of $\Delta E_{\rm VBM}(V)/\Delta E_{\rm CBM}(V)$. Since Rules Ⅱ and Ⅲ are for a-$\varepsilon^{q/q'}(\eta)$ and a-$E_{\rm pin}(\eta)$, in order to obtain the rules of $\varepsilon^{q/q'}(\eta)$ and $E_{\rm pin}(\eta)$ with respect to the band edge states under strain, the $\Delta E_{\rm VBM}(V)/\Delta E_{\rm CBM}(V)$ needs to be taken into account. As shown in Fig. 5(a), in most common semiconductors with VBM (CBM) as bonding (anti-bonding) states, the $\Delta E_{\rm VBM}(V)$ [$\Delta E_{\rm CBM}(V)$] will almost linearly increase (decrease) as a function of V from compression to tension.[36] Therefore, combining Rules Ⅱ, Ⅲ and Fig. 5(a), we can easily reach the conclusion that Rules Ⅱ and Ⅲ can also apply for the relative $\varepsilon^{q/q'}(\eta)$ of acceptors and $E_{\rm pin-p}(\eta)$, because the opposite trend of η-dependent a-$\varepsilon^{0/-1}(\eta)$ [a-$E_{\rm pin-p}(\eta)$] and $\Delta E_{\rm VBM}(V)$ can induce a novel synergistic effect to make the $\varepsilon^{q/q'}(\eta)$ [$E_{\rm pin-p}(\eta)$] of acceptors even shallower under a tensile strain. Indeed, as shown in Figs. 5(b) and 5(c), our calculations on $\varepsilon^{q/q'}$ and $E_{\rm pin-p}$ in GaN confirm our expectation. The calculated $\varepsilon^{0/-1}$ [$E_{\rm pin-p}$] of $\rm Mg_{Ga}$ is shifted from VBM +0.3 [VBM +0.7] eV to VBM −0.15 [VBM +0.27] eV under $\eta = +2\%$, significantly shallower than that of a-$\varepsilon^{0/-1}(\eta)$ [a-$E_{\rm pin-p}$]. The trend of $\varepsilon^{q/q'}$ shown in Fig. 5(b) is consistent with the similar experimental observations that the $\rm Mg_{Ga}$ level becomes deeper in $\rm Al_xGa_{1-x}N$ as the x increases, due to the Al-induced compressive strain and band edge shift.[47] We emphasize that there are no similar rules for $\varepsilon^{q/q'}(\eta)$ of donors and $E_{\rm pin-n}(\eta)$ in common semiconductors. Because of the η-dependent a-$\varepsilon^{0/+1}(\eta)$ [a-$E_{\rm pin-n}(\eta)$] and $\Delta E_{\rm CBM}(V)$ shift in the same direction, the relative shift will depend on their individual changes.

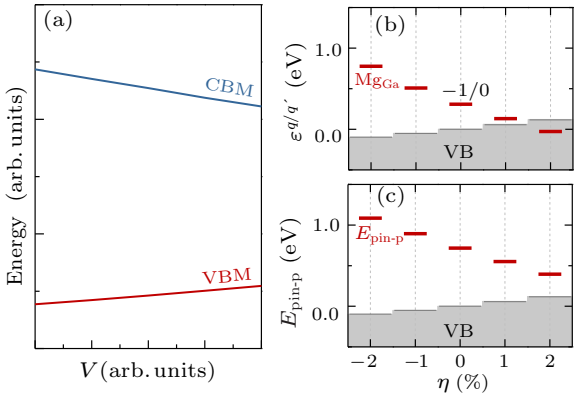

Fig. 5. Role of $\Delta E_{\rm VBM}(V)/\Delta E_{\rm CBM}(V)$. (a) Schematic plotting of the band edge changes as a function of V for common semiconductors with VBM (CBM) as bonding (anti-bonding) states. (b) The $\varepsilon^{0/-1}(\eta)$ of $\rm Mg_{Ga}$ and (c) $E_{\rm pin-p}(\eta)$ in GaN as a function of strain η, with consideration of $\Delta E_{\rm VBM}(V)$.

In summary, for the first time, we have developed a universal theory and consequently established three basic rules for understanding the diverse strain-dependent doping behaviors in semiconductors, which can be applied to tune their $H_{\rm f}^{D,q}$, $\varepsilon^{q/q'}$, and $E_{\rm pin}$, as successfully confirmed by the first-principles calculations on several exemplary semiconductors. Due to the robustness of our developed universal theory and Rule I, it is reasonable to expect that Rules Ⅱ and Ⅲ are also robust in different semiconductors. Therefore, these basic rules can be widely applied to control doping and simultaneously overcome the doping bottlenecks in semiconductors via simple strain engineering. Since the strain effects widely exist in different semiconductor films during the growths or operations, our theory can be adopted as a guideline to induce the "right" strain effects to improve the doping performance of semiconductors.

The authors thank L. Kang, L. Hu, and J. F. Wang for helpful discussions. The calculations were performed on Tianhe2-JK at CSRC.

References

[1] Freysoldt C, Neugebauer J and van de Walle C G 2011 *Advanced Calculations for Defects in Materials: Electronic Structure Methods* (New York: Wiley-VCH Verlag)

[2] Freysoldt C, Grabowski B, Hickel T *et al.* 2014 *Rev. Mod. Phys.* **86** 253
[3] Jacobsen R S, Andersen K N, Borel P I *et al.* 2006 *Nature* **441** 199
[4] Feng J, Qian X, Huang C W *et al.* 2012 *Nat. Photon.* **6** 866
[5] Jain J R, Hryciw A, Baer T M *et al.* 2012 *Nat. Photon.* **6** 398
[6] Yan Q, Rinke P, Janotti A *et al.* 2014 *Phys. Rev. B* **90** 125118
[7] Yang M M, Kim D J and Alexe M 2018 *Science* **360** 904
[8] Liu C, Song X, Li Q, Ma Y and Chen C 2021 *Chin. Phys. Lett.* **38** 086301
[9] Xie Y, Feng J, Xiang H and Gong X 2019 *Chin. Phys. Lett.* **36** 056801
[10] Yadav S K, Sadowski T and Ramprasad R 2010 *Phys. Rev. B* **81** 144120
[11] Chen Y, Lei Y, Li Y *et al.* 2020 *Nature* **577** 209
[12] Huang B, Jin K H, Cui B *et al.* 2017 *Nat. Commun.* **8** 15850
[13] Yu D, Zhang Y and Liu F 2008 *Phys. Rev. B* **78** 245204
[14] Zhou M, Liu Z, Wang Z *et al.* 2013 *Phys. Rev. Lett.* **111** 246801
[15] Harats M G, Kirchhof J N, Qiao M *et al.* 2020 *Nat. Photon.* **14** 324
[16] Zhou H B, Jin S, Zhang Y *et al.* 2012 *Phys. Rev. Lett.* **109** 135502
[17] Kalikka J, Zhou X, Dilcher E *et al.* 2016 *Nat. Commun.* **7** 11983
[18] Sadigh B, Lenosky T J, Caturla M J *et al.* 2002 *Appl. Phys. Lett.* **80** 4738
[19] Sun Y, Thompson S E and Nishida T 2007 *J. Appl. Phys.* **101** 104503
[20] Bennett N S, Smith A J, Gwilliam R M *et al.* 2008 *J. Vac. Sci. Technol. B* **26** 391
[21] Ahn C, Bennett N, Dunham S T *et al.* 2009 *Phys. Rev. B* **79** 073201
[22] Zhu J, Liu F, Stringfellow G B *et al.* 2010 *Phys. Rev. Lett.* **105** 195503
[23] Donner W, Chen C, Liu M *et al.* 2011 *Chem. Mater.* **23** 984
[24] Kan E, Wu F, Zhang Y *et al.* 2012 *Appl. Phys. Lett.* **100** 072401
[25] Aschauer U, Pfenninger R, Selbach S M *et al.* 2013 *Phys. Rev. B* **88** 054111
[26] Zhu J, Liu F and Scarpulla M A 2014 *APL Mater.* **2** 012110
[27] Zheng T, Lin W, Cai D *et al.* 2014 *Nanoscale Res. Lett.* **9** 40
[28] Zheng Y F, Chen S, Yang J H *et al.* 2019 *Phys. Rev. B* **99** 014113
[29] Chaudhuri R, Bader S J, Chen Z *et al.* 2019 *Science* **365** 1454
[30] Lu Y B, Dai Y, Wei W *et al.* 2013 *ChemPhysChem* **14** 3916
[31] Bean J C 1985 *Science* **230** 127
[32] Allred C L, Yuan X, Bazant M Z *et al.* 2004 *Phys. Rev. B* **70** 134113
[33] Zhang S B and Northrup J E 1991 *Phys. Rev. Lett.* **67** 2339
[34] Laks D B, Van de Walle C G, Neumark G F *et al.* 1992 *Phys. Rev. B Condens Matter* **45** 10965
[35] Wei S H 2004 *Comput. Mater. Sci.* **30** 337
[36] Li Y H, Gong X G and Wei S H 2006 *Phys. Rev. B* **73** 245206
[37] Lyons J L and Van de Walle C G 2017 *npj Comput. Mater.* **3** 12
[38] Miceli G and Pasquarello A 2016 *Phys. Rev. B* **93** 165207
[39] Moriya N, Feldman L C, Luftman H S *et al.* 1994 *J. Vac. Sci. Technol. B* **12** 383
[40] Ovsyannikov S V, Gou H, Karkin A E *et al.* 2014 *Chem. Mater.* **26** 5274
[41] Franz M, Pressel K and Gaworzewski P 1998 *J. Appl. Phys.* **84** 709
[42] Zhang S B, Wei S H and Zunger A 1998 *J. Appl. Phys.* **83** 3192
[43] Nakamura S, Senoh M and Mukai T 1991 *Jpn. J. Appl. Phys.* **30** L1708
[44] Tao I W, Jurkovic M and Wang W I 1994 *Appl. Phys. Lett.* **64** 1848
[45] Tanaka T, Hayashida K, Nishio M *et al.* 2003 *J. Appl. Phys.* **94** 1527
[46] Janotti A, Wei S H and Zhang S B 2003 *Appl. Phys. Lett.* **83** 3522
[47] Nakarmi M L, Nepal N, Lin J Y *et al.* 2009 *Appl. Phys. Lett.* **94** 091903

Supplementary Materials for

Universal Theory and Basic Rules of Strain-Dependent Doping Behaviors in Semiconductors

Xiaolan Yan[1], Pei Li[1], Su-Huai Wei[1,2], and Bing Huang[1,2]

[1] *Beijing Computational Science Research Center, Beijing 100193, China*

[2] *Department of Physics, Beijing Normal University, Beijing 100875, China*

List of Content

strain η. (b) a-$\varepsilon^{0/-1}(\eta)$ of Mg_{Al} as a function of strain η. Band edge positions are fixed at the values of unstrained AlN.

Part I: Method for defect calculations.

All the density functional theory (DFT) calculations are performed with Vienna Ab Initio Simulation Package (VASP) [1]. The exchange-correlation energy is treated using the generalized gradient approximation (GGA) in the PBE form [2]. The hybrid functional (HSE06) is adopted to correct the bandgaps of semiconductors [3]. The plane-wave cutoff energy is set to 520 eV and sufficient *k*-mesh are selected for all the systems. During the structural relaxations, a conjugate-gradient algorithm is used until the force on each atom was lower than 0.01 eV $Å^{-1}$, and the total energy is converged to 1.0×10^{-6} eV. For the defect calculations, the standard supercell approach is adopted [4, 5].

The formation energy (H_f) of a defect or dopant in semiconductors can be evaluated as [4, 6-8]

$$H_f^{D,q}(V, E_F) = E_t^{D,q}(V) - E_t^{host}(V) + \sum n_i \mu_i + qE_F \quad (1),$$

where $E_t^{host}(V)$ is the total energy of the supercell without a defect, and $E_t^{D,q}(V)$ is the total energy of a supercell with a defect in a charge state q. q is the number of electrons transferred from the supercell to the reservoirs in forming the defect cell and n_i is the number of atoms removed from or added into the supercell and μ_i is the chemical potential of atom i with respect to elemental solid/gas with energy E_i. E_F is the Fermi level.

For the defect calculation, sufficiently large supercell sizes have been adopted and we have tested our results using the different sizes of supercells and confirmed that our main conclusions maintain (see **Fig. 2** and **Figs. S1-S5**). For the strain calculations, the hydrostatic strain is considered in the main text, and our conclusion maintains for the biaxial strain based on our test calculations (see **Fig. S6**).

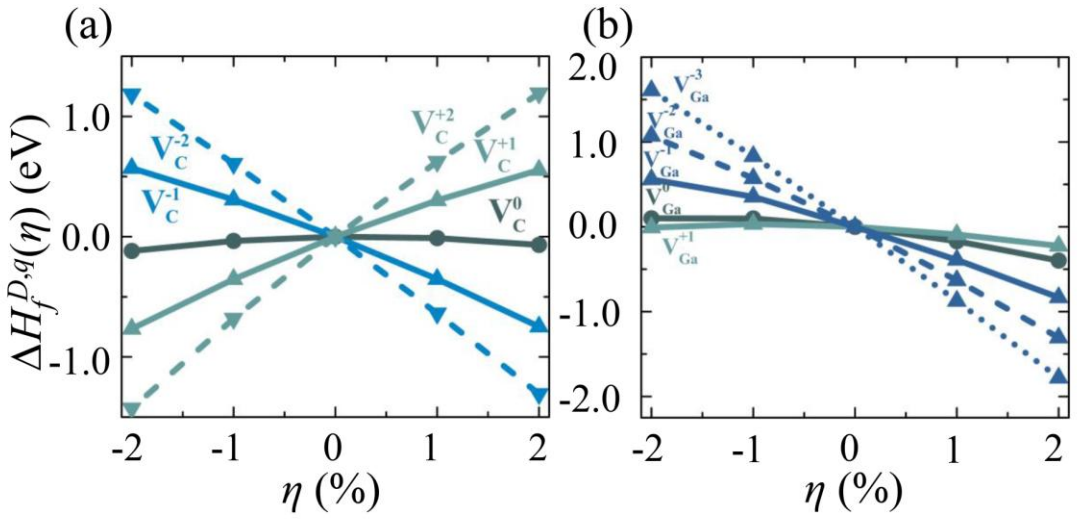

Fig. S1. (a) Change of formation energies $\Delta H_f^{D,q}(\eta)$ for C vacancy (V_C) in SiC as a function of strain η. (b) same for (a) but for Ga vacancy (V_{Ga}) in GaN.

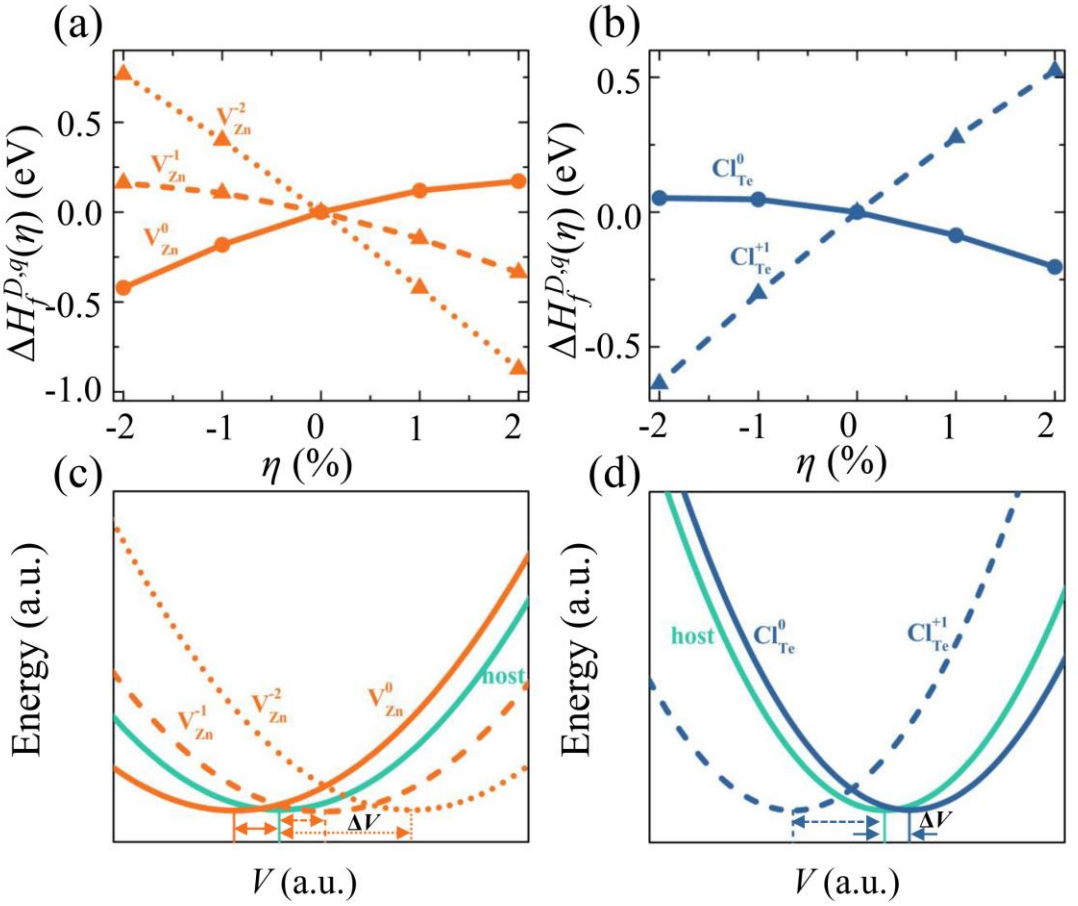

Fig. S2. (a) Change of formation energies $\Delta H_f^{D,q}(\eta)$ for Zn vacancy (V_{Zn}) in ZnTe as a function of strain η. (b) same for (a) but for *n*-type (Cl_{Te}) dopant in ZnTe. Schematic plotting of total energies $E_t^{D,q}(V)$ as a function of volume V for (c) V_{Zn} and (d) Cl_{Te}. $E_t^{host}(V)$ for host are also shown in (c)-(d) for comparison.

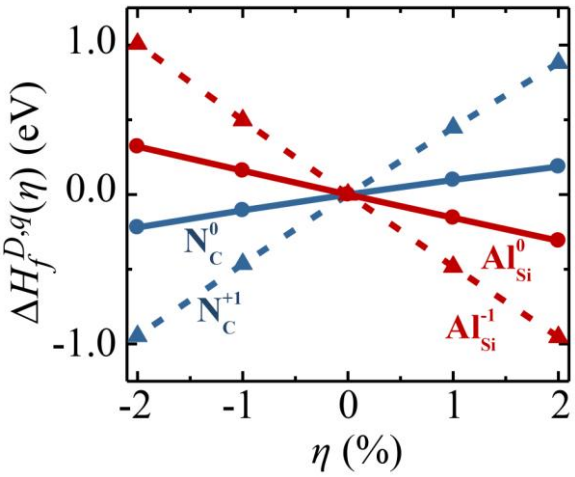

Fig. S3. Change of formation energies $\Delta H_f^{D,q}(\eta)$ for *n*-type (N_C) and *p*-type (Al_{Si}) dopants in SiC as a function of strain η.

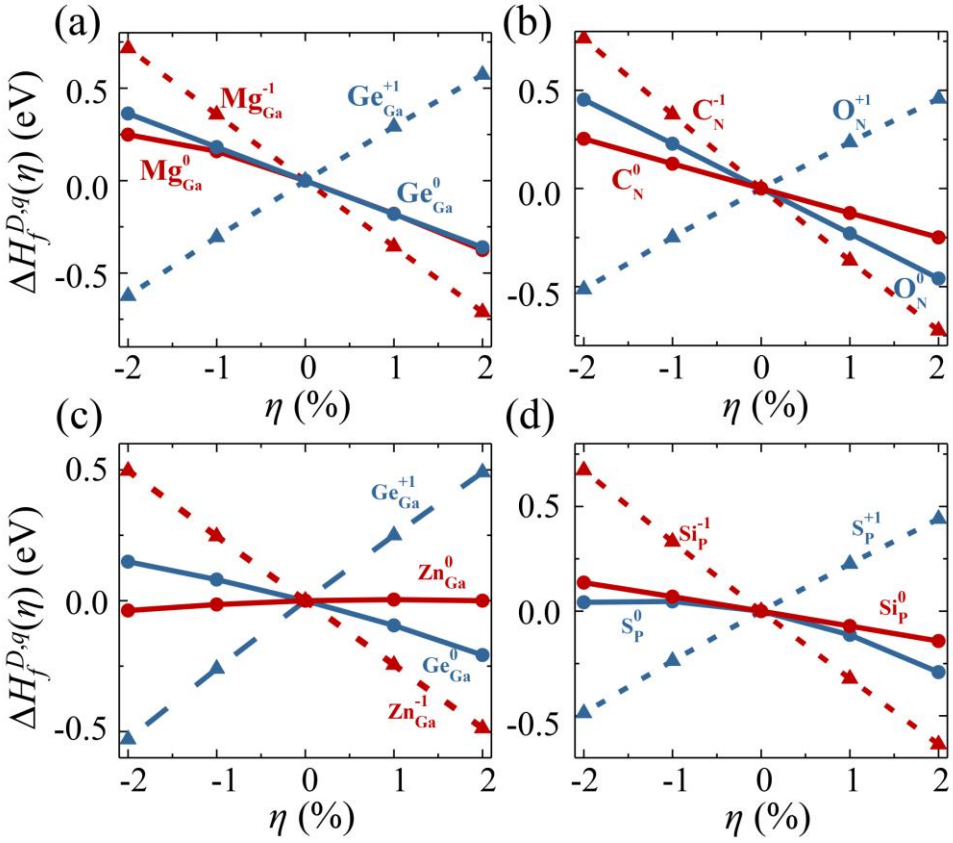

Fig. S4. (a) Change of formation energies $\Delta H_f^{D,q}(\eta)$ for *n*-type Ge_{Ga} and *p*-type Mg_{Ga} in GaN. (b) same for (a) but for *n*-type O_N and *p*-type C_N in GaN. (c) same for (a) but for *n*-type Ge_{Ga} and *p*-type Zn_{Ga} in GaP. (d) same for (a) but for *n*-type S_P and *p*-type Si_P in GaP.

Figure S4 also indicates that the trends are independent of the substitution positions either at anion or cation sites.

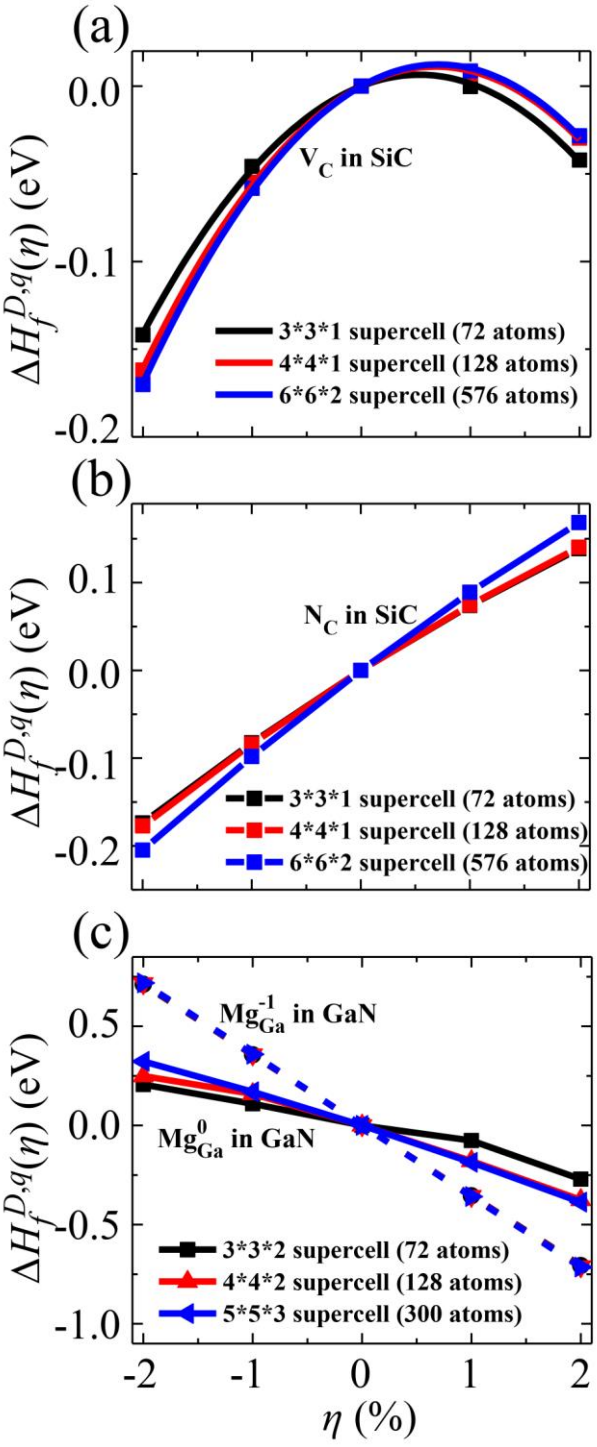

Fig. S5. Change of formation energies $\Delta H_f^{D,q}(\eta)$ for (a) V_C and (b) N_C in 4H-SiC with different supercell sizes, corresponding to different defect concentrations, under strain η. (c) Similar to (b) but for Mg_{Ga} dopants in both q=0 and q=-1 states in GaN.

Figure S5 indicates that the trends of $\Delta H_f^{D,q}(\eta)$ in different semiconductors is independent of the sizes of supercell calculations.

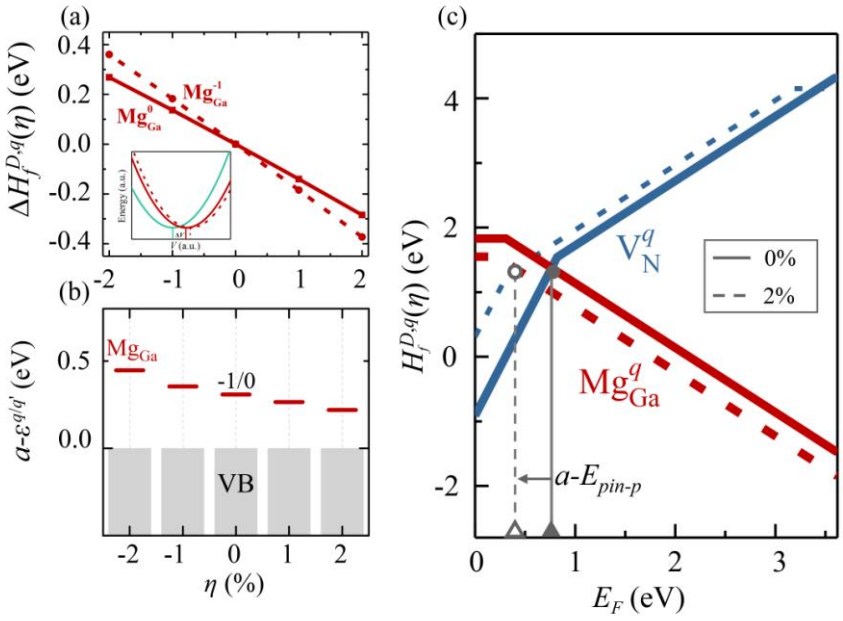

Fig. S6. (a) Change of formation energies $\Delta H_f^{D,q}(\eta)$ for Mg_{Ga} in GaN as a function of biaxial strain η. Inset: Schematic plotting of total energies $E_t^{D,q}(V)$ as a function of volume V for Mg_{Ga}, $E_t^{host}(V)$ for host is also shown for comparison. (b) $a\text{-}\varepsilon^{0/-1}(\eta)$ of Mg_{Ga} as a function of biaxial strain η. Band edge positions are fixed at the values of unstrained GaN. (c) Formation energies of external Mg_{Ga} and intrinsic compensating V_N in GaN without and with a +2% biaxial strain.

It indicates that the Rule Nos. I-III are valid for both uniform strain and biaxial strain.

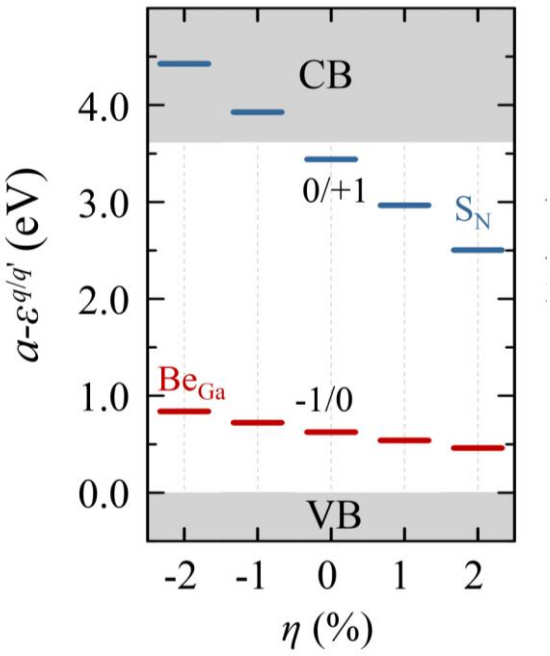

Fig. S7. $a\text{-}\varepsilon^{0/-1}(\eta)$ [$a\text{-}\varepsilon^{0/+1}(\eta)$] of Be_{Ga} (S_N) in GaN as a function of strain η. Band edge positions are fixed at the values of unstrained GaN.

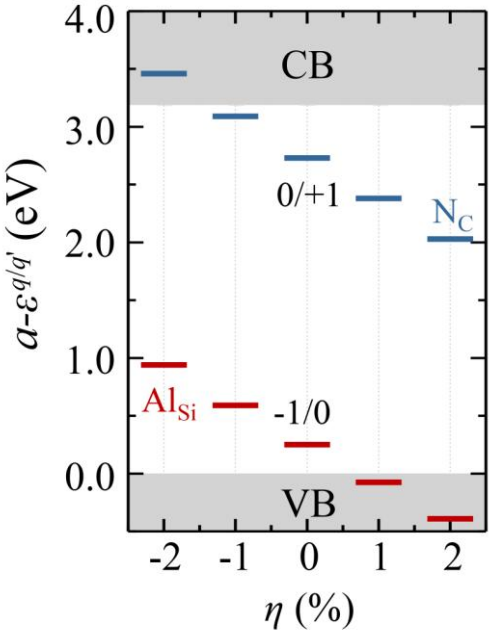

Fig. S8. a-$\varepsilon^{0/-1}(\eta)$ [a-$\varepsilon^{0/+1}(\eta)$] of Al_{Si} (N_C) in SiC as a function of strain η. Band edge positions are fixed at the values of unstrained SiC.

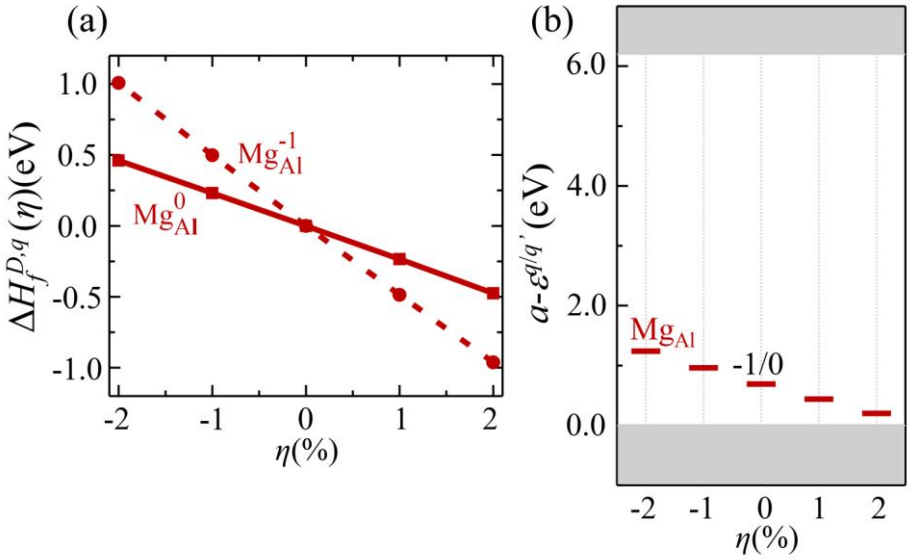

Fig. S9. (a) Change of formation energies $\Delta H_f^{D,q}(\eta)$ for Mg_{Al} in AlN as a function of strain η. (b) a-$\varepsilon^{0/-1}(\eta)$ of Mg_{Al} as a function of strain η. Band edge positions are fixed at the values of unstrained AlN.

References

[1] Kresse G, Furthmuller J 1996 *Phys. Rev. B Condens Matter* **54** 11169.

[2] Perdew J P, Burke K, Ernzerhof M 1996 *Phys. Rev. Lett.* **77** 3865.

[3] Heyd J, Scuseria G E 2003 *J. Chem. Phys.* **118** 8207.

[4] Wei S-H 2004 *Comput. Mater. Sci.* **30** 337.

[5] Freysoldt C, Neugebauer J, Van de Walle C G *Advanced Calculations for Defects in Materials:*
Electronic Structure Methods, (Wiley-VCH Verlag GmbH & Co. KGaA, 2011).

[6] Zhang S B, Northrup J E 1991 *Phys. Rev. Lett.* **67** 2339.

[7] Laks D B, Van de Walle C G, Neumark G F, *et al.* 1992 *Phys. Rev. B Condens Matter* **45** 10965.

[8] Freysoldt C, Grabowski B, Hickel T, *et al.* 2014 *Rev. Mod. Phys.* **86** 253.