15 AUGUST 1987-II

## Defect dynamics and the Staebler-Wronski effect in hydrogenated amorphous silicon

Sokrates T. Pantelides

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598 (Received 11 February 1987)

It is shown that four fundamental reactions involving floating bonds, dangling bonds, and H govern defect dynamics in a-Si:H under equilibrium and various nonequilibrium conditions. The Staebler-Wronski effect is a natural consequence of these reactions in the presence of excess electrons and holes. It is predicted that, under illumination or particle irradiation, both floating and dangling bonds can be created. Experimental data support this prediction and provide characteristic signatures for both defects. Though H may not be involved in the creation of the metastable defects, it plays a key role in their annealing.

Ten years ago, Staebler and Wronski<sup>1</sup> (SW) discovered that prolonged illumination of hydrogenated amorphous Si (a-Si:H) leads to a significant drop in conductivity that can be traced to an increase in the defect density. The metastable excess defects were found to anneal above 150 °C with an activation energy of  $\sim 1.5$  eV. Subsequent luminescence<sup>2</sup> and EPR<sup>3</sup> data led to the conclusion that the light-induced metastable defects responsible for the SW effect are the well-known D centers, commonly believed to be dangling bonds. Similar phenomena were reported in samples irradiated with electrons or ions.<sup>2,4</sup> Several models have been proposed to explain how dangling bonds are created by light, typically based on the notion that preexisting negative-UD centers capture an excess electron or hole and become magnetic or the notion that recombination of light-induced electron-hole pairs causes the breakup of weak Si-Si bonds. Recently, Stutzmann, Jackson, and Tsai<sup>5</sup> reviewed these models and concluded that the experimental data favor the breakup of weak Si-Si bonds. The difficulty with that model is the need to keep the two opposing dangling bonds from rebonding. Stutzmann et al. favored the suggestion<sup>3</sup> that an H atom from a neighboring Si-H bond switches over into the broken bond. Subsequent experiments designed to test for a role by H, however, revealed no positive evidence.<sup>5</sup> Furthermore, in this model, only half of the new defects are true dangling bonds. The other half are Si -H-Si centers, perhaps with the H shifted off center, whose EPR signature is likely to be different.

In addition to questions regarding the mechanisms for the creation of radiation-induced defects, questions also persist about the origins of the defects in as-grown samples. Are defects created only during growth or are there mechanisms for creating defects thermally after deposition so that equilibrium can be established by a balance of the creation and annealing rates? In a recent theory, Bar-Yam, Adler, and Joannopoulos<sup>6</sup> implicitly assumed that such mechanisms exist, but did not identify the specific reactions. Smith and Wagner<sup>7</sup> independently suggested that thermally generated electron-hole pairs create D centers at the same rate as light-induced electron-hole pairs in the SW effect. They calculated equilibrium concentrations from kinetic considerations without assuming a microscopic model for either the creation or annealing of defects and found them in agreement with experimental data. The microscopic processes that lead to equilibrium have not, however, been elucidated.

All attempts to explain the above phenomena have so far been based on the notion that the D centers are dangling bonds. Other defects involving undercoordinated Si atoms have occasionally been mentioned, but overcoordinated Si atoms have been universally overlooked. I recently pointed out<sup>8</sup> that overcoordination is possible and that, in principle, threefold- and fivefold-coordinated atoms are the primitive conjugate intrinsic defects in a-Si, like vacancies and self-interstitials are the primitive conjugate intrinsic defects in crystalline Si. It was shown that hyperfine EPR data favor the suggestion that the Dcenters are fivefold-coordinated Si atoms with an unpaired electron in a state called "floating bond." The existence of dangling bonds was not, however, ruled out. In a later paper,<sup>9</sup> I showed that overcoordination is also the key to understanding defect migration, which, in turn is the key to understanding many peculiar phenomena that occur in a-Si:H at low temperatures and have not been explained satisfactorily in terms of existing concepts. In this paper, it will be shown that overcoordination and floating-bond migration are also the keys to understanding defect generation, the SW effect, and various other nonequilibrium and equilibrium phenomena. It will be shown that, during prolonged illumination or particle irradiation, both dangling and floating bonds can be created and that available experimental data support this prediction. A characteristic signature for each defect is obtained.

We begin with a summary of the main results of Ref. 9 that are needed in this paper. As shown in Fig. 1, floating bonds can migrate very efficiently by bond switching. No particular fivefold-coordinated Si atom is migrating through the network. Instead, the extra bond is passed from one Si atom to another with only small movements by a few atoms. In contrast, dangling bonds pointing into voids have no obvious low-energy mode of migration. "Bulk" dangling bonds can migrate only by converting to floating bonds (see Refs. 8–10) so that, in effect, one needs to consider only immobile dangling bonds and mobile floating bonds, bearing in mind the possibility of interconversion. In the remainder of this paper, we will assume that D centers are floating bonds, an assumption that will be further strengthened by new conclusions.

In Ref. 9, it was shown that migrating floating bonds react with Si-H bonds and release H in an interstitial

3480



FIG. 1. Schematic illustration of floating-bond migration by bond switching (From Ref. 9).

configuration. In turn, the same reaction can proceed in the reverse direction or interstitial H will run into a floating bond and form an Si-H bond. Thus, the two basic reactions between floating bonds, Si-H bonds, and interstitial H are

$$FB+SiH \neq \Box + H_i, \tag{1}$$

$$H_i + FB \rightleftharpoons \Box + SiH.$$
 (2)

Here, FB denotes a floating bond,  $H_i$  denotes interstitial H, <sup>11</sup> and  $\Box$  denotes a fourfold-coordinated network.

For a complete description of defect dynamics we now supplement reactions (1) and (2) with two more reactions. The third reaction is spontaneous "Frenkel"-pair creation, i.e.,

$$\Box \rightleftharpoons FB + DB, \tag{3}$$

where DB stands for a dangling bond. The reverse reaction, which is perhaps easier to visualize, is the mutual annihilation that occurs when a migrating FB encounters a DB. The reaction, viewed in either direction, is illustrated schematically in Fig. 2. Since the two types of intrinsic defects can in principle convert into one another,<sup>8-10</sup> it could be argued that the net result may be two identical defects, either floating or dangling bonds. Thus, one can, in principle, write reactions similar to (3) where the right-hand side is 2FB or 2DB. Steric constraints, however, arising from ring sizes favor the creation of a FB-DB pair. For the same reason, it is sterically easier for a FB to run into a DB and annihilate it Ireaction (3) in the reverse direction], than it is for two FB's to run into and annihilate each other.

The fourth and last reaction is hydrogenation of dangling bonds, i.e.,

$$H_i + DB \rightleftharpoons \Box + SiH. \tag{4}$$

Note, however, that there is no analog of Eq. (1) for DB's



FIG. 2. Schematic illustration of dangling- and floating-bond pair creation or annihilation by bond switching.

because they are not mobile. Furthermore, in a more detailed theory, one can distinguish between two types of Si— H bonds, one that breaks up into a DB [reverse of reaction (4)] and one that breaks into a FB [reverse of reaction (2)]. That distinction will not be pursued here.

Let us now consider several nonequilibrium situations.

(i) When excess D centers (FB's) are present, reactions (1) and (2) are driven in the forward direction. This is the "self-destructive mediation" discussed in Ref. 9. If excess DB's are also present, they are annealed by reaction (4) being driven in the forward direction. The Si— H bond density remains constant during this process.<sup>12</sup> Thus, H acts as a catalyst for the mutual annihilation of excess FB's and DB's without a need for direct encounters among them, but self-mobility by at least one of the two intrinsic defects is essential for the process to occur.

(ii) During H evolution, reaction (1) is driven in the forward direction and reactions (2) and (4) are driven in the reverse direction (breakup of Si—H bonds).<sup>13</sup> Note, however, that FB's *created* by reaction (2) migrate and *self-destruct* via reaction (1), releasing more H<sub>i</sub>. Thus, reactions (1) and (2) provide an elegant explanation for the puzzling observation<sup>14</sup> that, when all H has evolved, the number of D centers is significantly smaller than the total amount of H that evolved.

(iii) If H is diffused into pure *a*-Si, reaction (1) is driven in the reverse direction, whereas reactions (2) and (4) are driven in the forward direction. Thus, in-diffusion of H in pure *a*-Si passivates DB's quite effectively; it also passivates FB's yielding Si—H bonds, but also *creates FB's*, *again yielding* Si—H bonds. (Physically, interstitial H attaches itself weakly to a fourfold-coordinated Si, making it fivefold coordinated; a Si—Si bond then migrates away, leaving behind a Si—H bond.) When H indiffusion stops, remaining FB's are annealed via selfdestructive mediation<sup>9</sup> [reactions (1) and (2)] without affecting the Si—H bond density. Thus, again, our basic reactions provide an elegant explanation for the observation<sup>15</sup> that large amounts of H are needed to passivate relatively small concentrations of D centers.

(iv) If at room temperature, where all four reactions are slow, we supply electron-hole (e-h) pairs via illumination or particle irradiation, the energy stored in these pairs can in principle be used to enhance either defect migration or the rate of a particular reaction. The suggestion discussed by Stutzmann et al.<sup>6</sup> as an explanation of the SW effect corresponds to e - h pairs enhancing reaction (4) in the reverse direction, where  $H_i$  is in the bond-centered configuration. By viewing the D centers as FB's, the analogous process would be an enhancement of reaction (2) in the reverse direction. In that case,  $H_i$  might immediately yield an FB and an SiH [reverse of reaction (1)]. The net result would be the creation of two FB's while moving an Si-H bond to an adjacent site. Such a process, however, requires that several distinct steps occur simultaneously. A more attractive possibility is the enhancement of reaction (3) in the forward direction which can be written as a two-step process:

$$0 \to e + h, \tag{3a}$$

$$\Box + e + h \longrightarrow FB + DB. \tag{3b}$$

The underlying physics is that the creation of an FB-DB pair is enhanced when there is a hole in a state that is significantly localized at a particular Si-Si bond, e.g., in one of the valence-band-tail states, and an electron in the corresponding antibonding state in the conduction-band tail. Under such conditions, that particular bond is effectively broken. Bond switching, as illustrated in Fig. 2, and migration of the floating bond by at least one step stabilize the pair as isolated defects. We propose that this is the essence of the SW effect. Once these defects can be created, they remain stable unless heated to the point when long-range migration of floating bonds is appreciable, which leads to their annealing via self-destructive mediation [reactions (1) and (2)]. Thus, the SW excess defects are created without H involvement, but their annealing is mediated by H. We also note the following about the precise role of the e-h pairs. It is generally believed that the SW effect is driven by the recombination of *e-h* pairs, which provides the energy needed to overcome the reaction barrier for bond breakup. Such a process would require that all of the recombination energy be channeled into a particular local phonon mode. A more attractive possibility is that the e-h pair places a weak bond in an excited state (one electron in the bonding state, one in the antibonding state) whose energy is roughly degenerate with that of a FB-DB pair. In other words, the local excitation effectively eliminates the barrier for reaction (3) in the forward direction. Kinetic studies of the SW effect cannot distinguish between the two possibilities because the defect creation rate is in both cases proportional to the pn product. The observation that excess holes alone induce a SW-like effect,<sup>5</sup> however, supports the local-excitation model: two holes in a weak bond (no electron in either the bonding or the antibonding state) also make its energy roughly degenerate with that of a FB-DB pair.

According to the above discussion, illumination and particle irradiation create both DB's and FB's. In Refs. 8 and 9 and in this paper we argued that the paramagnetic D centers are FB's. What is then the role of DB's? One possibility is that they all convert to FB's.<sup>8-10</sup> Careful examination of the experimental literature, however, reveals ample evidence that a second defect is created alongside the paramagnetic D centers in both the SW effect and during particle irradiation. The key observations<sup>2,4</sup> are the following: (a) tail-state luminescence ( $\sim 1.2 \text{ eV}$ ) in irradiated samples is significantly lower than in normal samples containing the same concentration of D centers; and (b) a new luminescence peak emerges at  $\sim 0.9$  eV. The simplest explanation of these data is a second nonmagnetic defect. Indeed, Street, Biegelsen, and Stuke<sup>4</sup> in 1979 noted that "the data indicate that there are defects that quench the luminescence but do not contain unpaired spins," and added that they have "no information as to the detailed structure of this spinless defect." A second defect was also invoked by Voget-Grote et al.<sup>4</sup> in 1980 in the context of luminescence and photoconductivity data. Schade and Pankove<sup>2</sup> in 1981 independently suggested that two defects are needed and attributed the  $\sim 0.9$ -eV luminescence peak to the "second" defect, without further identification. In later years, however, the two-defect idea remained dormant (see, e.g., Ref. 5), as no credible candidates were available. Our analysis (Refs. 8 and 9 and this paper) provides a natural and elegant solution of the elusive "second-defect" puzzle: The two defects are floating and dangling bonds. Well-annealed samples have a detectable density of only FB's (*D* centers) because DB's are more effectively annealed via reaction (4). During prolonged illumination and particle irradiation, FB's and DB's are created simultaneously via reaction (3b). The characteristic signature of FB's is the EPR signal at g = 2.0055, whereas the characteristic signature of DB's is the luminescence peak at ~0.9 eV. In order to preserve tradition, we shall continue referring to the paramagnetic FB's as *D* centers and introduce the name *L* center for the luminescent DB's.

The above identification of two defects in *a*-Si is consistent with extensive experimental information that has been examined by the author so far. For example, the *L* center is seen only in illuminated and irradiated samples [where it is created by reaction (3b)], after high-temperature evolution of H, which, as we saw in item (ii) above, leaves behind both floating and dangling bonds, and in *n*-type or *p*-type *a*-Si (not in co-doped *a*-Si), where we again predict<sup>16</sup> that doping (but not co-doping) induces both floating and dangling bonds.

The above analysis requires that L centers are not paramagnetic, which is consistent with recent calculations<sup>17</sup> that favor a negative correlation energy U. Thus, upon their creation, dangling bonds immediately capture an electron or hole and become negatively or positively charged, respectively. When a sufficient number of them is present, light-induced EPR, especially hyperfine structure in <sup>29</sup>Si-enriched samples might reveal an EPR signature for these defects as well.

We should caution, however, that the evidence for two defects is not systematic and comes primarily from particle-irradiated samples. Because of the possibility of interconversion of "bulk" dangling bonds, the only dangling bonds that can survive may be those of voids. Clearly, simultaneous monitoring of the EPR signal and the 0.9-eV luminescence peak during prolonged illumination and comparisons with normal samples containing the same spin densities would be crucial in establishing whether the *D* and *L* centers are two distinct centers or interconversion leads to only one of the two defects being responsible for both the EPR and luminescence signals.

Finally we briefly discuss the question of equilibrium that is raised by the work of Smith and Wagner.<sup>7</sup> These authors postulated that thermally generated electron-hole pairs generate D centers at the rate that occurs during the SW effect. Their kinetic modeling did not require them to assume a microscopic model for either the creation or annealing process, except for an assumption about the order of the annealing reaction. The present analysis confirms their basic postulate regarding the defect creation rate, but also reveals a more complex annealing process. It would be desirable to carry out kinetic modeling of data based on reactions (1)–(4), but such a task is beyond the scope of this paper.

If equilibrium is achieved, the mass-action law yields equilibrium concentrations of floating bonds, dangling 3482

bonds, and interstitial H (compare also Ref. 6):

$$[FB] = Np_{FB} \exp(-E_{FB}/kT), \qquad (5)$$

$$[DB] = Np_{DB} \exp(-E_{DB}/kT), \qquad (6)$$

$$[H_i] = [SiH]_{p_H} \exp(-E_H/kT), \qquad (7)$$

where N is the density of fourfold-coordinated Si atoms;  $E_{\rm FB} = (E_2 - E_1)/2; \ E_{\rm DB} = (E_3 - E_{\rm FB})/2; \ E_{\rm H} = (E_2 + E_1)/2$ 2;  $E_1$ ,  $E_2$ , and  $E_3$  are the equilibrium energies of reactions (1), (2), and (3), respectively, and the p's are entropy factors. Using a variety of low-temperature experimental data we arrive at the following *tentative* values:  $E_{FB} \sim 0.6$  eV,  $E_{DB} \sim 0.8$  eV,  $E_{H} \sim 1$  eV.<sup>18</sup> Thus, the predicted density of  $H_i$  is too small to be detectable by EPR, in agreement with experiments (for 10 *at*. % H, at 300 °C,  $[H_i] \sim 10^{13}$  cm<sup>-3</sup>). The predicted equilibrium *D*-center densities (e.g.,  $\sim 10^{17}$  cm<sup>-3</sup> at 300 °C) are in good agreement with data quoted in Ref. 7. Equilibrium L-center densities are one to two orders of magnitude smaller. We should caution, however, that even quasiequilibrium may be difficult to achieve in practice. At temperatures below 300 °C where no H evolution occurs, the reverse of reactions (2) and (4) may occur too slowly for equilibrium to be reached on practical time scales. At higher temperatures, H evolution occurs so that H<sub>2</sub> formation should be taken into account and the solid must be at equilibrium with the gas.

In summary, we provided a description of defect dynamics in a-Si and, on the basis of available data, we as-

- <sup>1</sup>D. L. Staebler and C. R. Wronski, Appl. Phys. Lett. **31**, 292 (1977); J. Appl. Phys. **51**, 3263 (1980).
- <sup>2</sup>J. I. Pankove and J. E. Berkeyheiser, Appl. Phys. Lett. 37, 705 (1980); H. J. Schade and J. I. Pankove, J. Phys. (Paris) Colloq. 42 C4-327 (1981).
- <sup>3</sup>H. Dersch, J. Stuke, and J. Beichler, Appl. Phys. Lett. 38, 456 (1981).
- <sup>4</sup>R. Street, D. Biegelsen, and J. Stuke, Philos. Mag. B 40, 451 (1979); U. Voget-Grote, W. Kuemmerle, R. Fischer, and J. Stuke, *ibid.* 41, 127 (1980).
- <sup>5</sup>M. Stutzmann, W. B. Jackson, and C. C. Tsai, Phys. Rev. B **32**, 23 (1985); M. Stutzmann *et al.* Appl. Phys. Lett. **48**, 62 (1986).
- <sup>6</sup>Y. Bar-Yam, D. Adler, and J. D. Joannopoulos, Phys. Rev. Lett. 57, 467 (1986).
- <sup>7</sup>Z. E. Smith and S. Wagner, Phys. Rev. B 32, 5510 (1985);
  Z. E. Smith *et al.*, Phys. Rev. Lett. 57, 2450 (1986).
- <sup>8</sup>S. T. Pantelides, Phys. Rev. Lett. 57, 2979 (1986).
- <sup>9</sup>S. T. Pantelides, Phys. Rev. Lett. 58, 1344 (1987).
- <sup>10</sup>P. A. Fedders and A. E. Carlsson, Phys. Rev. Lett. **58**, 1156 (1987), showed that conversion of DB's into FB's is favored energetically. However, DB's in voids do not convert and some bulk DB's may be sterically constrained. There may also be defects of intermediate structure, but the width of the hyperfine lines (see Ref. 8) suggests that the dominant defect is quite distinct.

cribed a distinct role to both dangling and floating bonds. The suggestion that the paramagnetic center in *a*-Si is a floating bond, first made on the basis of hyperfine data,<sup>8</sup> is now corroborated by our ability to explain a broad range of phenomena which could not be explained by conventional concepts. Additional data are needed to test further the existence of two defects. In either case, the overall conclusion is that overcoordination and the resulting ability of *D* centers to migrate efficiently are the key to understanding *a*-Si.

Note added in proof. Recently, H. Yokomichi and K. Morigaki (unpublished) have reported ENDOR-detected hyperfine EPR data (ENDOR dentoes electron-nuclear double resonance) that suggest that the EPR signal of Dcenters is sometimes a superposition of two distinct signals. They propose that one of these may be due to Si -H-Si centers (one of the possible  $H_i$  configurations in our notation). Definitive identification (e.g., by deuterium substitution or data in intrinsic a-Si) is needed. In any case, the four fundamental reactions described here remain intact. Their relative significance, however, needs to be updated as new data become available. For example, evidence that  $H_i$  are, indeed, created during the SW effect would constitute evidence that light drives reactions (2) and/or (4) in the reverse direction, which is equivalent to the model described in Refs. 3 and 5.

This work was supported in part by U. S. Office of Naval Research Contract No. N00014-84-C-0396.

- <sup>11</sup>Several paths are possible for  $H_i$ . For example, through the bonds, through the *n*-member (n = 5, 6, 7) rings, etc. In reaction (1), the reverse reaction may proceed immediately, so that in effect the Si— H bond is moved one step while the FB goes on.
- <sup>12</sup>The experimental data mentioned in Ref. 9 are consistent with this prediction but are probably not sensitive enough to detect changes in the Si-H bond density because the initial *D*center density is about two orders of magnitude smaller.
- <sup>13</sup>The early H evolution stage (see Ref. 9) is initially due to reaction (1), but, as the FB supplies are depleted, H evolution is fueled by reactions (2)-(4). Early H evolution may also be due to H<sub>2</sub> trapped in voids.
- <sup>14</sup>H. Fritzsche, M. Tanielian, C. C. Tsai, and P. J. Gaczi, J. Appl. Phys. **50**, 3366 (1979).
- <sup>15</sup>D. Kaplan, N. Sol, G. Velasco, and P. A. Thomas Appl. Phys. Lett. **33**, 440 (1978).
- <sup>16</sup>S. T. Pantelides (unpublished).
- <sup>17</sup>Y. Bar-Yam and J. D. Joannopoulos, Phys. Rev. Lett. 56, 2203 (1986).
- <sup>18</sup>Migration of H<sub>i</sub> yielding these values is likely to be through the bonds. The high-temperature break-up of Si— H bonds is more likely to release H<sub>i</sub> migrating through the *n*-member rings, but the activation energy of  $\sim 3.5$  eV reported by Zellama *et al.* (see Ref. 9) is an overestimate because reaction (1) was not taken into account in the data modeling.