

Order in disorder: Increased carrier mobility of downscaled amorphous semiconductors as exemplified by hydrogenated amorphous silicon

Yuezhou Luo^{✉*} and Andrew John Flewitt[✉]

Electrical Engineering Division, Department of Engineering, *University of Cambridge*, Cambridge CB3 0FA, United Kingdom



(Received 20 July 2025; revised 19 September 2025; accepted 17 October 2025; published 21 November 2025)

Amorphous semiconductors are important channel semiconductors in thin-film transistors (TFTs), which serve not only active-matrix displays, but also flexible electronics for Internet of things applications. Nevertheless, a great limitation of amorphous semiconductors is their low carrier mobilities relative to their monocrystalline counterparts. Based on a recently established band-fluctuation framework [Y. Luo and A. Flewitt, *Phys. Rev. B* 109, 104203 (2024)], this paper shows that the intrinsic carrier mobility of amorphous semiconductors can significantly increase simply through device downscaling, without any material-level optimizations. Specifically, it is revealed that the intrinsic electron mobility of hydrogenated amorphous silicon in a 10-nm-long gap can increase by about 12 times, and this does not compromise device-to-device uniformity. This mobility improvement is a result of reduced localized band-tail states due to the ultrashort gap length relative to the band-fluctuation length scale before downscaling; the latter is determined by the short- and medium-range structural order of the amorphous semiconductor.

DOI: [10.1103/n66s-qw44](https://doi.org/10.1103/n66s-qw44)

I. INTRODUCTION

Thin-film transistors (TFTs), traditionally known as the building blocks of active-matrix displays, now play an increasingly important role in Internet of things (IOT) hardware that is seamlessly integrating into the objects in our daily life [1]. In contrast to silicon-based metal-oxide-semiconductor field effect transistor devices used in mature integrated circuits (ICs), many of the fascinating merits that TFTs have offered to the IOT, such as large-area low-cost fabrication and low thermal budget, result from the properties and fabrication process of their channel semiconductors, which are typically noncrystalline materials [2,3], with amorphous semiconductors being the most commercially influential representatives. Well-known examples include hydrogenated amorphous silicon (*a*-Si:H), which demonstrated its success in early liquid-crystal displays in the last century [4], as well as amorphous oxides (such as amorphous indium gallium zinc oxide, *a*-IGZO), which have been increasingly dominant since 2004 [5]. Speed, which was not a top priority for TFTs in the past, is now of growing significance in light of

the soaring development of IOT technologies, as exemplified by flexible electronics [6,7]; this is especially so when it comes to emerging fields, such as flexible microprocessors [8], flexible machine learning engines [9], and novel memory architectures for edge computing [10]. Amorphous semiconductors, however, are intrinsically disadvantaged in speed due to their low carrier mobility. For instance, the electron mobility of *a*-Si:H is typically limited to about 1 cm²/(V s) [11]. Even for amorphous oxides that exhibit higher mobilities [such as ~ 10 cm²/(V s) mobility of *a*-IGZO [5]], their mobility values are still very far from the expectations of standard IC electronics. Other material options available for TFTs, such as high-mobility polycrystalline and crystalline thin films, are at the cost of poor device uniformity, higher expenses, higher thermal budget, and process complexity.

Great efforts have been made to improve the mobility of amorphous semiconductors. Some endeavors focused on the material level to improve the intrinsic mobility. These, for *a*-Si:H, include the use of disilane with a controlled flow rate [12], the technique of expanding thermal plasma [13], and the process of catalytic chemical vapor deposition [14]. For amorphous oxides such as *a*-IGZO, increasing the carrier concentration through elemental controls [15,16] and decreasing the defect density through metal capping [17] were attempted. Other endeavors focused on the device level and aimed to improve the apparent field effect mobility through, for instance, optimizing semiconductor-insulator interface conditions [18–23] and

*Contact author: YL778@cantab.ac.uk

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International* license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

decreasing the parasitic resistance [24]. Nevertheless, none of the reported methods so far present an effective and reproducible way to fundamentally overcome the mobility issue; also, many of them often compromise other material merits such as stability, thermal budget, and process simplicity. An alternative strategy is thus needed in light of the technological importance of amorphous semiconductors in the era of the IOT.

One well-known factor causing low mobilities of amorphous semiconductors is the localized states in the density of states (DOS) distribution band tails; this was elaborated in our recent work [25], henceforth named “Paper 1.” In Paper 1, we used *a*-Si:H as an example and reinvestigated the essence of localized band-tail states from the perspective of excess valence-band and conduction-band (CB) charges (termed “excess delocalized charges”), which induced extra potentials that forced the crystalline silicon (*c*-Si) band edge to fluctuate. Figure 1(a) shows the band-edge fluctuation modeled in Paper 1, from which the fluctuation along a representative direction is extracted and shown in Fig. 1(b). As revealed in Paper 1, because of the existence of short- and medium-range structural order in amorphous semiconductors, the band fluctuation is slow. Here, the Fourier decomposition in Fig. 1(c) reveals that the predominant (>95%) Fourier components are below

a cutoff spatial frequency of $0.012a_0^{-1}$, where a_0 is the short-range order length (3 Å) defined in Paper 1. This corresponds to a length scale of 25 nm. We therefore envisage that, similar to a high-pass filter, downscaling below such a length scale would selectively retain only the high-frequency fluctuations with low amplitudes, therefore mitigating the overall band fluctuation. *a*-Si:H would then contain fewer localized band-tail states, according to the analysis in Paper 1, and thus, exhibit higher mobility.

The above hypothesis of band-tail engineering by means of downscaling also meets the general trend in the semiconductor industry; beside ICs, downscaling has also received much attention for TFTs. Nevertheless, existing studies did not straightforwardly corroborate our hypothesis (e.g., Refs. [26–28]). This is due to two facts. First, it was the apparent field effect mobility (rather than the intrinsic mobility) that was investigated; the former is affected by nonideality factors such as contact resistance. Second, the staggered TFT structures were adopted, as illustrated in Fig. 1(d), where the actual channel length is partly correlated with the unscaled gate length, and thus, is not being *strictly downscaled*. The *intrinsic mobility* after *strict downscaling* remains an underexplored topic. It should be highlighted that, with numerous existing strategies, nonideality factors could eventually be suppressed

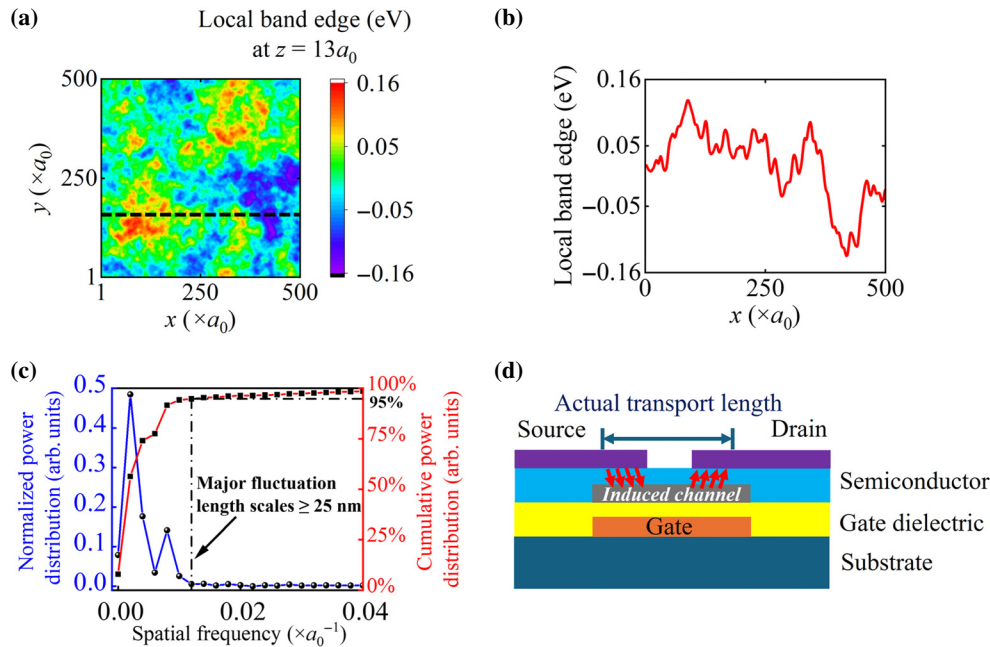


FIG. 1. Preliminary justification of increased mobility through downscaling. (a),(b) Band edge spatial distribution of the model developed in Paper 1 [25]. $150 \times 150 \times 7.5$ nm³ model simulates a *a*-Si:H thin film; (a) shows the result across the plane at $z = 13a_0$, where $a_0 = 3$ Å is the short-range order length. One-dimensional fluctuation along the dashed line in (a) is extracted in (b). (c) Fourier decomposition of (b), showing that more than 95% contribution to the fluctuation in (b) comes from components with spatial frequencies less than $0.012a_0^{-1}$. (d) Illustration of one of the reasons why many previous researchers did not observe increased mobility after downscaling. In staggered TFTs, the gate is not effectively downscaled, leading to a longer carrier path than the source-drain separation.

to a satisfactorily low level; it is then the *intrinsic mobility* that would be the limitation. Compared with the large engineering space of nonideality factors, the intrinsic property is indeed much more difficult to deal with. Our hypothesis appears to be a promising strategy, but validation is needed.

Therefore, a quantitative and rigorous bottom-up effort is made in this paper. First, a basic device configuration is defined in Sec. II, in the context of which the downscaling of amorphous semiconductors, as exemplified by a -Si:H, is investigated. Section III resolves the motion and transport of carriers in downscaled a -Si:H through statistical analysis, leading to analytical quantification of the intrinsic carrier mobility. Section IV uses the analytical results in Sec. III and the modeling framework from Paper 1, to simulate the mobility after downscaling. A remarkable mobility improvement of about 12 times at a length scale of 10 nm is shown, and a satisfactorily low device-to-device variation is revealed. In the end, it is concluded that the essence of the mobility improvement is a result of reduced localized band-tail states due to the ultrashort material length scale relative to the band-fluctuation length scale set by the short- and medium-range order of the amorphous semiconductor.

II. THE CONTEXT OF DOWNSCALING: THE DEVICE STRUCTURE

As shown in Fig. 2(a), a basic device structure named the “nanogap” is proposed. This is a structure where a pair of metal electrodes are separated by a nanoscale length L (ideally <25 nm, as envisaged earlier), in which a -Si:H is filled. The width (y direction) and thickness (z direction) are of the same scales as the source and drain electrodes in an ordinary TFT. a -Si:H exists only inside the gap, and thus, is being *strictly downscaled*. To implement such strict downscaling in practical TFTs, only coplanar structures, as exemplified by Fig. 2(b), may be adopted. The nanoscale spacing may be fabricated using recent low-cost scalable processes, such as adhesion lithography [29–32], as an

alternative to slow and expensive electron-beam lithography. The gate field penetration in the TFT should be satisfactory, on condition that the dielectric layer is thin and that the source and drain thicknesses are comparable to or less than L . For the purpose of this paper, we focus only on the intrinsic dc carrier mobility in Fig. 2(a) under a voltage U .

III. CHARGE-TRANSPORT ANALYSIS

For a -Si:H, electron carriers play a much more important role in conduction than hole carriers [11], and so the focus of this paper is only on electron carriers. Given that hydrogen passivates the majority of dangling bonds in high-quality a -Si:H [33], and that the volume of a -Si:H in the nanogap is too small to be able to involve even a single defect, in this paper we approximately treat a -Si:H as being *defect free*. In the absence of defect states, the localized states in a -Si:H are, in essence, band-tail states that originate from band fluctuation, as analyzed in Paper 1 [25]. In our other previous work (henceforth named Paper 2 [34]), we found that, instead of being *completely confined*, electron carriers still undergo local motion, even when they are in localized band-tail states. Such microscopic motion is not fundamentally different from that of extended-state electrons, both of which are being scattered from disorder. Therefore, regardless of the state of the electron carriers, their paths take a zigzag shape due to frequent scattering, the mean free path of which is of interatomic spacing [11] ($\approx$ the short-range order length a_0). This is illustrated in Fig. 3(a). Such random thermal motion acts as noise that does not contribute to net conduction. An imagined “drift center” is thus introduced in Fig. 3(b) to describe the net drift motion of every electron carrier under an applied field. Acting as a reference point, the drift center of an electron carrier is defined in such a way that it is exactly at the location of the actual carrier when the x coordinate of the actual carrier first reaches the middle of the nanogap (i.e., $x = x_0$); this is illustrated in Fig. 3(b). Except the case of $x = x_0$,

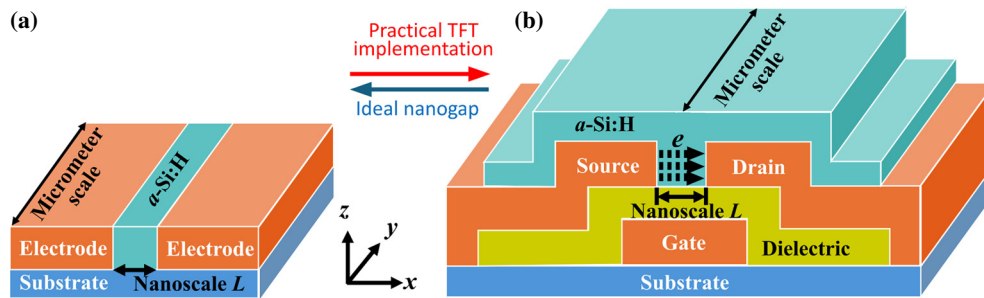


FIG. 2. Device structure and practical implementation. (a) Structure of the nanogap where a -Si:H is strictly downscaled. Length of the nanogap is on the nanoscale, whereas the width and thickness are of standard scales as in ordinary TFTs. (b) Practical implementation of the nanogap in TFTs. Coplanar structure must be used to produce a channel that is strictly downscaled (illustrated by dashed arrows).

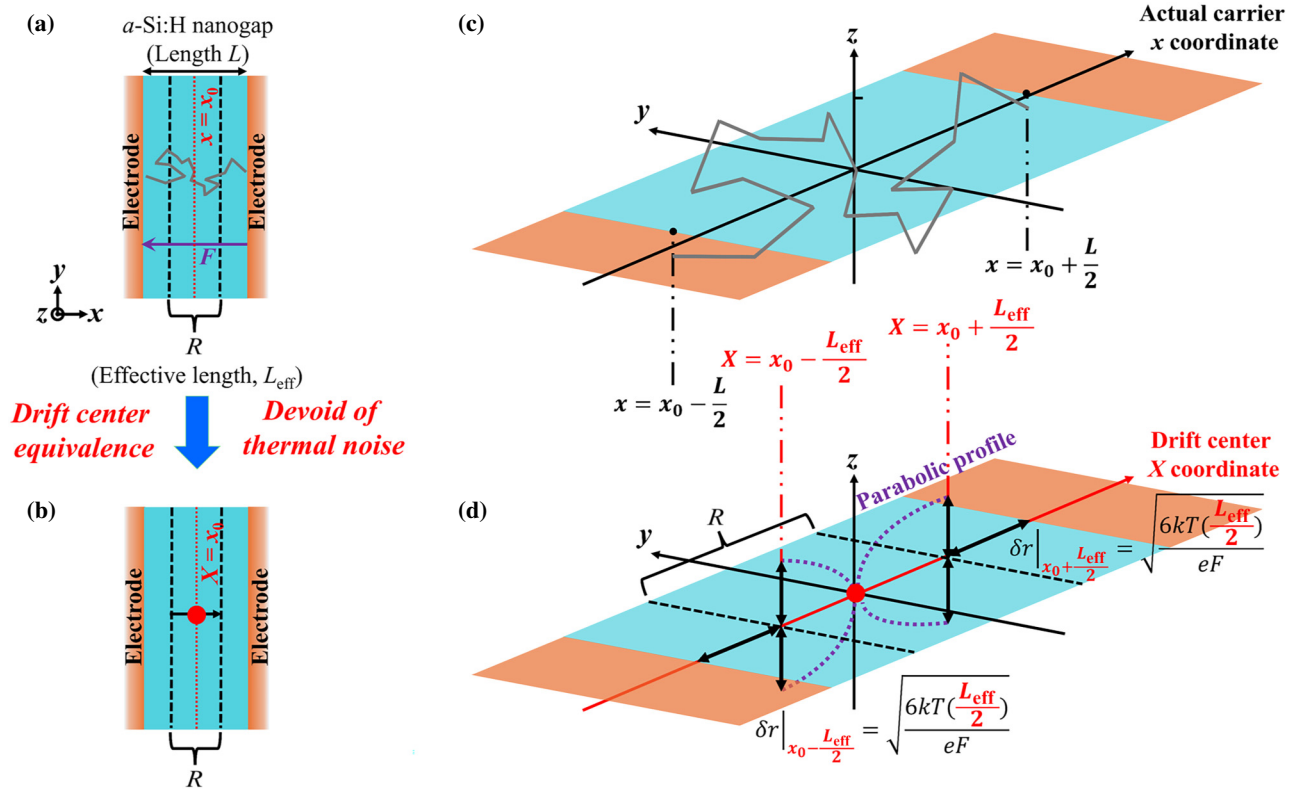


FIG. 3. Analysis of the random thermal motion and drift motion of electron carriers in the nearly defect-free a -Si:H nanogap. (a) Illustration of the actual zigzag path of a carrier under an applied field F . (b) Description of net drift motion based on an idealized drift center. Drift center (red sphere) is defined in such a way that it is at the exact location of the actual carrier when the latter first reaches the middle of the gap (*i.e.*, $X = x_0$). (c) Three-dimensional (3D) illustration of the actual carrier path. During the period when the drift center drifts in a region named R [see also (a) and (b)], the carrier is narrowly able to reach the metal-semiconductor interfaces. (d) Quantification of the uncertainty of carrier location. Before and after the drift center reaches $X = x_0$, the root-mean-square (RMS) deviation (δr) of the actual carrier from the drift center depends on the location of the drift center in a parabolic fashion (dotted purple curves). Region R is specified in such a way that, when the drift center reaches either of the boundaries of R , the farthest RMS deviation is just able to reach the metal-semiconductor interfaces.

the actual location of the carrier is uncertainly distributed around the drift center.

To quantitatively analyze this uncertainty, the classic random walk theorem is utilized. As revealed in Paper 2 [34], electrons at different energies experience different extents of spatial confinement. Thus, if the transit of a carrier through the a -Si:H nanogap includes more than one energy state, such a transit should be modeled as a random walk with multiple phases, each of which lasts for time t_m and possesses a unique diffusion coefficient, D_m , reflecting the extent of spatial confinement. Based on Einstein's diffusion theory [35] and probability theories, the RMS displacement of a carrier from its initial location in 3D space is

$$\delta r = \sqrt{\sum_m 6D_m t_m}. \quad (1)$$

Under an applied electric field (F), the diffusion coefficient (D_m) is related to the drift mobility (μ_m) of the carrier in the m th phase via the Einstein relationship:

$$D_m = \frac{kT}{e} \mu_m, \quad (2)$$

where k is the Boltzmann constant, T denotes temperature, and e is the elementary charge. Furthermore, the drift mobility is associated with the drift velocity (v_m) via

$$v_m = \mu_m F. \quad (3)$$

In total, the carrier completes a *net drift* opposite to the field by a distance X that is

$$X = \sum_m v_m t_m. \quad (4)$$

Substituting Eqs. (2)–(4) back into Eq. (1), the total RMS displacement is expressed as

$$\delta r = \sqrt{\frac{6kTX}{eF}}, \quad (5)$$

which relates the net drift to the superimposed diffusion. This is illustrated by the dotted evolution contour in Fig. 3(d) in a 3D view [to be viewed alongside Fig. 3(c)]. The period before the drift center reaches $X = x_0$ is simply treated as a reverse process, and so, the location uncertainty is similarly quantified, showing a mirrored RMS profile about $X = x_0$. It should be noted that the symmetry shown in Fig. 3(d) results from how we define the location of the drift center initially. Although, in principle, different definitions would not change the result of this paper, the adopted definition provides the best convenience, as will be demonstrated.

It is now apparent that electron carriers may be in the metal electrodes when their drift centers are very near the metal-semiconductor interfaces. This introduces complexity into our subsequent analysis on charge transport, because the electric field in the metal is noticeably weaker than that in *a*-Si:H; the intrinsic property of *a*-Si:H becomes vague if the effect of the metal is included. In fact, the latter should be recognized as a parasitic factor in the form of, e.g., contact resistance. To avoid such complexity and reveal the *intrinsic property*, within *a*-Si:H a region named *R* with a length of L_{eff} is defined. The expectations are that, first, *R* needs to be kept away from the electrodes, as illustrated in Figs. 3(a) and 3(b); in this way, as long as the drift center is within *R*, the actual carrier should not reach the metals in most cases. Second, *R* should be as large as possible such that the intrinsic property of the entire *a*-Si:H is revealed, instead of a small portion of it. These requirements are analyzed in Fig. 3(d), where the RMS deviation at either side of *R* exactly reaches the metal-semiconductor interfaces (see the arrows). Consequently, the length of *R* can be analytically determined via

$$\frac{L - L_{\text{eff}}}{2} = \sqrt{\frac{3kT}{eF}} L_{\text{eff}}. \quad (6)$$

The mobility of the entire *a*-Si:H in the nanogap can now be analyzed based on the motion of drift centers across *R*. It is assumed that, after applying voltage *U*, the dc conduction lasts for a sufficiently long time, T_0 , relative to the time for the system to reach a steady current. During T_0 , N_d drift centers pass through *R*. In this setting, the average electron mobility of *a*-Si:H in the nanogap is given by the

parallel conduction analysis detailed in Appendix A as

$$\bar{\mu} = \frac{\int_{E_{F,\text{NG}}}^{+\infty} g_R(E)f(E)dE}{\int_{E_{F,\text{NG}}}^{+\infty} g(E)f(E)dE} \frac{N_d}{\sum_{i=1}^{N_d} \frac{1}{\mu_i}}. \quad (7)$$

This takes the form of a harmonic mean. *i* indexes each of the N_d drift centers, each of which exhibits unique mobility (μ_i) as it drifts across *R*. $g_R(E)$ is the DOS distribution of *a*-Si:H inside *R*. $g(E)$ is the overall DOS of *a*-Si:H in the entire nanogap. $E_{F,\text{NG}}$ is the Fermi level of *a*-Si:H inside the nanogap. $f(E)$ is the Fermi-Dirac distribution.

Multiple trapping and release (MTR) is the predominant charge-transport mechanism of room-temperature *a*-Si:H [36] (and many other amorphous semiconductors). According to Paper 2, the average dc mobility of electrons in a *macroscale* *a*-Si:H thin film is [34]

$$\mu_a = \mu_{\text{Ceff}} \frac{n_{\text{ext,eff}}}{n_{\text{ext,eff}} + n_{\text{loc,eff}}}, \quad (8)$$

where μ_{Ceff} is the effective extended-state mobility. $n_{\text{ext,eff}}$ and $n_{\text{loc,eff}}$, respectively, denote the densities of electrons that are in effective extended states and effective localized states, as demarcated by the effective CB mobility edge E_{Ceff} ; these densities are calculated based on the inner product of DOS and Fermi distribution.

It is tempting to evaluate the mobilities, μ_i , in Eq. (7) using Eq. (8). However, as analyzed in Paper 2, Eq. (8) is valid only when both of the following assumptions are true [34]. First, uniformity: extended states and localized states are approximately uniform in space such that all carriers end up interacting with generally similar electronic structures. Second, sufficiency: carriers undergo numerous trapping and release events during their migration to allow the statistical nature to be exhibited. However, these assumptions collapse in the downscaled nanogap. It is most evident that the sufficiency assumption collapses, as an electron is unlikely to undergo many trapping and release events during its ultrashort migration time across the gap. The uniformity assumption also no longer holds true. The dimension along the *x* axis is too small to leverage any macroscale uniformity. Though there is macroscopic uniformity along the *y* axis, the diffusion of electrons along this direction is limited given the analysis in Fig. 3(d). Thus, electrons that are well-separated in the *y*-axis direction may end up interacting with noticeably different electronic structures.

To solve the issue of nonuniformity, in Fig. 4(a), the nanogap is *virtually* divided along the *y* and *z* axes into totally $N_{\perp}$ “lanes” with a reasonable cross section. The expectation is that, in this way, carriers can be treated as being *confined* within the respective lanes along which their drift centers migrate; despite the nonuniformity at the

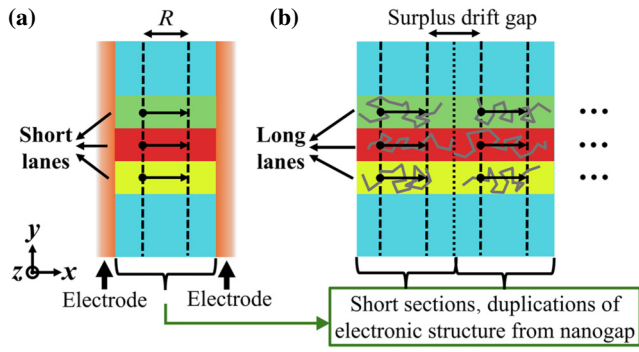


FIG. 4. Discretized transport analysis and its equivalence. (a) Division of lanes in the nanogap (z -axis division not shown). Each lane is given a specific color to demonstrate its unique local electronic structure. Motion of the drift centers of carriers are illustrated by black arrows. (b) Virtual thin film constructed through duplicating the electronic structures of a -Si:H in (a) numerous times along the x axis. Actual carrier motions are denoted by the zigzag gray lines. Repeated drifts in (a) are equivalently reproduced in (b) by discontinuous drifts along the respective long lanes. Continuous drift in (b) also includes drift processes in the surplus gaps, which are similar to the discontinuous drifts if the length of R is around half of the nanogap length.

scale of the entire nanogap, uniformity is still achieved inside the individual lanes. To meet this expectation, the lane cross section should be the *average* transverse deviation of carriers away from their drift centers. According to Eq. (5) and Fig. 3(d), this is defined based on the volume average, i.e.,

$$\frac{4}{3}\pi \langle (\delta r)^3 \rangle = \frac{\int_{x_0 - (L_{\text{eff}}/2)}^{x_0 + (L_{\text{eff}}/2)} (4/3)\pi \left(\sqrt{6kT|X - x_0|/eF} \right)^3 dX}{L_{\text{eff}}}, \quad (9)$$

through which the cross-section side length of the lanes is set as

$$d_{\text{CS}} = 2\sqrt[3]{\langle (\delta r)^3 \rangle} = \sqrt[3]{\frac{16}{5}} \sqrt{\frac{3kT}{eF}} L_{\text{eff}}. \quad (10)$$

The issue of insufficiency can now be addressed. Assume that throughout the dc conduction time, T_0 , there are n_s total drift centers that migrate across R along the s th lane. Neglecting the influence of thermal fluctuations in the material, there should not be more than one drift center existing in R at any single moment because of the limited carrier density and extremely small volume of R . As a result, basic statistics reveals that the migrations of the n_s drift centers are *independent random experiments*, while the drift time (thus, the mobilities) of these electrons are *outcomes*. As electrons are indistinguishable, it is equivalent to conduct these random experiments using a

single electron; this corresponds to a scenario where a single drift center migrates across R along the s th lane for n_s times.

This scenario is reproduced virtually. In Fig. 4(b), the *electronic structures* (not the real material) of a -Si:H in the nanogap are duplicated M times along the x axis, where, for $\forall s \in [1, N_{\perp}]$, $M > n_s$. A “virtual macroscale thin film” forms, which comprises “long lanes” duplicated from the electronic structure of the corresponding short lanes in Fig. 4(a). The scenario of “a single drift center migrating across R for n_s times” is now equivalently reproduced in Fig. 4(b) as the discontinuous drift of a single drift center across n_s short sections along the s th long lane under the same electric field F . The drift is discontinuous, in the sense that the drift center is expected to instantly “jump over” the surplus gaps of length $(L - L_{\text{eff}})$, even though the actual carrier may still interact with the electronic structures in these surplus gaps. Note that because the long lanes in the virtual macroscale thin film form via the duplication of the electronic structure of the short lanes in the nanogap, the carrier will also be confined within a single long lane in the virtual film, just as it would be in a short lane of the nanogap. Such a limited RMS deviation over a long drift distance is different from the situation in a real macroscale film, which may be imaginarily understood as a result of the existence of certain “boundaries” between adjacent long lanes, which have no other effects except elastically scattering electrons.

To resolve the discontinuous drift, continuous drift along the long lane is considered, which includes the drifts in surplus gaps. “Lane mobility” $\bar{\mu}_s$ may be defined, quantifying the most probable mobility that a drift center exhibits when it continuously drifts throughout the s th long lane. Appendix B reveals that, on condition that L_{eff} is around half of L (to be validated in Sec. IV A retrospectively), $\bar{\mu}_s$ could be used to describe the discontinuous drift as well, based on which the nanogap mobility defined in Eq. (7) transforms into a form of a weighted mean:

$$\bar{\mu} = \frac{\int_{E_{F,\text{NG}}}^{+\infty} g_R(E)f(E)dE}{\int_{E_{F,\text{NG}}}^{+\infty} g(E)f(E)dE} \frac{\sum_{s=1}^{N_{\perp}} \bar{N}_{R,s} \bar{\mu}_s}{\sum_{s=1}^{N_{\perp}} \bar{N}_{R,s}}. \quad (11)$$

Because each of the lanes in Fig. 4(b) is uniform and sufficiently long in the drift direction to allow the statistical nature to be exhibited, $\bar{\mu}_s$ can now be directly quantified via the MTR theory shown in Eq. (8). The nanogap mobility, $\bar{\mu}$, can then be calculated using Eq. (11). In Eq. (11), $\bar{N}_{R,s}$ is the average number of drift centers inside the s th lane within R , which is

$$\bar{N}_{R,s} = \frac{L_{\text{eff}} S}{N_{\perp}} \int_{E_{F,\text{NG}}}^{+\infty} g_{R,s}(E)f(E)dE, \quad (12)$$

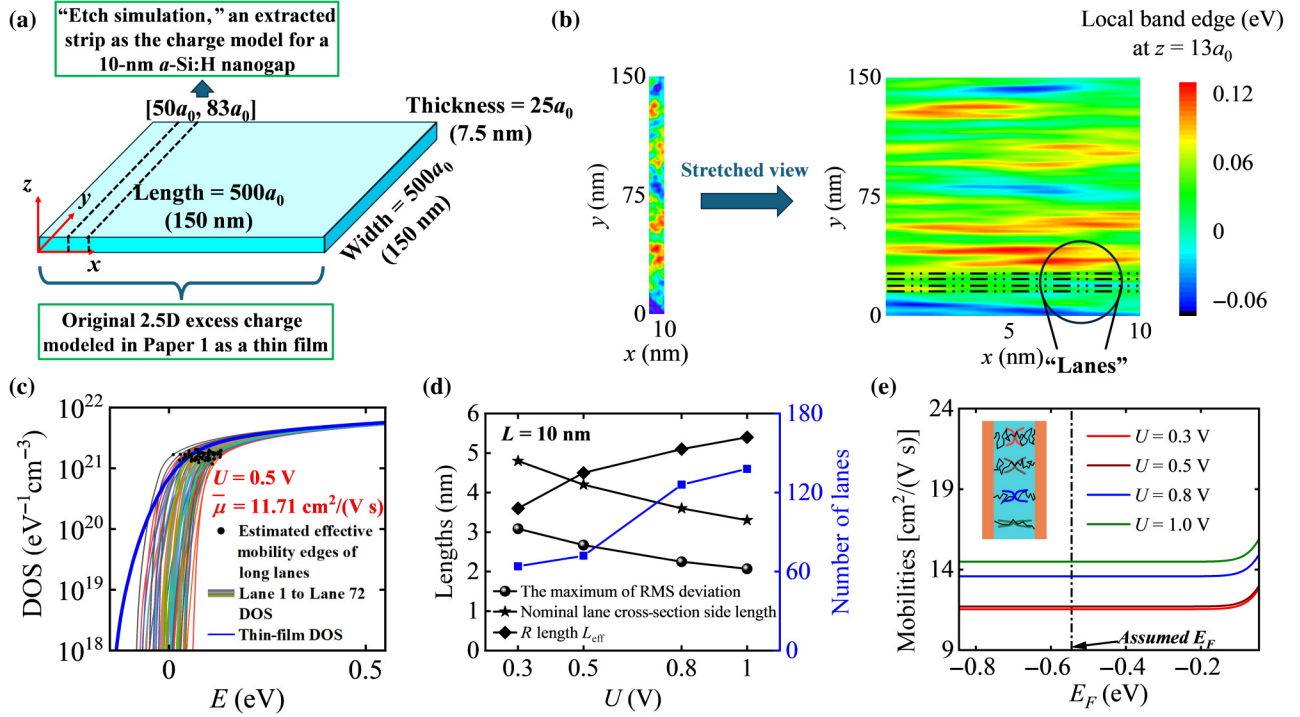


FIG. 5. Quantification of nanogap mobility. (a) Modeling *a*-Si:H nanogap based on the 2.5D excess charge model from Paper 1 [25]. (b) Band edge distribution of *a*-Si:H in the nanogap. Distribution on the middle plane ($z = 13a_0$) is shown in both the original and stretched aspect ratio. Division of lanes is illustrated in the latter. (c) Calculated DOS distributions of individual lanes under 0.5 V. These DOS curves are much steeper compared with that of the macroscale thin film. Overestimated mobility edges (black dots) are used to underestimate the 72 lane mobilities. (d) Investigation of the effect of voltage. Higher voltage confines the deviation of carriers from their drift centers, shrinks the lane cross sections (thus increasing the number of lanes), and increases the length of R ; these are illustrated in the inset of (e). (e) Dependence of nanogap mobility on voltage and Fermi position. As envisaged based on (d), a higher voltage increases nanogap mobility. It is reasonably assumed that the material does not breakdown under these voltages. Fermi-level position (within a reasonable range) has almost zero influence on nanogap mobility. Energies are relative to the conduction band edge of crystalline silicon. Arrow indicates the choice of Fermi level in this paper. Parabolic curves in the inset show the evolution of RMS deviation under different voltages.

where S is the cross-section area of the nanogap and $g_{R,s}(E)$ is the local DOS distribution specific to the s th lane within R .

IV. RESULTS AND DISCUSSION

A. Simulation of a 10-nm nanogap

As a comparative study, we base this paper on the 2.5-dimensional (2.5D) excess delocalized charge model developed in Paper 1 [25]. This model approximates a defect-free macroscale *a*-Si:H thin film, the electron mobility of which is to be compared with that after downscaling in the nanogap. Note that due to computational cost, the length scale of the 2.5D model was only 150 nm (rather than on the micrometer scale as in real situations). Hence, to produce a significant contrast, here a 10-nm nanogap is first simulated, though in practice a slightly longer gap enhances gate controllability. It should be mentioned that, in real devices, *a*-Si:H in the nanogap is fabricated through

etching from a macroscale thin film, so the extent of bond-angle and bond-length distortion of *a*-Si:H in the nanogap, which determines the excess delocalized charges therein, is the same as that of the same region before downscaling. Thus, *a*-Si:H in the nanogap is simply modeled, from the perspective of excess delocalized charges, as a “strip” that is “cut” from the larger 2.5D model [see Fig. 5(a)], assuming trivial metal-semiconductor interfacial charges.

As the first trial, we extract the strip between $x = 50a_0$ and $x = 83a_0$ from the 2.5D model. The band-edge distribution for the $z = 13a_0$ plane is mapped in Fig. 5(b), based on which $g_R(E)$ and $g(E)$ can be calculated using the method described in Paper 1 [25]. Lane division is specified according to Eq. (10), assuming that $U = 0.5$ V, based on which $g_{R,s}(E)$ is obtained. For computational simplicity, the energy integrations in Eqs. (7) and (12) start from the energy position at which the corresponding DOS is equal to 10^{18} eV $^{-1}$ cm $^{-3}$. This is reasonable because, for a nearly-defect-free *a*-Si:H, DOS below 10^{18} eV $^{-1}$ cm $^{-3}$ continues to drop exponentially and would be much

more trivial compared with the DOS at higher energies; for example, the latter is typically $>10^{21} \text{ eV}^{-1} \text{ cm}^{-3}$ at extended-state energies [25]. Note that although the electric field at $U=0.5 \text{ V}$ (i.e., 0.5 MV/cm) is high enough to cause the onset of velocity saturation of extended-state electrons in α -Si:H due to acoustic phonon emission, the effect of this saturation is not strong, as shown in Ref. [37]; meanwhile, a high field is able to promote thermal release [37,38], which compensates for the velocity saturation effect. Therefore, if combining these two counteracting factors into the single quantity, μ_{Ceff} , in Eq. (8), this quantity may be treated as being field independent. Furthermore, μ_{Ceff} , which is a quantity related to carrier scattering, may also be treated as being dimension independent here; in other words, it does not change after downscaling. This is because the mean free path of extended-state electrons in amorphous semiconductors is of the interatomic spacing scale [11], which is still much shorter than the gap length, despite downscaling; ballistic transport will not occur. As a consequence, μ_{Ceff} will be canceled if it is the relative mobility increase that is focused upon. Still, we continue to use the value of $46 \text{ cm}^2/(\text{V s})$, as revealed in Paper 2 [34], to straightforwardly show the absolute mobilities.

Figure 5(c) shows $g_{R,s}(E)$, the local DOSs of individual lanes, which are compared with the DOS of their macroscale thin-film counterpart before downscaling. It is evident that downscaling fundamentally changes the steepness of these constituent DOSs, which validates our initial hypothesis that high-magnitude low-frequency band-fluctuation components are filtered through downscaling. To evaluate room-temperature lane mobilities, $\bar{\mu}_s$, using Eq. (8), the effective mobility edge (E_{Ceff}) of the lanes needs to be known in addition to μ_{Ceff} . A worst-case scenario is considered; E_{Ceff} is overestimated to be at the highest local band edge in the individual lanes. This is shown by the black dots on the DOS curves in Fig. 5(c). Then, the nanogap intrinsic mobility, $\bar{\mu}$, is calculated to be $11.71 \text{ cm}^2/(\text{V s})$ through Eq. (11), i.e., about 12 times that before downscaling.

Voltage plays a role in that it determines the magnitude of electric field in the nanogap and thereby determines L_{eff} through Eq. (6) and the cross-section length of divided lanes through Eq. (10). Additional voltages of 0.3, 0.8, and 1 V are simulated. Figure 5(d) shows that, as the voltage increases, the lateral diffusion of carriers is more restricted, leading to a smaller lane cross section, more lanes, and a longer R length, L_{eff} . These are schematically illustrated in the inset of Fig. 5(e). Note that the earlier assumption that L_{eff} is around half of L is now deemed acceptable.

There may be a concern that the voltages studied herein produce high electric fields up to the order of 1 MV/cm , which may lead to the breakdown of α -Si:H. Nevertheless, there is evidence that avalanche multiplication in α -Si:H via impact ionization does not occur, unless the field is even higher than 1 MV/cm [39]; meanwhile, since α -Si:H

in this paper is assumed to be of high quality and free from defects and contaminations, breakdown via defect states is not possible, which also ensures a high field tolerance [40]. High fields are involved in this paper for two reasons. First, according to Eq. (10), high fields are necessary to yield lane cross sections that are small relative to the entire cross section of the nanogap, such that the number of divided lanes is sufficient to yield statistically accurate mobility results. Second, for real nanogaps and nanoscale TFTs, voltages on the orders of 0.1 and 1 V should still be applied, as is common in the literature [41–44]. These voltages are compatible with ordinary driving circuitries and ensure a sufficient signal-to-noise ratio. Fields on or slightly above the order of 0.1 MV/cm are not unexpected.

In addition to showing that the nanogap mobility increases with an increasing voltage U , Fig. 5(e) also shows the influence of the Fermi-level position. The Fermi level of α -Si:H in the nanogap is not precisely known at present, but within a reasonably large range of values around the Fermi level of typical macroscale α -Si:H [25,36] (indicated by the arrow in the figure), the mobility result is not changed.

B. Simulation of device-to-device (D2D) variation

“Cutting” from different parts of the 2.5D model is analogous to batch-fabricating nanogaps in an industrial setting. Although all these nanogaps share the same geometry and are fabricated under identical conditions, the stochasticity of the detailed material structure would show up and lead to property variation from one nanogap to another. This is also known as “D2D variation.” The D2D variation of intrinsic mobility is simulated here through extracting charges at multiple randomly chosen locations, for instance, at $x \in [100a_0, 133a_0]$, $[200a_0, 233a_0]$, $[290a_0, 323a_0]$, $[350a_0, 383a_0]$, and $[420a_0, 453a_0]$. This is illustrated in Fig. 6(a). Together with the first trial in Sec. IV A and under the same $U=0.5 \text{ V}$, six nanogap mobilities are shown in Fig. 6(b); the mobility of the macroscale thin-film counterpart (μ_{TF}) and μ_{Ceff} are indicated for comparison. On average, these nanogaps exhibit an average mobility of $11.86 \text{ cm}^2/(\text{V s})$, with the highest mobility being about $14\mu_{\text{TF}}$ and the lowest mobility being $9\mu_{\text{TF}}$.

Coefficient of variation (CV), which quantifies the D2D mobility variation, is 16.26%. To quantitatively make sense of this CV value, the variation of the 72 lane mobilities ($\bar{\mu}_s$) within each single trial is investigated for comparison. These lane mobility values are indicated by the scatters in Fig. 6(b); their CV values are recorded in the top table. Compared with these, the CV of nanogaps is significantly smaller.

The width of nanogaps plays an important role in D2D variation; to better demonstrate this, nanogaps with five more different widths are simulated. The way that these nanogaps are produced is illustrated in Fig. 6(c). Figure

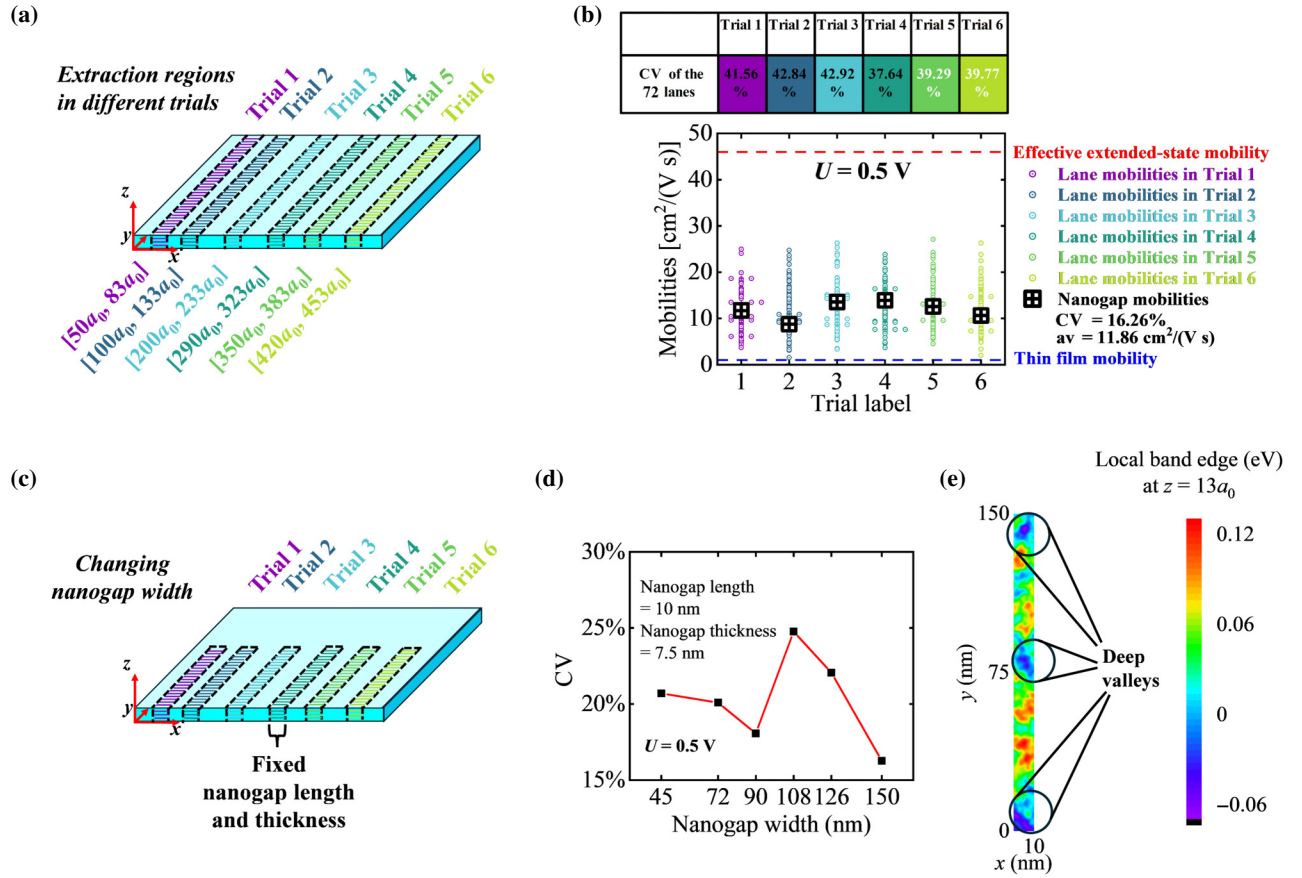


FIG. 6. Study of D2D variation at the material level. (a) Illustration of the production of six random trials. (b) Mobility variation. Average mobility of the six trials is 11.97 times the mobility (μ_{TF}) of their macroscale thin-film counterpart (dashed blue line). CV, which reflects the D2D variation, is 16.26%. Individual lane mobilities in each trial are denoted by hollow circles. Top table lists the CV of the lane mobilities for each trial; 16.26% is significantly lower in comparison. (c) Illustration of the production of nanogaps with different widths. (d) CV calculated at different widths. (e) Examples of deep energy valleys in the band-fluctuation landscape.

6(d) shows that the CV of the nanogap mobility is indeed the lowest at the largest width, which may be understood from the perspective of the following mechanism.

Figure 6(e) shows the band fluctuation of one example nanogap [which is studied in Fig. 5(b)], where deep energy valleys are identified. If a lane of the nanogap contains one such electronic structure, its lane mobility would be considerably lower, because a valley produces significant localized states, which retard the drift of carriers. In Eq. (11), the weight of lane mobility is proportional to $\bar{N}_{R,s}$. For the lane that contains a valley, its local DOS distribution would be closer to the Fermi level than those of the others [see Fig. 5(c)]; this indicates a higher occupation probability, f , and a consequentially higher $\bar{N}_{R,s}$. The above facts mean that a low lane mobility takes a relatively large weight in the calculation of Eq. (11), both of which are the consequences of the energy valley. As a result, batch- and identically fabricated nanogaps would behave differently if some did not contain a deep valley while others contained one or more.

Such differences, however, would be lowered if the batch-fabricated nanogaps were sufficiently wide. A wider nanogap contains more lanes, such that, through the weighted-mean operation in Eq. (11), the property fluctuation caused by a single (or a few) lane(s) would be suppressed. It is expected that for real nanogaps with much larger widths (e.g., in the micrometer scale) than the simulation in this paper, the D2D variation of intrinsic mobility should be considerably lower.

Furthermore, viewing device-level properties, a sufficient nanogap width is also beneficial because it suppresses the variation of metal-semiconductor contact resistance, thanks to the long-range uniformity along the y -axis direction. This topic could be quantitatively investigated in the future as the next step towards device-level simulations.

C. Dependence of mobility on the nanogap length

Five more length scales, 20, 30, 40, 50, and 60 nm, are studied. To obtain statistically reliable results without being affected by stochastic variations, six nanogaps

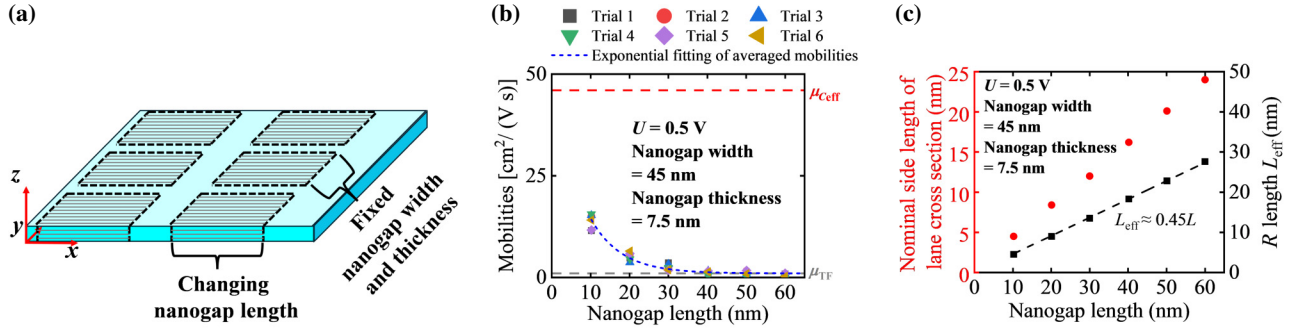


FIG. 7. Dependence of nanogap mobility on nanogap length. (a) Illustration of the production of nanogaps with different lengths. Due to the limit of simulation size, each nanogap is only 45 nm wide, such that six trials can be produced from a single thin-film model. (b) Simulation results and fitting. For each of the six nanogap lengths, six trials are conducted. In total, 36 mobility results are obtained. Average mobility of the six trials at each of the six lengths is calculated and fitted using an exponential decay function. Mobility of their thin-film counterpart (μ_{TF}) and the effective extended-state mobility (μ_{eff}) are shown for comparisons. (c) Dependence of the length of R (right axis) and the side length of the lane cross section (left axis) on the length of nanogap. Both scale linearly with the nanogap length under a fixed voltage of $U = 0.5$ V.

are simulated for each of these lengths. Because the 2.5D model is limited in size due to computational load, also in light of the space occupations of the extraction regions in the 2.5D model, the width of the six nanogaps at each length decreases to only 45 nm in this subsection; they are extracted according to the illustration in Fig. 7(a). Together with the results of the 10-nm nanogaps, a total of 6×6 nanogap mobilities under the same voltage, $U = 0.5$ V, are calculated in Fig. 7(b). The average mobility of the six nanogaps at each length is then calculated, fitting well with an exponential function of the form $\mu_{const} \exp(-L/L_0) + \mu_{TF}$, where L_0 is a length constant and μ_{const} is a mobility constant. This is shown by the dashed blue curve in the figure. Note that extrapolation below 10 nm is not shown because these lengths are too short to be practically meaningful. The above results further validate our hypothesis that downscaling acts as a spatial frequency filter. The greater the downscaling, the fewer the low-frequency components, the lower the amplitudes of band fluctuation, and the higher the mobility.

It is worthwhile noting that a size of 20 nm (length) $\times$ 45 nm (width) $\times$ 7.5 nm (thickness) is simulated in this subsection, showing an average mobility of $5.16 \text{ cm}^2/(\text{V s})$. In comparison, a size of 10 nm (length) $\times$ 90 nm (width) $\times$ 7.5 nm (thickness) was simulated in the last subsection [Fig. 6(d)]; this shares the same volume as the former case, but it exhibits a much higher mobility of $12.27 \text{ cm}^2/(\text{V s})$. The fundamental reason why the same volume corresponds to different mobilities lies in the *directionality* defined by the carrier drift direction. To increase mobility, it is important to note which dimension is to be downscaled; this must be in the direction along which carriers are drifting (i.e., opposite to the direction of the applied electric field).

Shown in Fig. 7(c) is a demonstration that, under the same U , the length of R scales linearly as the nanogap length increases and the nominal lane cross-section side length also scales linearly. Here, the term “nominal” is used because the discretized division could result in a few “incomplete” lanes at the model margins, the cross sections of which are smaller, but these are minorities under most circumstances.

D. The role of medium-range order (MRO)

According to Paper 1, MRO of a -Si:H may be quantified by a modeling parameter named *window dimension* (w) [25]. Viewing this parameter as a variable that is related to the conditions and quality of material deposition, the dependence of nanogap mobility on w is investigated in this subsection. Here, in addition to $w = 5a_0$, which was discussed in previous subsections, another four choices of w (i.e., $w = 0, 2a_0, 10a_0$, and $20a_0$) are studied (for details of the modeling method, see Appendix C). Similar to Fig. 1(c), the Fourier decompositions of the band edge fluctuation along the dashed line in Fig. 1(a) (but under different choices of w) are calculated in Figs. 8(a) and 8(b). It can be observed that, with a larger w , high-frequency components are more trivial. It can thus be envisaged that, after downscaling, these remnant high-frequency components will possess lower amplitudes; this will lead to steeper lane DOSs and, consequently, higher nanogap mobility.

The above envisagement is confirmed by the results in Fig. 8(c), which show that, at any fixed nanogap length L , the mobility increases nearly linearly with the increase of w . It can thus be concluded that, fundamentally, it is the *relative length scale* of the nanogap compared with the MRO length that determines the mobility improvement after downscaling.

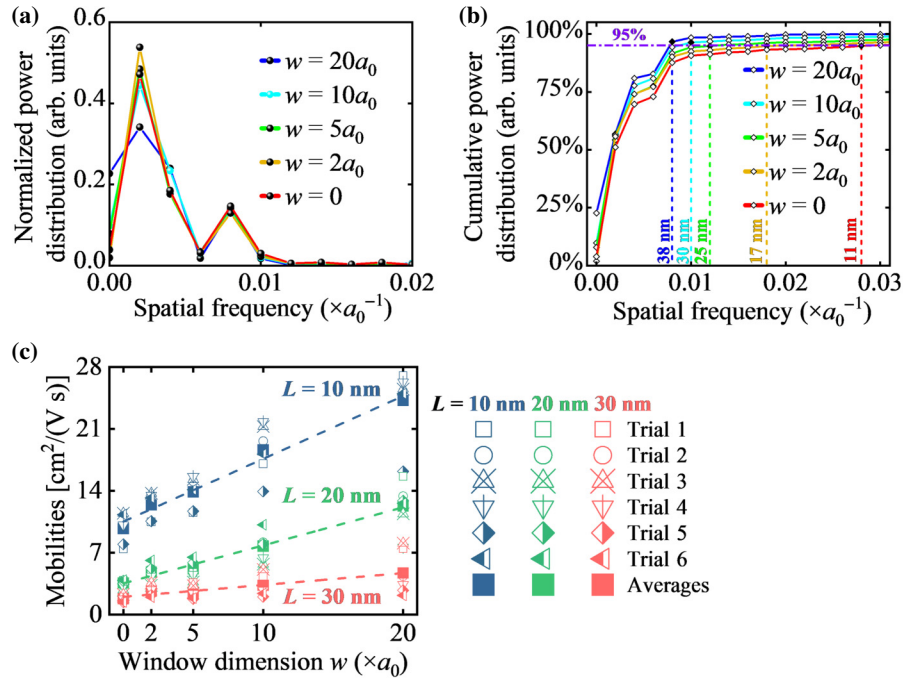


FIG. 8. Influence of MRO on the spatial frequencies of band fluctuation. (a) Normalized distribution of Fourier components extracted from the band edge fluctuation along the dashed line in Fig. 1(a) but under different choices of window dimension (w). (b) Cumulative distribution of spatial frequencies in (a). Like Fig. 1(c), the spatial frequency at which the cumulation reaches 95% is defined as the cutoff frequency. As w increases, the cutoff frequency decreases, and the length scale that corresponds to the cutoff frequency increases, as labeled in the figure. (c) Dependence of nanogap mobility on the nanogap length and w . Three nanogap lengths and five values of w are studied. For each case, six trials are conducted, as in Fig. 7. For a fixed nanogap length, the average nanogap mobility of the respective six trials increases nearly linearly with increasing w . Mobility is ultimately a combined result of nanogap length and the length scale of medium-range order, which is associated with w .

E. Discussion

From the above analyses and results, it is now clear that downscaling improves the intrinsic carrier mobility of an amorphous semiconductor for two reasons. First, the nanoscale length leads to an ultrashort migration time of carriers across the gap; unlike in a macroscale film where the lateral diffusion of carriers can be noticeable, the carriers in the nanogap predominantly move along the downscaled length direction. Second, along this length direction, the band fluctuation of the amorphous semiconductor is smoother after downscaling, resulting in a lower density of localized band-tail states and consequentially enabling a higher carrier mobility. This mitigated fluctuation stems from the size shrinking, which is analogous to a high-pass filter that selectively blocks low-frequency fluctuation components that otherwise contribute to most of the band fluctuation. Ultimately, the extent of mobility increase depends on the detailed spatial frequency compositions of band fluctuation before downscaling and how their corresponding length scales compare to the downscaled material length; this is where the medium-range order of the amorphous semiconductor plays a role in part.

One of the superiorities of amorphous semiconductors to other noncrystalline and crystalline semiconductors was their excellent large-area uniformity, but little was known when it came to the nanoscale. It is shown in this paper that, to retain uniformity even after downscaling, it is essential that at least one dimension is unscaled.

V. CONCLUSIONS

This paper is a significant theoretical endeavor that elucidates the improvement of intrinsic carrier mobility in downscaled amorphous semiconductors, serving as a reliable guide for experiments in the future. To start with, the concept of spatial frequency filtering is introduced, which proposes that the band fluctuation (which determines carrier mobility) can be decomposed into various Fourier components and that downscaling acts as a “high-pass filter,” which leaves only high-frequency components with limited amplitudes that indicate a higher mobility. This preliminary hypothesis is validated in the context of a nanogap structure, using the model of *a*-Si:H established in Paper I [25]. It is shown that the intrinsic electron mobility of *a*-Si:H in a 10-nm nanogap can be about 12 times that of a standard macroscale *a*-Si:H thin film. Device-to-device

variation is studied, which reveals the importance of maintaining at least one of the other two dimensions unscaled. Furthermore, it is analyzed that the mobility improvement depends on both the extent of downscaling and the length scale of medium-range order in the amorphous semiconductor. Combined with future device-level optimizations that resolve nonideality factors, the improvement of intrinsic carrier mobility will eventually show up in the proposed nanoscale TFTs.

The results in this paper are significant in light of the need to improve the performance of electronic devices based on amorphous semiconductors. For instance, high-density high-speed flexible microprocessors may be developed based on the concept proposed in this paper, which may find extensive applications in next-generation wearable electronics with edge computing functionalities. Hardware development like this is increasingly important in supporting the soaring development of the IOT.

ACKNOWLEDGMENTS

The authors are grateful to Dr. Gwenhvir Wyatt-Moon and Dr. Kham Niang for their insights from experimental perspectives. This work is supported by the UKRI Engineering and Physical Sciences Research Council (Grant No. EP/W009757/1), the Rank Prize Return to Research Grant, and a Ph.D. scholarship from the Cambridge Commonwealth, European and International Trust.

DATA AVAILABILITY

The data that support the findings of this article are openly available [45].

APPENDIX A: ANALYSIS OF NANOGAP MOBILITY VIA PARALLEL CONDUCTION

From the definition, the average conductivity ($\bar{\sigma}$) of a -Si:H in the nanogap is

$$\bar{\sigma} = \frac{I_{SS}L}{US}, \quad (A1)$$

where I_{SS} denotes the current at a steady state. As defined earlier, L is the length of the nanogap, U is the applied voltage, and S is the cross-section area of the nanogap. It should be noted that the measured I_{SS} is deemed to be the *net current* that is contributed to by only the drift motion of electrons and is exclusive of noise caused by random thermal motion; it can thus be calculated as

$$I_{SS} = \frac{eN_d}{T_0}, \quad (A2)$$

assuming that, as mentioned in Sec. III, there are a total of N_d drift centers passing through R within a sufficient time T_0 .

Excluding thermal noise and statistical fluctuation, the *average number* of drift centers inside R , $\bar{N}_R$, is

$$\begin{aligned} \bar{N}_R &= L_{\text{eff}} S \bar{n}_R, \\ \bar{n}_R &= \int_{E_{F,\text{NG}}}^{+\infty} g_R(E) f(E) dE, \end{aligned} \quad (A3)$$

where, as defined in Sec. III, $g_R(E)$ is the DOS distribution of a -Si:H inside R , $E_{F,\text{NG}}$ is the Fermi level of a -Si:H in the nanogap, and $f(E)$ is the Fermi-Dirac distribution. It is noted that $\bar{N}_R$ is far less than unity, considering the nanoscale L_{eff} and limited DOS. Therefore, the actual number of drift centers within R temporally varies between 0 and 1. Suppose that the drift center of the i th of the N_d electrons drifts across R for a duration t_i , where

$$t_i = \frac{L_{\text{eff}}}{\mu_i F} = \frac{L_{\text{eff}} L}{\mu_i U}, \quad (A4)$$

with $i \in [1, N_d]$ and μ_i being the unique mobility of the drift center as it crosses R . Note that although the drift center only exists in R in our analysis, the corresponding actual electron carrier is able to span wider over the entire nanogap. Consequently, the average electric field experienced by the actual electron is still the total field, F , across the entire nanogap.

An interpretation of the statistical average, $\bar{N}_R$, from a temporal perspective shows that the cumulation of the drift time of all these N_d electrons is $\bar{N}_R T_0$. Thus,

$$\sum_{i=1}^{N_d} t_i = \bar{N}_R T_0. \quad (A5)$$

Substituting Eqs. (A2) and (A4) into Eq. (A5) yields

$$\sum_{i=1}^{N_d} \frac{L_{\text{eff}} L}{\mu_i U} = \bar{N}_R \frac{e N_d}{I_{SS}}. \quad (A6)$$

The average electron mobility of a -Si:H in the entire nanogap is defined by Eq. (A1) as

$$\bar{\mu} = \frac{I_{SS} L}{US} / e \bar{n}, \quad (A7)$$

where $\bar{n}$ represents the average density of electron carriers in the entire nanogap, which is

$$\bar{n} = \int_{E_{F,\text{NG}}}^{+\infty} g(E) f(E) dE, \quad (A8)$$

with $g(E)$ being the DOS of the entire nanogap [which may be slightly different from $g_R(E)$]. Integrating Eqs. (A3) and (A7) into Eq. (A6) leads to

$$\bar{\mu} = \frac{\bar{n}_R}{\bar{n}} \frac{N_d}{\sum_{i=1}^{N_d} 1/\mu_i}. \quad (A9)$$

APPENDIX B: EQUIVALENT TRANSPORT ANALYSIS BASED ON Fig. 4

In Fig. 4(a), assume that, throughout the sufficiently long time T_0 , there are a total of n_s drift centers that migrate across R along the s th lane, the j th of which exists in R for a duration $t_{s,j}$ and exhibits a mobility of $\mu_{s,j}$ with $j \in [1, n_s]$. Note that n_s should be sufficiently large in time T_0 , so the statistical nature,

$$\sum_{j=1}^{n_s} t_{s,j} = \bar{N}_{R,s} T_0, \quad (\text{B1})$$

can be exhibited, where, as introduced in Sec. III, $\bar{N}_{R,s}$ is the average number of drift centers inside the s th lane within R , which is quantified by Eq. (12).

As analyzed in Sec. III, the migration of the n_s drift centers across R is statistically equivalent to the discontinuous drift of a single drift center along the s th long lane in Fig. 4(b). This drift center requires time $t_{s,k}$ to cross the k th section (of length L_{eff}) in the s th long lane and exhibits a mobility $\mu_{s,k}$, with $k \in [1, n_s]$. Given that n_s is sufficiently large, the statistical equivalence can be exhibited, which yields

$$\sum_{k=1}^{n_s} t_{s,k} = \sum_{j=1}^{n_s} t_{s,j}, \quad (\text{B2})$$

with both equal to $\bar{N}_{R,s} T_0$, considering Eq. (B1). Furthermore, given the same total electric field F in both Figs. 4(a) and 4(b), U/L_{eff} is multiplied to both sides of Eq. (B2), and so

$$\sum_{k=1}^{n_s} \frac{1}{\mu_{s,k}} = \sum_{j=1}^{n_s} \frac{1}{\mu_{s,j}}. \quad (\text{B3})$$

Note that the drift along the s th long lane has so far been considered to be discontinuous, as illustrated in Fig. 4(b). Now, instead, consider an imagined scenario where the drift center does continuously drift along the s th long lane in Fig. 4(b). Such an imagined continuous drift process includes both the discontinuous drift and the drift across the surplus gaps, so invoking the lane mobility, $\bar{\mu}_s$, defined in Sec. III, we have

$$\sum_{k=1}^{n_s} \frac{L_{\text{eff}}}{(U/L)\mu_{s,k}} + \sum_{k'=1}^{n_s} \frac{L - L_{\text{eff}}}{F'\mu_{s,k'}} = \frac{n_s L}{(U/L)\bar{\mu}_s}, \quad (\text{B4})$$

where k' indexes the drift across each of the surplus gaps, during which the drift center exhibits a unique mobility, $\mu_{s,k'}$. F' denotes the electric field that the carrier experiences while the drift center is in these surplus gaps. As shown in Fig. 5(d), L_{eff} is around half of L , and so it is

reasonably assumed that the migrations of the drift center within the surplus gaps are similar to the discontinuous drift in Fig. 4(b); this leads to the approximations that

$$F' = \frac{U}{L}, \quad \sum_{k=1}^{n_s} \frac{1}{\mu_{s,k}} = \sum_{k'=1}^{n_s} \frac{1}{\mu_{s,k'}}. \quad (\text{B5})$$

These simplify Eq. (B4) and lead to

$$\sum_{k=1}^{n_s} \frac{1}{\mu_{s,k}} = \frac{n_s}{\bar{\mu}_s}, \quad (\text{B6})$$

which can be linked to Eq. (B3). Then Eq. (A9) is equivalently:

$$\bar{\mu} = \frac{\bar{n}_R}{\bar{n}} \frac{\sum_{s=1}^{N_{\perp}} n_s}{\sum_{s=1}^{N_{\perp}} n_s / \bar{\mu}_s}. \quad (\text{B7})$$

It should be noted that, for each part of the discontinuous drift,

$$t_{s,k} = \frac{L_{\text{eff}}}{\mu_{s,k}(U/L)}, \quad (\text{B8})$$

and so, summing up k on both sides yields

$$\sum_{k=1}^{n_s} t_{s,k} = \frac{L_{\text{eff}} L}{U} \sum_{k=1}^{n_s} \frac{1}{\mu_{s,k}}. \quad (\text{B9})$$

Combining this with Eq. (B6),

$$n_s = \bar{\mu}_s \frac{U}{L_{\text{eff}} L} \sum_{k=1}^{n_s} t_{s,k}. \quad (\text{B10})$$

Thus, Eqs. (B1), (B2), and (B10) simplify Eq. (B7) to

$$\bar{\mu} = \frac{\bar{n}_R}{\bar{n}} \frac{\sum_{s=1}^{N_{\perp}} \bar{N}_{R,s} \bar{\mu}_s}{\sum_{s=1}^{N_{\perp}} \bar{N}_{R,s}}. \quad (\text{B11})$$

APPENDIX C: SIMULATION PROCEDURES

Figures 5–7 are based on the 2.5D excess delocalized charge model developed in Paper 1 [25]. The mobilities in these figures are obtained in four steps.

First, produce the excess delocalized charge model for individual nanogaps. This is straightforward; simply extract the charge values from the original 2.5D model according to the specifications of nanogap width and length, and repeat this at different positions in the 2.5D model to mimic the stochasticity and device-to-device variation in real fabrication processes.

Second, follow the perturbation strategy developed in Paper 1 and produce the band fluctuation of α -Si:H in

individual nanogaps. This relies on the three-dimensional fast multipole method algorithm (FMM3D, developed by the Flatiron Institute [46] and adopted in Paper 1), to accelerate the calculation.

Third, produce the DOS distribution. The DOS distributions of a -Si:H in the entire nanogap and in R are easily obtained based on the band-fluctuation result (using the same perturbation method as that in Paper 1). Afterwards, the division of lanes is specified according to Eqs. (6) and (10). The DOS distribution of a -Si:H in each lane is then similarly calculated.

Fourth, calculate the nanogap mobility. This relies on the virtual film construction in Fig. 4(b), based on which the nanogap mobility is derived using Eq. (11). The integrations are approximated by summations using an energy interval of 0.001 eV.

Figure 8 requires extra 2.5D charge models to simulate different extents of medium-range order. Generating these charge models entirely follows the procedures detailed in Paper 1. (i) Start with the generation of a Gaussian array of random charge values (with standard deviation σ). (ii) Smooth the array using the moving average algorithm; different window dimensions (w) represent different medium-range orders. (iii) Calculate the band fluctuation using the FMM3D algorithm. (iv) Produce the DOS distribution of the respective 2.5D models. (v) Retrospectively adjust σ to alter the respective DOS curves until they match the DOS curve of the $w = 5a_0$ model. This ensures the same macroscale mobility and has the purpose of exclusively revealing the influence of MRO.

The decomposition in Figs. 1 and 8 is completed using the built-in Fourier decomposition analysis of OriginPro. Focus is given to the normalized power distribution, based on which the cumulative distribution is obtained. To define what spatial frequencies are “predominant,” a criterium of a 95% cumulative distribution is used.

Except the FMM3D algorithm, which is conducted in FORTRAN, all the other computations are done in MATLAB. Fittings are completed in OriginPro.

-
- [1] D. Geng, K. Wang, L. Li, K. Myny, A. Nathan, J. Jang, Y. Kuo, and M. Liu, Thin-film transistors for large-area electronics, *Nat. Electron.* **6**, 963 (2023).
 - [2] R. A. Street, Thin-film transistors, *Adv. Mater.* **21**, 2007 (2009).
 - [3] K. Myny, The development of flexible integrated circuits based on thin-film transistors, *Nat. Electron.* **1**, 30 (2018).
 - [4] A. Snell, K. Mackenzie, W. Spear, P. LeComber, and A. Hughes, Application of amorphous silicon field effect transistors in addressable liquid crystal display panels, *Appl. Phys. A* **24**, 357 (1981).
 - [5] K. Nomura, H. Ohta, A. Takagi, T. Kamiya, M. Hirano, and H. Hosono, Room-temperature fabrication of transparent

flexible thin-film transistors using amorphous oxide semiconductors, *Nature* **432**, 488 (2004).

- [6] H. Oh, J.-Y. Oh, C. W. Park, J.-E. Pi, J.-H. Yang, and C.-S. Hwang, High density integration of stretchable inorganic thin film transistors with excellent performance and reliability, *Nat. Commun.* **13**, 4963 (2022).
- [7] S.-H. Kang, J.-W. Jo, J. M. Lee, S. Moon, S. B. Shin, S. B. Choi, D. Byeon, J. Kim, M.-G. Kim, Y.-H. Kim, J.-W. Kim, and S. K. Park, Full integration of highly stretchable inorganic transistors and circuits within molecular-tailored elastic substrates on a large scale, *Nat. Commun.* **15**, 2814 (2024).
- [8] J. Biggs, J. Myers, J. Kufel, E. Ozer, S. Craske, A. Sou, C. Ramsdale, K. Williamson, R. Price, and S. White, A natively flexible 32-bit arm microprocessor, *Nature* **595**, 532 (2021).
- [9] E. Ozer, J. Kufel, J. Myers, J. Biggs, G. Brown, A. Rana, A. Sou, C. Ramsdale, and S. White, A hardwired machine learning processing engine fabricated with sub-micron metal-oxide thin-film transistors on a flexible substrate, *Nat. Electron.* **3**, 419 (2020).
- [10] W. Wang, K. Li, J. Lan, M. Shen, Z. Wang, X. Feng, H. Yu, K. Chen, J. Li, F. Zhou, L. Lin, P. Zhang, and Y. Li, CMOS backend-of-line compatible memory array and logic circuitries enabled by high performance atomic layer deposited ZnO thin-film transistor, *Nat. Commun.* **14**, 6079 (2023).
- [11] R. A. Street, *Hydrogenated Amorphous Silicon* (Cambridge University Press, New York, 1991).
- [12] S. Yamamoto, J. N. J. Nakamura, and M. M. M. Migitaka, High-mobility thin-film transistor fabricated using hydrogenated amorphous silicon deposited by discharge of disilane, *Jpn. J. Appl. Phys.* **35**, 3863 (1996).
- [13] M. Brinza and G. Adriaenssens, Electronic properties of hydrogenated amorphous silicon prepared in expanding thermal plasmas, *J. Optoelectron. Adv. Mater.* **7**, 73 (2005).
- [14] M. Sakai, T. Tsutsumi, T. Yoshioka, A. Masuda, and H. Matsumura, High performance amorphous-silicon thin film transistors prepared by catalytic chemical vapor deposition with high deposition rate, *Thin Solid Films* **395**, 330 (2001).
- [15] H. Hosono, Ionic amorphous oxide semiconductors: Material design, carrier transport, and device application, *J. Non-Cryst. Solids* **352**, 851 (2006).
- [16] K. Nomura, A. Takagi, T. Kamiya, H. Ohta, M. Hirano, and H. Hosono, Amorphous oxide semiconductors for high-performance flexible thin-film transistors, *Jpn. J. Appl. Phys.* **45**, 4303 (2006).
- [17] H.-W. Zan, C.-C. Yeh, H.-F. Meng, C.-C. Tsai, and L.-H. Chen, Achieving high field-effect mobility in amorphous indium-gallium-zinc oxide by capping a strong reduction layer, *Adv. Mater.* **24**, 3509 (2012).
- [18] J.-L. Lin, W.-J. Sah, and S.-C. Lee, Amorphous-silicon thin-film transistors with very high field-effect mobility, *IEEE Electron Device Lett.* **12**, 120 (1991).
- [19] B. C. Lim, Y. J. Choi, J. H. Choi, and J. Jang, Hydrogenated amorphous silicon thin film transistor fabricated on plasma treated silicon nitride, *IEEE Trans. Electron Devices* **47**, 367 (2000).

- [20] H. Okada, Y. Uchida, and M. Matsumura, High-mobility amorphous-silicon MOS transistors, *Jpn. J. Appl. Phys.* **25**, L718 (1986).
- [21] L. Han, P. Mandlik, K. H. Cherenack, and S. Wagner, Amorphous silicon thin-film transistors with field-effect mobilities of $2 \text{ cm}^2/\text{Vs}$ for electrons and $0.1 \text{ cm}^2/\text{Vs}$ for holes, *Appl. Phys. Lett.* **94**, 162105 (2009).
- [22] E. Fortunato, L. Pereira, P. Barquinha, A. M. Botelho do Rego, G. Gonçalves, A. Vilà, J. R. Morante, and R. F. Martins, High mobility indium free amorphous oxide thin film transistors, *Appl. Phys. Lett.* **92**, 222103 (2008).
- [23] H. Chiang, J. Wager, R. Hoffman, J. Jeong, and D. A. Keszler, High mobility transparent thin-film transistors with amorphous zinc tin oxide channel layer, *Appl. Phys. Lett.* **86**, 013503 (2005).
- [24] M. Ando, M. Wakagi, and T. Minemura, Effects of back-channel etching on the performance of a-Si:H thin-film transistors, *Jpn. J. Appl. Phys.* **37**, 3904 (1998).
- [25] Y. Luo and A. Flewitt, Understanding localized states in the band tails of amorphous semiconductors exemplified by *a*-Si:H from the perspective of excess delocalized charges, *Phys. Rev. B* **109**, 104203 (2024).
- [26] M. Kim, J. H. Jeong, H. J. Lee, T. K. Ahn, H. S. Shin, J.-S. Park, J. K. Jeong, Y.-G. Mo, and H. D. Kim, High mobility bottom gate InGaZnO thin film transistors with SiO_x etch stopper, *Appl. Phys. Lett.* **90**, 212114 (2007).
- [27] I. Song, S. Kim, H. Yin, C. J. Kim, J. Park, S. Kim, H. S. Choi, E. Lee, and Y. Park, Short channel characteristics of gallium-indium-zinc-oxide thin film transistors for three-dimensional stacking memory, *IEEE Electron Device Lett.* **29**, 549 (2008).
- [28] J. Guo, *et al.*, A New Surface Potential and Physics Based Compact Model for a-IGZO TFTs at Multinanoscale for High Retention and Low-Power DRAM Application, in *2021 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2021), pp. 8.5.1.
- [29] D. J. Beesley, J. Semple, L. Krishnan Jagadamma, A. Amassian, M. A. McLachlan, T. D. Anthopoulos, and J. C. Demello, Sub-15 nm patterning of asymmetric metal electrodes and devices by adhesion lithography, *Nat. Commun.* **5**, 3933 (2014).
- [30] J. Semple, Dimitra G. Georgiadou, Gwenhivir Wyatt-Moon, Minh Yoon, Akmaral Seitkhan, Emre Yengel, Stephan Rossbauer, Francesca Bottacchi, Martyn A. McLachlan, Donal D. C. Bradley, and Thomas D. Anthopoulos, Large-area plastic nanogap electronics enabled by adhesion lithography, *npj Flexible Electron.* **2**, 1 (2018).
- [31] G. Wyatt-Moon, K. Niang, C. Rider, and A. Flewitt, Air stable indium-gallium-zinc-oxide diodes with a 6.4 GHz extrinsic cutoff frequency fabricated using adhesion lithography, *IEEE Electron Device Lett.* **41**, 175 (2019).
- [32] G. Wyatt-Moon, G. Saravanel, S. Sambandan, and A. Flewitt, Nanodiodes on a digestible substrate, *IEEE Electron Device Lett.* **44**, 337 (2022).
- [33] M. Zeman, Advanced amorphous silicon solar cell technologies, in *Thin Film Solar Cells: Fabrication, Characterization and Applications*, edited by J. Poortmans and V. Arkhipov (John Wiley & Sons, Ltd., West Sussex, England, 2006), pp. 173.
- [34] Y. Luo and A. Flewitt, Revisiting multiple trapping and release charge transport in amorphous semiconductors exemplified by hydrogenated amorphous silicon, *J. Non-Cryst. Solids* **654**, 123436 (2025).
- [35] A. Einstein, On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat, *Ann. Phys.* **17**, 208 (1905).
- [36] P. Le Comber and W. Spear, Electronic transport in amorphous silicon films, *Phys. Rev. Lett.* **25**, 509 (1970).
- [37] G. Juška, K. Arlauskas, J. Kočka, M. Hoheisel, and P. Chabloz, Hot electrons in amorphous silicon, *Phys. Rev. Lett.* **75**, 2984 (1995).
- [38] X. Wang and A. Dodabalapur, Carrier velocity in amorphous metal-oxide-semiconductor transistors, *IEEE Trans. Electron Devices* **68**, 125 (2020).
- [39] J. Chévrier and B. Equer, High electric-field amorphous silicon *p-i-n* diodes: Effect of the *p*-layer thickness, *J. Appl. Phys.* **76**, 7415 (1994).
- [40] K. Sawada, C. Mochizuki, S. Akata, and T. Ando, Photocurrent multiplication in hydrogenated amorphous silicon *p-i-n* photodiode films, *Appl. Phys. Lett.* **65**, 1364 (1994).
- [41] S. Matsuda, *et al.*, 30-nm-channel-length c-axis aligned crystalline In-Ga-Zn-O transistors with low off-state leakage current and steep subthreshold characteristics, in *2015 Symposium on VLSI Technology (VLSI)* (IEEE, 2015), pp. T216.
- [42] K. Chen, *et al.*, Scaling Dual-Gate Ultra-thin a-IGZO FET to 30 nm Channel Length with Record-high G_m , max of $559 \mu\text{S}/\mu\text{m}$ at $V_{DS}=1 \text{ V}$, Record-low DIBL of 10 mV/V and Nearly Ideal SS of 63 mV/dec , in *2022 IEEE Symposium on VLSI Technology and Circuits (VLSI)* (IEEE, 2022), pp. 298.
- [43] S. Subhechha, *et al.*, First demonstration of sub-12 nm Lg gate last IGZO-TFTs with oxygen tunnel architecture for front gate devices, in *2021 Symposium on VLSI Technology* (IEEE, 2021), pp. 1.
- [44] S. Samanta, K. Han, C. Sun, C. Wang, A. Kumar, A. V.-Y. Thean, and X. Gong, Amorphous InGaZnO thin-film transistors with sub-10-nm channel thickness and ultrascaled channel length, *IEEE Trans. Electron Devices* **68**, 1050 (2021).
- [45] Y. Luo, University of Cambridge Repository (2025), <https://doi.org/10.17863/CAM.122621>.
- [46] Flatiron Institute Fast Multipole Libraries, <https://github.com/flatironinstitute/FMM3D> (Accessed August 2023).