How good are 2D transistors? An application-specific benchmarking study

Cite as: Appl. Phys. Lett. 118, 030501 (2021); https://doi.org/10.1063/5.0029712
Submitted: 15 September 2020 • Accepted: 04 January 2021 • Published Online: 21 January 2021

Hattan Abuzaid, Nicholas X. Williams and Aaron D. Franklin

ARTICLES YOU MAY BE INTERESTED IN

Perspective on ferroelectric, hafnium oxide based transistors for digital beyond von-Neumann computing
Applied Physics Letters 118, 050501 (2021); https://doi.org/10.1063/5.0035281

Optoelectronic intelligence
Applied Physics Letters 118, 160501 (2021); https://doi.org/10.1063/5.0040567

Perspective on the future of silicon photonics and electronics
Applied Physics Letters 118, 220501 (2021); https://doi.org/10.1063/5.0050117
How good are 2D transistors? An application-specific benchmarking study

Hattan Abuzaid, Nicholas X. Williams, and Aaron D. Franklin

AFFILIATIONS
1Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708, USA
2Department of Chemistry, Duke University, Durham, North Carolina 27708, USA

Author to whom correspondence should be addressed: aaron.franklin@duke.edu

ABSTRACT
The research community has invested heavily in semiconducting two-dimensional (2D) materials, such as transition metal dichalcogenides (TMDs). Their stability when scaled down to a few atoms thick makes them attractive candidates to replace or supplement silicon in many future technologies. Although this sentiment is prevalent, demonstrations of 2D field-effect transistors (FETs) often do not present their data in a way that enables a straightforward comparison. For example, some papers solely use mobility as the figure of merit, while others focus on unnormalized device on-current. Here, we benchmark the performance of a selection of 2D FETs with field-corrected metrics that allow a more accurate projection of their potential; while the demonstrated methods are by no means comprehensive, they provide insight into improved benchmarking of 2D FETs going forward. Importantly, we show that appropriate benchmarking requires consideration of the specific application, with the three dominant potential application areas of front-end-of-line (FEOL) high-performance FETs, back-end-of-line (BEOL) 3D-integrated FETs, and low-cost thin-film FETs (or TFTs) each demonstrated. We find that 2D materials have the potential to compete with silicon as the channel in scaled FEOL high-performance devices. Meanwhile, in BEOL applications, FETs from in situ synthesized 2D materials have performance limited by their low crystal quality – a result of the stringent thermal budget of BEOL fabrication, which necessitates the use of transferred 2D materials. In the TFT area, 2D materials are simpler to fabricate than their silicon-based counterparts and they are competitive with other material alternatives. As promising as these findings are, there remain many hurdles for 2D materials to overcome, including poor reliability, performance variability, and fabrication scalability. Continuous research effort, combined with appropriate benchmarking, is strongly encouraged.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0029712

Semiconducting two-dimensional (2D) materials, including transition metal dichalcogenides (TMDs) and X-enes such as black phosphorous (BP), are promising candidates for next-generation, aggressively scaled field-effect transistors (FETs). Their layered nature preserves their general behavior down to a monolayer, and their atomic thinness enables excellent electrostatic gating control of the channel. Among 2D materials, the TMD, molybdenum disulfide (MoS2), stands out as a strong choice for n-type transistors toward complementary logic, as evidenced by the numerous experimental demonstrations in the literature. Being the front-runner of 2D FETs, high-performance MoS2 FETs are often employed as the indicator for progress of 2D FETs toward replacing or supplementing state-of-the-art silicon technology. However, published reports frequently make that comparison using metrics that touch upon limited aspects of device performance instead of forming a holistic picture. For example, mobility is widely used as an ultimate figure of merit for the intrinsic quality of 2D channels even though devices with the highest reported mobilities do not necessarily have the highest on-state performance in terms of on-current (Ion).

There are two main concerns for placing a high value on mobility in scaled, high-performance 2D transistors. First, the process of extracting field-effect mobility is known for being unreliable due to the high contact resistance (forming at metal-2D interfaces), which dominates the total resistance of the device and obscures the intrinsic channel performance. Although most reports point to the likelihood of mobility underestimation, some also predict mobility overestimation from contact gating and other effects. These complications manifest in a broad range of experimentally reported mobilities from as low as 0.02 cm²/(V s) to as high as 320 cm²/(V s) for monolayer MoS2. Second, because mobility describes the frequency of scattering events during carrier transport, its relevance in ultra-scaled devices with ballistic channels is questionable. Altogether, these two
issues demonstrate that the widely adopted mobility metric is inadequate for singularly describing the promise of 2D FETs.

For evaluating the off-state performance of 2D devices, it is useful to look at the subthreshold swing (SS) and the on/off-current ratio $I_{\text{ON}}/I_{\text{OFF}}$. Nevertheless, these two metrics alone are not comprehensive indicators. The IEEE International Roadmap for Devices and Systems (IRDS) specifies absolute current value requirements such that $I_{\text{OFF}}$ is small enough for an acceptable leakage power consumption and $I_{\text{ON}}$ is large enough for a sufficient switching speed. Hence, ratios and swings are most meaningful when the terminal values are considered. Reporting $I_{\text{ON}}$ as a measure for active device performance is a good starting point, but there needs to be a thoughtful normalization to enable comparison across various technology platforms and different device configurations. For instance, it is unrealistic to compare the $I_{\text{ON}}$ magnitude of a dual-gated, high-k dielectric encapsulated, short-channel device with a SiO$_2$ back-gated, long-channel one. It is also unmethodical to compare the performance of devices at largely disparate gate voltage overdrives (gate-source voltage minus threshold voltage, $V_{\text{GS}} - V_{\text{th}}$) or drain-source voltages.

In addition to the important considerations for on- and off-state performance metrics, it is also critical to focus on the most relevant deliverables for a particular application. 2D devices have been motivated as having broad applicability in the transistor space; yet, their performance requirements differ significantly between areas from high-performance computing to thin-film applications. The most effective benchmarking approach must make appropriate comparisons among 2D device options while also putting these in the context of the target application.

In this work, we propose and demonstrate a benchmarking approach utilizing relevant and cross-compatible metrics to investigate the potential of 2D materials for use in high-performance (HP) transistors, back-end-of-line (BEOL, 3D-integrated) transistors, and thin-film transistors (TFTs), all of which have disparate processing and performance requirements (see Fig. 1). HP transistors power the most demanding applications, like server chipsets or state-of-the-art CPUs—maximizing performance and minimizing size are of paramount importance in these applications. BEOL transistors are embedded during the final chip processing steps for added functionality or enhanced performance. By utilizing 3D monolithic integration, BEOL transistors are added in the upper interconnect layers on top of the finished front-end-of-line (FEOL) stack. Process compatibility is the main consideration for this device category because of the process thermal budget and fabrication cost limits. Finally, TFTs sacrifice high performance and miniaturization for low cost and versatility, with fabrication simplicity being a major requirement. TFTs are better suited for applications like large-area and flexible electronics. An example of improved benchmarking for each of these 2D device application areas is provided herein, drawing from recently reported advances in the literature, including from the Applied Physics Letters special collection.

It has been a common practice to normalize $I_{\text{ON}}$ by dividing it by the device width. Expressing $I_{\text{ON}}$ in mA/µm or µA/µm makes it possible to correlate devices with an unequal width or integration density and is a critical part of proper benchmarking; however, this normalization is still insufficient for a straightforward comparison. Si FinFETs are extremely miniaturized owing to state-of-the-art fabrication capabilities, which are not available to scientists creating research-grade nanoelectronics; thus, 2D FETs typically have longer channels. This device geometry disparity is still not captured by expressing current per unit width. On the other hand, densely integrated, commercial Si FinFETs must adhere to the power density limit, which caps their drain-source voltage supply to a low value (0.7 V at the 5-nm technology node as per IEEE IRDS), while 2D FETs are frequently reported with large drain-source voltages to display their saturation behavior or compensate for their high total device resistance. We propose that

![Fig. 1](https://example.com/fig1.png)
these disparities can be normalized through extracting $I_{ON}$ from the saturation regime and including the source-to-drain electric field ($E_{SD}$) averaged across the channel length, which is expressed in $V/\mu m$, in the on-current metric. Hence, the maximum width-normalized on-current ($I_{Max}/\mu A/\mu m$) over $E_{SD}$ has units of ($\mu A/\mu m$)/$(V/\mu m)$ = $\mu A/V$. Although $I_{ON}$ can also be extracted from the linear region, that approach would introduce subjectivity into the selection of the point of extraction (the point of highest $I_{Max}/E_{SD}$ in the linear regime could have impractically low $I_{ON}$). Devices that do not saturate will also have their performance exaggerated vs. the ones that saturate as desired for logic transistors. While there are shortcomings to this metric, including the nonuniformity of the electric field from source-to-drain particularly in short-channel devices, the improvements it provides in including the impact of applied fields and relative lengths are considerable. This proposed benchmarking performance metric is used throughout this paper to analyze the potential of 2D FETs in each of the previously mentioned device categories.

To begin, we investigated 2D FETs targeted for applications as scaled, HP transistors. Benchmarking the performance of a selection of HP 2D FET demonstrations is shown in Fig. 2 against the 2020 IEEE IRDS projection for Si FinFETs. $I_{Max}/E_{SD}$ is presented (as extracted from the saturation regime) vs. with regard to the average sheet carrier concentration in the channel, $n_{2D} = C_{ox} \times (V_{GS} - V_{th}) / q$, where $C_{ox}$ is the gate oxide capacitance and $q$ is the elementary charge. The use of $n_{2D}$ streamlines comparison of devices with a dissimilar gate dielectric thickness, gate dielectric constant, gating configuration, and gate voltage overdrive. Finally, the legend in Fig. 2 highlights the device’s on/off-current ratio. This whole benchmarking scheme provides a basic, yet insightful, way to examine how devices with quite distinctive dimensions and configurations compare to one another.

The BP devices (Refs. 62–64) reported in Fig. 2 clearly outperform silicon in terms of normalized drive current; unfortunately, they suffer from a poor on/off-current ratio (less than $10^4$ for these specific devices) due to the small bandgap of BP. This has been a fundamental limitation of BP along with its instability in air. There is still promise for BP as researchers tackle these problems,40–43 and further investment in BP is warranted due to its unique position as a proven p-type 2D material. Quantum transport simulations on sub-10 nm monolayer BP FETs predict that they have the potential to meet the requirements of future technology nodes.

Another promising p-type material is WSe$_2$, which unlike BP, has a sizable bandgap. WSe$_2$ exhibits ambipolar conduction (having similar electron and hole Schottky barriers) that can be pushed toward favorable hole transport using high-work function metals such as Pd and Au.48–50 The monolayer WSe$_2$ device reported by Liu et al.63 comfortably exceeds 2020 IRDS HP requirements, albeit at a higher $n_{2D}$. The majority of MoS$_2$ devices depicted in Fig. 2 had high $I_{Max}/E_{SD}$ ratios that met or exceeded the 2020 IRDS HD requirement. Notably, the device demonstrated by Das et al.12 had substantial normalized drive current that outperformed its silicon counterpart. Further, that device12 is still expected to maintain its superior position at the same reduced $n_{2D}$ of IRDS projections. This is all assuming its current does not saturate as its $E_{SD}$ increases (moving downward in the plot). This assumption seems to be feasible considering that MoS$_2$ FETs are projected to have $I_{Max} > 1 \text{ mA}/\mu \text{m}$ for $n_{2D} > 20 \times 10^{12} \text{ cm}^{-2}$ at room temperature.51 This is higher than the $I_{ON}$ value of 0.854 mA/$\mu$m for 2020 IRDS HP.

While this benchmarking example for HP 2D FETs is not comprehensive, it does have distinct advantages compared to traditional approaches; for instance, if we were to use mobility as a sole performance indicator, the device described by Liu et al.63—with a reported field-effect mobility of $517 \text{ cm}^2/\text{V} \cdot \text{s}$—would be deemed the best device even though it has a low active on-state performance, as per the $y$-axis in Fig. 2. This inflated mobility likely stems from an extraction error, which is the main drawback for the adoption of mobility as a reliable metric despite its appeal as a material-related property. It is important to note that this benchmarking exercise ideally utilizes transistors that are operating in the same regime. If all benchmarked devices had their $I_{Max}/E_{SD}$ extracted from the linear regime of the output characteristics, then the relationship between the $y$-axis and the $x$-axis in the plot in Fig. 2 would include isometric lines of constant mobility, which should be extracted in the linear regime. However, most devices in Fig. 2 have $I_{Max}/E_{SD}$ extracted from the saturation regime, making their comparison to the saturated silicon benchmarks reasonable and any comparison to mobility unrealistic. Only the devices proposed by Zhang et al.54, Liu et al.51, Liu et al.51, and Wang et al.51 have metrics from the linear regime, which could suggest that their performance is overestimated; nevertheless, their incorporation in the comparison does not alter the findings. The use of the $I_{Max}/E_{SD}$ metric does carry the risk of negatively impacting devices where $I_{Max}$ is extracted at $V_{DS} > V_{DS,sat}$, particularly if the current truly does completely saturate (i.e., zero output resistance). However, there are several factors that mitigate

![FIG. 2. Performance benchmarking of a selection of 2D FET demonstrations in the category of HP transistors using the width-normalized on-current divided by the source-to-drain electric field vs the gate field-induced carrier concentration. 1L: monolayer, >3L: more than three layers thick. 2020 IRDS HP: IRDS projected specifications for high-performance logic transistors at the 5 nm node. 2020 IRDS HD: IRDS projected specifications for high-density or low-power logic transistors at the 5 nm node. The data points with a superimposed “×” indicate cases where $I_{Max}/E_{SD}$ was extracted from the linear region of the output characteristics ($I_{Max}$ vs $V_{DS}$) and, thus, may be exaggerated compared to the other points, which are from the saturation regime. Inset: schematic of a bottom-gated 2D FET (with an added top gate that is used in some reported devices) with key parameters highlighted.](image-url)
this risk: (1) most 2D FETs do not completely saturate; (2) extraction of $I_{\text{Max}}$ in the saturation regime is most often done just after $V_{\text{DS,Sat}}$ (true for all devices in Fig. 2 except for Bolshakov et al.); and (3) even when changing the extraction of $I_{\text{Max}}/E_{\text{SD}}$ for the 5 nm node Si FinFETs to occur at $V_{\text{DS}} = V_{\text{Sat}}$, they still fall below 50 on the $I_{\text{Max}}/E_{\text{SD}}$ axis, which is still below many of the 2D FET demonstrations.

Overall, it is best if $I_{\text{Max}}$ is extracted as close to $V_{\text{DS,Sat}}$ as possible for this benchmarking approach, but it is also not absolutely critical. Hence, the use of this $I_{\text{Max}}/E_{\text{SD}}$ vs $n_{2D}$ approach alleviates some of the confusion stemming from incorrectly extracted mobility values and focuses more on on-current performance for a particular drain and gate field, with some consideration of different channel lengths.

There is still a long road ahead with hurdles and obstacles for 2D materials to overcome before they are considered a worthy replacement to incumbent HP technologies. In fact, in the near future, it might be more practical for 2D materials to augment rather than supplant silicon technology. Scaled 2D FETs could initially find their way in applications with less-stringent performance requirements, such as memory (e.g., SRAM and eDRAM). One prominent area to integrate 2D FETs on silicon platforms is BEOL fabrication. 2D transistors could be embedded within the top layers of a chip during the final metallization processes to interconnect all the individual components in an integrated circuit (IC). Being in the final steps in the process, BEOL implementation is a low-temperature step with a strict thermal budget of 400–500 °C.95,96 At elevated temperatures, the already fabricated FEOL and other components would be adversely affected. This limitation is detrimental to the crystalline quality of 2D materials grown directly onto the BEOL layers, as evidenced by the compromised condition of 2D materials grown at low temperatures using scalable synthesis techniques like chemical vapor deposition (CVD).97–99

Nevertheless, there are some reports of 2D devices that are synthesized at temperatures below the process thermal limit.100–103

The active performance of a collection of BEOL-compatible FETs is benchmarked in Fig. 3 in relation to their maximum processing temperature. Here, BEOL compatibility is defined in two ways: (1) either the device can be processed at a temperature below the process limit, (2) it can be grown with full coverage on a sacrificial substrate and subsequently transferred and patterned on the target chip, (3) it can be grown with full coverage on a sacrificial substrate and subsequently transferred and patterned on the target chip (right side of the figure). The transfer approach appears to be more promising for 2D materials due to the aforementioned temperature constraint. The BEOL stack performance requirement depends on the FEOL stack it is integrated onto, and there is more focus on functionality for this type of application. Therefore, it is more informative to benchmark 2D FETs against other alternatives rather than a fixed roadmap projection.

The left-hand side of Fig. 3 (with maximum processing temperatures <500 °C) indicates the difficulty of in situ synthesis of 2D channels for BEOL FETs. It is important to note that the y-axis in Fig. 3 is on the logarithmic scale due to the wide range of reported performance. ITO seems like an attractive option for low-temperature, low-performance BEOL implementations with its high $I_{\text{ON}}/I_{\text{OFF}}$ ratio. Li et al.101 achieved their competitive result at an efficient $I_{\text{ON}}$ value of $8.4 \times 10^{12} \text{ cm}^{-2}$. Gretyak et al.102 scored the highest active performance while observing thermal budget limits, albeit at a lower $I_{\text{ON}}$ than ITO that is still acceptable for high-performance applications. Moving to the higher processing temperatures in the figure, an apparent upward trend in performance is observed for MoS$_2$ devices as processing temperature goes up. This suggests that ex situ synthesis is a more viable path for 2D materials in this category, where improved crystalline quality is achievable thanks to the higher thermal energy syntheses. Note that these high-temperature synthesized 2D materials will require further work to realize sufficient coverage, consistency, and transferability for BEOL FET applications.

Beyond their suitability for aggressively scaled transistors, 2D layered materials enjoy a multitude of unique electronic, optoelectronic, and mechanical properties.103–106 Those traits, coupled with the ability to form functional, solution-processed 2D films, paves the way for their incorporation into thin-film transistors (TFTs).95–96 TFTs do not have to meet stringent high-performance requirements; instead, they need to offer benefits such as low-cost fabrication, large-area synthesis, or substrate agnosticism. With cost being a dominant factor for TFT relevance, we present a similar benchmarking analysis as with the HP and BEOL FETs, but with fabrication complexity as the variable against active performance, as shown in Fig. 4.

Fabrication complexity for reported TFTs was derived by breaking down the major processing steps in each paper and scoring those quantitatively for difficulty. Scoring was done based on the sophistication of equipment used in the fabrication, along with the required time and energy for processing (see the supplementary material for detailed scoring of steps in each reference).

Interestingly, a spin-coated MoS$_2$ TFT accomplished the highest performance in this survey of TFTs. Although the authors used a simple spin coating step to deposit the films initially, their use of high-temperature annealing and e-beam evaporation for contact metal formation pushed their fabrication complexity toward the mid-range. The superiority of spin-coated MoS$_2$ TFTs here is made a bit uncertain as other demonstrations in the collection (Refs. 106 and 107) fall short in active performance and $I_{\text{ON}}/I_{\text{OFF}}$. While the dip-coated MoS$_2$ device in the study by Xi et al.107 did not reach the same performance level, its simple fabrication process could make it desirable for specific applications. As a less-established process, further work is needed to optimize dip-coated MoS$_2$-based devices and investigate proper post-processing.
steps to increase $I_{\text{max}}/E_{\text{GS}}$. The same could be said about the in-place printed carbon nanotube (CNT) demonstration by Lu et al. It is important to highlight that the devices reported by Gomes et al.,97 Xi et al.,98 and Higgins et al.100 were controlled using a global gate. While that configuration is common in research demonstrations due to its simplicity, it is not viable in commercial circuits where individual control via a local gate is compulsory. Therefore, the fabrication complexity of these demonstrations might actually be higher if they follow that mandate. Moreover, globally gated devices falsely enjoy lower contact resistance through the effect of contact gating. Consequently, their performance is slightly exaggerated. The widely used amorphous silicon (a-Si) (Refs. 102 and 103) provided reasonable performance yet suffered from a high fabrication complexity with a lengthy and complicated procedure. Overall, 2D materials show some promise for TFT applications owing to their fabrication simplicity; however, there are some key factors still to overcome in terms of reproducibility and scalability of the processes. Hence, significant further research is needed to optimize their processing to elevate their performance to viable levels.

To summarize, we benchmarked the performance of 2D FETs across three distinct device categories using proposed, field-corrected metrics. Our proposed scheme made it possible to compare devices with a dissimilar structure and under varying bias conditions. Even though these benchmarking exercises were not exhaustive and do have shortcomings, we were able to extract useful insight into the competitiveness of 2D-layered materials in distinct future technologies by taking the specific needs of each technology into consideration. An even more comprehensive approach to appropriate benchmarking that may address lingering challenges with the approach herein would be welcome; in the meantime, the proposed benchmarking methods provide a distinct improvement over benchmarking with a single performance metric, such as unnormalized on-current or mobility.

When 2D FETs are scaled down to the same dimensions as state-of-the-art Si FinFETs, we expect them to be a viable contender in HP transistors. MoS$_2$ appears to be the front-runner for n-type devices, while more work is needed to pinpoint a p-type material that offers similar performance. While this says nothing of the challenges related to synthesis, reproducibility, and process integration, at least from a device performance Perspective, the vision is clearer with this benchmarking approach.

A more feasible implementation for 2D materials in the nearer term is in BEOL applications. However, the crystal quality of 2D materials suffers from the thermal constraint of a BEOL process. Ex situ synthesis at higher temperatures and subsequent transfer appear to be a promising route toward commercial implementation. With growing interest in the added functionality of monolithic 3D-integrated devices, 2D materials are strong contenders for continuous consideration.

The advantageous properties of 2D materials and their compatibility with solution-phase processing make them a strong candidate for future applications owing to their fabrication simplicity; however, there are many requisites for technological success that could not be captured in our analysis. 2D materials are notorious for their performance variability, and their scalable fabrication techniques are not completely mature. A paradigm shift in current fabrication approaches might be needed instead of striving to fit the mold of incumbent methods. Nonetheless, the results reported so far by the community of researchers are encouraging and they warrant substantial investment into the betterment of this exciting class of semicon-ductor materials. We propose that researchers perform more targeted benchmarking in the analysis of their 2D devices, by considering the impact of relative electric fields and focusing on the appropriate metrics for a specific application.

SUPPLEMENTARY MATERIAL

See the supplementary material for a detailed breakdown on fabrication complexity scoring for reported TFTs in Fig. 4.

The authors would like to acknowledge the usefulness of the 2D Device Trends database (http://2d.stanford.edu/2D_Trends) created by researchers in the Pop Lab at Stanford while preparing this work. This work was supported in part by the National Science Foundation (Grant No. ECCS-1915814). This work was performed in part at the Duke University Shared Materials Instrumentation Facility (SМIF), a member of the North Carolina Research Triangle Nanotechnology Network (RTNN), which was supported by the National Science Foundation (Grant No. ECCS-1542015) as part of the National Nanotechnology Coordinated Infrastructure (NNCI).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES


Published under license by AIP Publishing
D. Lembke and A. Kis, ACS Nano 6(11), 10070 (2012).
D. Neuamaier, S. Pindil, and M. C. Lemme, Nat. Mater. 18(6), 525 (2019).


95. Z. Zhi, I. Murtaza, H. Meng, and W. Huang, BSC Adv. 7(28), 17387 (2017).


