

# SCIENTIFIC REPORTS



OPEN

## Material-Device-Circuit Co-optimization of 2D Material based FETs for Ultra-Scaled Technology Nodes

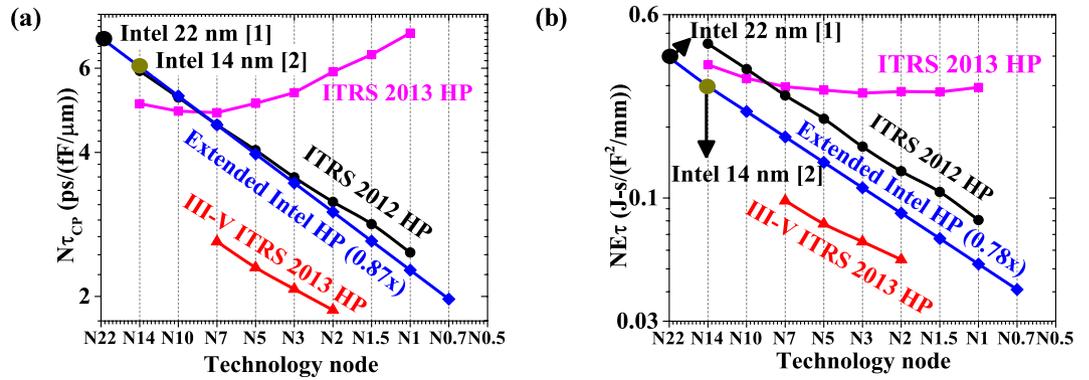
Tarun kumar Agarwal<sup>1,2</sup>, Bart Soree<sup>1,2</sup>, Iuliana Radu<sup>1</sup>, Praveen Raghavan<sup>1</sup>, Giuseppe Iannaccone<sup>3</sup>, Gianluca Fiori<sup>3</sup>, Wim Dehaene<sup>1,2</sup> & Marc Heyns<sup>1,2</sup>

Two-dimensional (2D) material based FETs are being considered for future technology nodes and high performance logic applications. However, a comprehensive assessment of 2D material based FETs has been lacking for high performance logic applications considering appropriate system level figure-of-merits (FOMs) e.g. delay, and energy-delay product. In this paper, we present guidelines for 2D material based FETs to meet sub-10 nm high performance logic requirements focusing on material requirement, device design, energy-delay optimization for the first time. We show the need for 2D materials with smaller effective mass in the transport direction and anisotropy to meet the performance requirement for future technology nodes. We present novel device designs with one such 2D material (monolayer black-phosphorus) to keep Moore's alive for the HP logic in sub-5 nm gate length regime. With these device proposals we show that below 5 nm gate lengths 2D electrostatics arising from gate stack design becomes more of a challenge than direct source-to-drain tunneling for 2D material-based FETs. Therefore, it is challenging to meet both delay and energy-delay requirement in sub-5 nm gate length regime without scaling both supply voltage ( $V_{DD}$ ) and effective-oxide-thickness (EOT) below 0.5V and 0.5 nm respectively.

To keep Moore's law alive, silicon based tri-gate FinFETs are being used for high performance logic at current technology nodes. With each technology generation, these devices achieve 15% boost in ON current, 50% reduction in energy-delay product, and 0.5x area scaling<sup>1,2</sup>. To further continue this trend, alternative channel materials e.g. SiGe, Ge, III-V, and novel device architectures e.g. gate-all-around nanowire (NW) FETs are being explored for future technology nodes. III-V materials due to its lower effective mass and electron-phonon scattering promise higher mobilities, thus higher ON currents for logic applications. But, the lower effective mass also poses challenges such as losing control on electrostatics with the scaling of channel length, and lower charge concentrations owing to limited density-of-states (DOS)<sup>3</sup>. Device architectures such as gate-all-around (GAA) FETs promise to achieve better electrostatics at scaled gate lengths.

Alternatively, 2D materials are considered for high performance logic roadmap due to their atomic thickness, which offer better scalability in comparison to Si and III-V channel FETs<sup>4</sup>. Within the 2D materials family, monolayer black phosphorus based FET has recently gained popularity as a promising high-performance (HP) logic device option at the end of the semiconductor roadmap due to its superior transport properties<sup>5</sup>. Monolayer (ML) BP shows anisotropic properties such as lower effective mass in armchair direction and 8x higher effective mass in zigzag direction. By aligning the ML BP channel length in armchair direction and channel width in zigzag direction we can achieve higher carrier velocity (mobility) and higher density-of-states (i.e. inversion charge density) respectively, which can effectively result in higher on-state currents. With full-band dissipative simulations, currents in monolayer black phosphorus (ML BP) FETs are reported to be significantly higher than other ML TMD based FETs<sup>6,7</sup>. The dissipative current in ML BP FET is shown to be around 90% of the ballistic current for 10 nm gate length. Having said that, the lower effective mass in the transport direction pose challenges in maintaining good sub-threshold slope below 10 nm gate lengths in ML BP FET in comparison to other TMD based FETs. Moreover, the efforts to have the stable BP under ambient condition are ongoing<sup>8,9</sup>. Nevertheless, to

<sup>1</sup>imec, Leuven, Belgium. <sup>2</sup>KUL, Leuven, Belgium. <sup>3</sup>University of Pisa, Pisa, Italy. Correspondence and requests for materials should be addressed to T.k.A. (email: [tarun.agarwalkumar@imec.be](mailto:tarun.agarwalkumar@imec.be))



**Figure 1.** High-performance logic roadmap, (a) Normalized delay ( $N\tau_{CP}$ ) and, (b) Normalized energy-delay product ( $NE\tau$ ) with technology nodes for  $\alpha = 0.1$ , and  $L_D = 10$  (HP logic).

evaluate real potential of such materials for high performance logic in sub-10 nm technology nodes, we need to co-optimize material and different device designs to achieve the required circuit-level metrics such as delay, and energy-delay product.

In this paper, using quantum transport simulations of monolayer 2D materials-based-FETs, we analyze the cause of such performance degradation in sub-10 nm gate length regime due to both material and device parameters. We propose device structures with ML BP to enable scaling of gate length in sub-5 nm regime. Further, for a given technology node, we show the selection of supply voltage ( $V_{DD}$ ) to achieve the required delay and an optimum energy-delay product (EDP).

## Results and Discussion

**Circuit-Level Requirements.** In order to benchmark the 2D FETs at the circuit level and understand the energy-delay tradeoff, we choose delay and energy per operation as circuit level figure-of-merits (FOMs). We estimate these circuit level metrics for a simplified version of critical path in CMOS logic, with a CMOS inverter chain and balanced 2D FETs for both p- and n-type transistors. The first-order equations for delay and energy per operation can be written as ref. 10:

$$\tau_{CP} = \frac{C_{node} V_{DD}}{I_{ON}} \cdot L_D; \quad E_{tot} = C_{tot} V_{DD}^2 (\alpha + L_D \frac{I_{OFF}}{I_{ON}}) \quad (1)$$

where  $\tau_{CP}$  is the delay of the critical path with a logic depth  $L_D$  and the total capacitance of each node  $C_{node}$ . Total energy ( $E_{tot}$ ) per operation can be written as sum of dynamic and leakage energy. Here,  $\alpha$ , and  $C_{tot}$  denote the activity factor and the total capacitance of the logic design respectively.

Further, we normalize the total energy and delay by the capacitance of the chip, which is reasonable for the sub-10 nm technology nodes, when the total capacitance is dominated by interconnect capacitances instead of intrinsic device capacitance, given as:

$$N\tau_{CP} = \frac{V_{DD}}{I_{ON}} \cdot L_D; \quad NE_{tot} = V_{DD}^2 (\alpha + L_D \frac{I_{OFF}}{I_{ON}}); \quad NE\tau = E_{tot} \cdot \tau_{CP} \quad (2)$$

here  $N\tau_{CP}$ , and  $NE_{tot}$  denote the normalized delay and total energy per operation respectively, while the energy-delay product ( $NE\tau$ ) signifies that energy and speed are equally weighed for an optimized logic design.

As shown in Fig. 1, we extend the circuit-level high performance logic roadmap for sub-5 nm technology nodes by extrapolating scaling of normalized delay and energy-delay product from Intel 22 nm<sup>1</sup> to Intel 14 nm technology nodes<sup>2</sup> with reported 15% boost in ON current (for same supply voltage) and required 50% reduction in the energy-delay product. Thus, the scaling of normalized delay by 0.87x and normalized energy-delay product by 0.78x results in total capacitance scaling of around 0.8x with each technology node. Figure 1 shows that extended Intel HP requirements seems most reasonable in comparison to ITRS HP requirements while III-V ITRS HP requirements are quite ambitious.

**Technology Requirements.** To achieve area scaling of 0.5x with each technology node, the technology parameters such as contacted gate pitch ( $C_{GP}$ ) and metal pitch (MP) are scaled by 0.7x with each technology generation. To scale  $C_{GP}$ , gate length ( $L_G$ ) scaling has been the primary driver for past technology generations. But, due to process constraints, scaling of  $C_{GP}$  below 25 nm is not forseen<sup>11</sup>. Therefore, for future technology nodes it is imperative to scale gate lengths in sub-10 nm to relax constraints on spacer thickness and contact openings. Alternatively, technology options such as monolithic 3D integration are sought to further scale the area per function<sup>12</sup>. The technology parameters listed in Table 1 (till N2) are taken from III-V ITRS HP roadmap<sup>13</sup>.

**Device-Level Figure-of-Merits.** We consider a double-gate monolayer 2D material based FET as shown in Fig. 2a. The electrical characteristics of 2D material based FETs are calculated using the framework described in methods section. The effect of different transport effective mass and channel lengths on ON current ( $I_{ON}$ ) is

	N7	N5	N3	N2	N1.5	N1	N0.7
$C_{GP}$ (nm)	42	32	25	?	?	?	?
$L_G$ (nm)	11.7	9.3	7.4	5.8	4.5	3.5	2.7
$V_{DD}$ (V)	0.61	0.58	0.56	0.54	?	?	?
$EOT$ (nm)	0.62	0.56	0.5	0.45	?	?	?
$I_{OFF}$ (A/m)	0.1	0.1	0.1	0.1	0.1	0.1	0.1

**Table 1.** III–V ITRS 2013 HP Roadmap extended for gate lengths below 5 nm. Supply voltage is denoted by  $V_{DD}$ . EOT denotes effective-oxide-thickness. Here, OFF current target ( $I_{OFF}$ ) for HP logic is chosen to be 100 nA/ $\mu$ m.

shown in Fig. 2b. We can clearly see that a smaller effective mass 2D material is the preferred choice for high performance logic. Smaller transport mass 2D materials with anisotropic properties can offer higher carrier injection velocity and higher inversion charge density, resulting in higher on-state current provided we can maintain good electrostatics with gate length scaling. To get physical insights in electrostatics of shorter gate length devices, we study the effect of transport effective mass on sub-threshold slope (S.S.) behavior for different gate lengths as shown in Fig. 2c. We break-down Fig. 2c into two regions: 1) At lower effective masses S.S. degrades due to direct S/D tunneling (due to material property); 2) At higher effective masses where the increase in S.S. with downscaling of gate lengths is attributed to 2D electrostatics.

**Circuit-Level Figure-of-Merits.** Figure 3 shows the combined effect of sub-threshold and super-threshold behavior on delay and energy-delay product for a given  $I_{OFF}$  of 100 nA/ $\mu$ m. We observe that till N3, 2D materials with smaller transport effective mass outperform the 2D materials with higher ones. It can be also seen that monolayer BP ( $m_x^* = 0.15 m_0$ ,  $m_y^* = 1.2 m_0$ ) FET can meet both extended Intel HP and III-V ITRS 2013 HP delay and energy-delay requirements for N7, and N5.

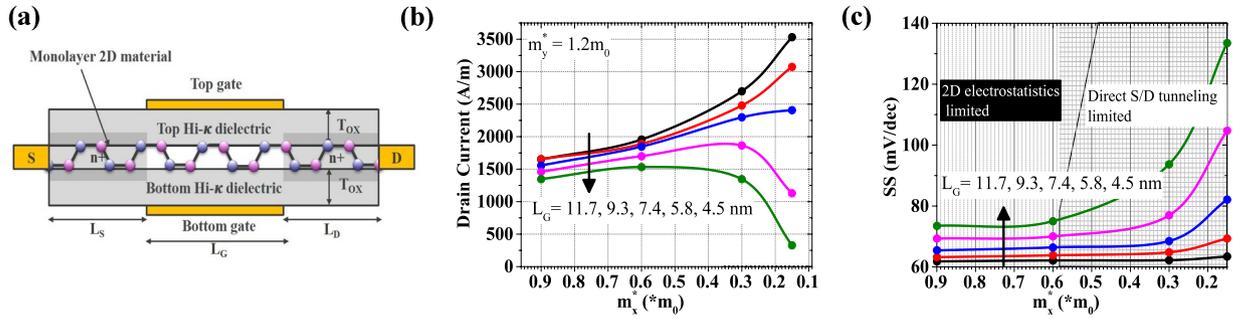
**Proposed Device Structures.** To further enable the HP logic roadmap with ML BP FETs, we need to improve electrostatics for sub-10 nm channel lengths with novel device designs. We propose device designs which address improving both 2D electrostatics and direct source-to-drain (S/D) tunneling.

*Improving 2-D Electrostatics.* We introduce a low- $\kappa$  interfacial layer (IL) between ML BP and High- $\kappa$  dielectric to reduce fringing fields due to the gate stack at shorter gate lengths<sup>14</sup>. As shown in Fig. 4a, by reducing fringing fields we improve both the gate control (i.e. slope of gate capacitance with gate voltage) and effective gate capacitance in ON state. Further, Fig. 4b shows that the performance of ML BP FETs at  $L_G = 7.4$  nm improves by more than 50% for the same effective-oxide-thickness (EOT) and physical thickness of the gate oxide ( $T_{OX}$ ). For effective-oxide-thicknesses above 0.5 nm, we consider low- $\kappa$  IL ( $SiO_2$ ) to be between 0.4–0.6 nm and High- $\kappa$  dielectric ( $HfO_2$ ,  $ZrO_2$ ,  $La_2O_3$ ) to be 1–1.5 nm thick. To meet both extended Intel HP and III-V ITRS 2013 HP delay requirement with the device structure having low- $\kappa$  IL, we can relax EOT requirements of the N3 technology node. Further, to see prospects of such gate stack with gate length scaling, we consider equivalent direct S/D tunneling probability i.e.  $\sim \exp(-L_G \cdot \sqrt{m_x^*})$  as shown in Fig. 4c. It shows that to achieve reasonable 2D electrostatics below 4.5 nm gate length, we require EOT scaling below 0.5 nm irrespective of the direct S/D tunneling.

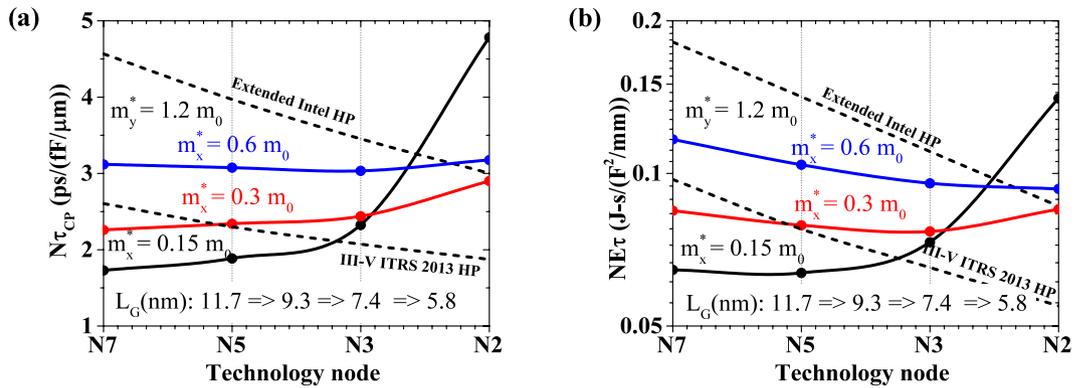
*Reducing Direct Source-to-Drain Tunneling.* We consider different device concepts (as shown in Fig. 5a) which employ depletion at the source/drain extension-to-channel junction in OFF state, resulting in larger tunneling lengths by modifying the potential profile at the junctions. Although, underlap (UL) and junctionless (JL) 2D material based FETs have been shown to improve direct source-to-drain tunneling at scaled gate lengths<sup>15</sup>, such designs alone can't provide required performance below 5 nm gate lengths as shown in Fig. 5b. To achieve the required performance for sub-5 nm gate lengths, we propose extended back-gate device architecture in conjunction with UL/JL FET, which makes it possible to meet the performance requirements till N0.7 ( $L_G = 2.7$  nm) for a fixed  $V_{DD}$ , and EOT. It is important to note that due to back-gate overlap in the extended back-gate architecture, an extra parasitic capacitance component as gate overlap capacitance comes in picture which may affect the total capacitance scaling, thus delay and energy-delay scaling. Nevertheless, Fig. 5c shows the need to scale  $V_{DD}$  to meet energy-delay requirement although the performance (delay) requirement is met till N0.7 for a fixed  $V_{DD}$ .

**Energy-Delay Optimization.** As shown in Fig. 6a, it is very challenging to meet both energy-delay and delay requirement even for smaller supply voltages for N1.5 and beyond. On the other hand, we see that the EOT requirement for N1.5 can be relaxed as shown in Fig. 6b, while Fig. 6c shows that we need to scale EOT below 0.5 nm to meet N1 requirements which scales the supply voltage. As EOTs below 0.5 nm become challenging to achieve using High- $\kappa$  dielectric with IL layer; it requires the advent of two-dimensional oxides with higher dielectric constant, and higher tunneling barrier with ML BP.

**Effect of contact resistance and scattering.** Lastly, to understand the limit on different contact resistances and different ballistic ratios, we first optimize the device structure consisting of High- $\kappa$  with IL and extended back-gate with underlap for technology node N3. The device parameters are taken from Table 1 and the optimized  $L_{UN}$  comes out to be 1 nm. As shown in Fig. 7a, both  $I_{ON}$  and  $N\tau_{CP}$  degrades by increasing contact resistance ( $R_c$ ). We notice the upper limit of contact resistance to be 125  $\Omega$ - $\mu$ m considering no scattering in the



**Figure 2.** Device Simulations, (a) Double-gate monolayer 2D material based FET with  $L_S = L_D = 15$  nm, and  $n^+$  doping of source/drain =  $4 \times 10^{13} \text{ cm}^{-2}$ , and for a fixed transverse effective mass ( $m_y^*$ ) effect of transport direction effective mass ( $m_x^*$ ) and gate length ( $L_G$ ) on, (b) ON current including effect of scattering and contact resistances, and, (c) Extracted subthreshold slope near OFF state.



**Figure 3.** Performance of DG monolayer 2D FET at different technology nodes (a) Normalized delay and (b) Energy-delay product, showing the effect of gate length scaling of a smaller effective mass material such as monolayer black phosphorus (BP) ( $m_x^* = 0.15 m_0, m_y^* = 1.2 m_0$ ) w.r.t. III-V ITRS 2013 HP and Extended Intel HP requirements.

channel. Further, Fig. 7b,c show that for  $R_C$ , ranging between 60 to  $100 \Omega\text{-}\mu\text{m}$ , we need to have ballisticity in the channel material between 85% to 60% respectively.

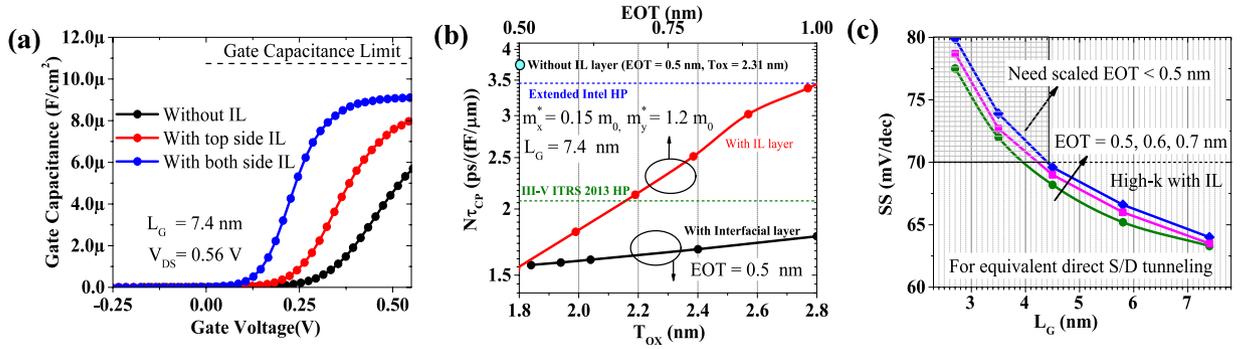
### Conclusions

In this paper, we show that monolayer black phosphorus based FETs with different device designs can fulfill the high-performance logic energy-delay requirements till sub-5 nm gate lengths. Although the monolayer black phosphorus is reported to be unstable under ambient conditions and efforts to have the stable BP are ongoing, we infer that lower transport effective mass 2D material such as monolayer BP (with proposed device designs) perform better than higher effective mass 2D materials. To boost the performance of 2D material FET for advanced technology nodes, we propose device structures consisting of High- $\kappa$  with IL (to increase the effective device gate capacitance), and extended back-gate with underlap (to curb direct source-to-drain tunneling). To meet the HP logic requirements, Table 2 lists the choice of device structure, and technology/device/circuit level parameters such as  $EOT/I_{ON}/V_{DD}$ . We see that for N1 and beyond, scaling of  $V_{DD}$  below 0.5 V becomes increasingly hard in order to meet both delay and energy-delay requirements, due to 60 mV/dec sub-threshold slope limit of FETs. It instigates the requirement of steep sub-threshold slope transistors with effective ON currents  $\sim 2000 \mu\text{A}/\mu\text{m}$ .

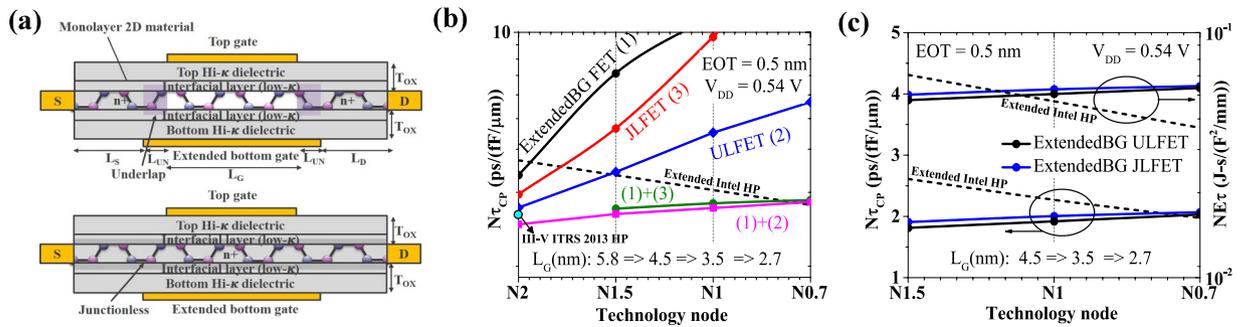
### Methods

The electrical characteristics of 2D material based FETs in the ballistic limit are calculated using a two-band tight binding (TB) Hamiltonian with a quantum transport simulation framework based on self-consistent solution of Poisson and Schrödinger equation with non-equilibrium Green's function within the NanoTCAD ViDES suite<sup>16</sup>. The two-band Hamiltonian for an anisotropic effective mass two-dimensional material with hexagonal lattice can be written as a  $2 \times 2$  Hamiltonian matrix:

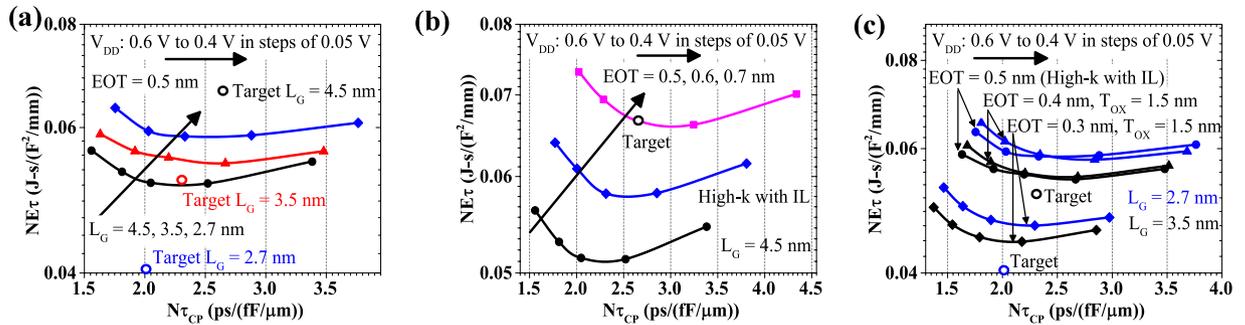
$$H_{2D} = \begin{bmatrix} E_{cm} & f(k) \\ f^*(k) & E_{vm} \end{bmatrix} \quad (3)$$



**Figure 4.** Effect of an interfacial layer (IL) with High-k dielectric, (a) On effective gate capacitance with gate voltage, (b) On performance (delay) of monolayer BP for N3 technology node i.e.  $L_G = 7.4$  nm in two different cases (i) with different EOT, and (ii) with varying physical oxide thickness for same EOT, (c) On the scalability of such gate-stack for different EOT considering equivalent direct source/drain tunneling.



**Figure 5.** Reducing direct S/D tunneling for N2 and beyond (a) Device structures showing extended back-gate (BG) with underlap (ULFET) and extended BG with junctionless doping profile. (JLFET) (b) Effect of different combination of device structures on performance of monolayer BP FET for N2 and beyond for fixed EOT and supply voltage, showing the need of extended back-gate with UL/JLFET for  $L_G = 4.5$  nm and beyond, (c) Delay and energy-delay product for extended BG UL/JLFET ( $L_{UN} = 2, 3, 4$  nm for N1.5, N1, and N0.7 respectively) showing that although we meet the performance requirement for N1 and N0.7, the energy-delay product doesn't scale.



**Figure 6.** Supply voltage for optimum energy delay product, (a) Effect of supply voltage on energy and delay for fixed EOT = 0.5 nm, and  $L_G = 3.5$ , and 2.7 nm. (b) Effect of EOT scaling on energy-delay Vs delay plot for  $L_G = 4.5$  nm, (c) For  $L_G = 3.5$  nm, and  $L_G = 2.7$  nm.

where  $E_{cm}$  and  $E_{vm}$  denote the bottom of conduction band, and top of the valence band. Further, bandgap ( $E_G$ ) of the material can be expressed as:  $E_G = E_{cm} - E_{vm}$ . Here, the  $f(k)$  function, due to nearest neighbors, can be written as:

$$f(k) = t_1 e^{-ik_x a / \sqrt{3}} + 2t_2 e^{ik_x a / 2\sqrt{3}} \cos\left(\frac{k_y a}{2}\right). \quad (4)$$



3. Del Alamo, J. A. Nanometre-scale electronics with III-V compound semiconductors. *Nature* **479**, 317–323 (2011).
4. Liu, L., Kumar, S. B., Ouyang, Y. & Guo, J. Performance Limits of Monolayer Transition Metal Dichalcogenide Transistors. *IEEE Transactions on Electron Devices* **58**, 3042–3047 (2011).
5. Li, L. *et al.* Black phosphorus field-effect transistors. *Nature nanotechnology* **9**, 372–377 (2014).
6. Szabo, A., Rhyner, R., Carrillo-Nunez, H. & Luisier, M. Phonon-limited performance of single-layer, single-gate black phosphorus n- and p-type field-effect transistors. In *2015 IEEE International Electron Devices Meeting (IEDM)*, 12–1 (IEEE, 2015).
7. Luisier, M. *et al.* First-principles simulations of 2-D semiconductor devices: Mobility, IV characteristics, and contact resistance. In *Electron Devices Meeting (IEDM), 2016 IEEE International*, 5–4 (IEEE, 2016).
8. Pei, J. *et al.* Producing air-stable monolayers of phosphorene and their defect engineering. *Nature communications* **7** (2016).
9. Illarionov, Y. Y. *et al.* Long-term stability and reliability of black phosphorus field-effect transistors. *ACS nano* **10**, 9543–9549 (2016).
10. Agarwal, T. *et al.* Effect of material parameters on two-dimensional materials based TFETs: An energy-delay perspective. In *2016 46th European Solid-State Device Research Conference (ESSDERC)*, 47–50 (2016).
11. Badaroglu, M. ITRS 2.0 - More Moore update. In *Focus Team Presentations* <http://www.itrs2.net/itrs-reports.html> (2015).
12. Batude, P. *et al.* 3-D Sequential Integration: A Key Enabling Technology for Heterogeneous Co-Integration of New Function With CMOS. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* **2**, 714–722 (2012).
13. Table PIDS 4, *ITRS updates online* <http://www.itrs2.net/2013-itrs.html> (2013).
14. Salmani-Jelodar, M. *et al.* Optimum High-k Oxide for the Best Performance of Ultra-Scaled Double-Gate MOSFETs. *IEEE Transactions on Nanotechnology* **15**, 904–910 (2016).
15. Cao, W., Kang, J., Sarkar, D., Liu, W. & Banerjee, K. 2D semiconductor FETs—Projections and design for sub-10 nm VLSI. *IEEE Transactions on Electron Devices* **62**, 3459–3469 (2015).
16. Fiori, G. & Iannaccone, G. *NanoTCAD ViDES* <http://vides.nanotcad.com> (2008).

## Acknowledgements

The authors acknowledge support through the imec's industrial affiliation programs on core CMOS. G.F. and G.I. gratefully acknowledge European Commission under Contract No. 696656 (Project GRAPHENE FLAGSHIP Core 1).

## Author Contributions

T.K.A. performed the device simulations with contributions from G.F. in the definition of the anisotropic 2D material and benchmarked the devices with proposed figure-of-merits with contributions from W.D. B.S., I.R., P.R., G.I. and M.H. helped interpreting the simulation results. T.K.A. wrote the manuscript with contributions from all authors.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017