Research Article | Vol. 26, No. 12 | 11 Jun 2018 | OPTICS EXPRESS 15445

Optics EXPRESS

Waveguide-based electro-absorption modulator performance: comparative analysis

Rubab Amin,[1] Jacob B. Khurgin,[2] and Volker J. Sorger[1,*]

[1] *Department of Electrical and Computer Engineering, George Washington University, 800 22nd St., Science & Engineering Hall, Washington, DC 20052, USA*
[2] *Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA*
[*] *sorger@gwu.edu*

Abstract: Electro-optic modulators perform a key function for data processing and communication. Rapid growth in data volume and increasing bits per second rates demand increased transmitter and thus modulator performance. Recent years have seen the introduction of new materials and modulator designs to include polaritonic optical modes aimed at achieving advanced performance in terms of speed, energy efficiency, and footprint. Such ad hoc modulator designs, however, leave a universal design for these novel material classes of devices missing. Here we execute a holistic performance analysis for waveguide-based electro-absorption modulators and use the performance metric switching energy per unit bandwidth (speed). We show that the performance is fundamentally determined by the ratio of the differential absorption cross-section of the switching material's broadening and the waveguide effective mode area. We find that the former shows highest performance for a broad class of materials relying on Pauli-blocking (absorption saturation), such as semiconductor quantum wells, quantum dots, graphene, and other 2D materials, but is quite similar amongst these classes. In this respect these materials are clearly superior to those relying on free carrier absorption, such as Si and ITO. The performance improvement on the material side is fundamentally limited by the oscillator sum rule and thermal broadening of the Fermi-Dirac distribution. We also find that performance scales with modal waveguide confinement. Thus, we find highest energy-bandwidth-ratio modulator designs to be graphene, QD, QW, or 2D material-based plasmonic slot waveguides where the electric field is in-plane with the switching material dimension. We show that this improvement always comes at the expense of increased insertion loss. Incorporating fundamental device physics, design trade-offs, and resulting performance, this analysis aims to guide future experimental modulator explorations.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (230.4110) Modulators; (250.7360) Waveguide modulators; (250.5403) Plasmonics; (160.2100) Electro-optical materials.

References and links

1. D. A. B. Miller, "Attojoule optoelectronics for low-energy information processing and communications," J. Lightwave Technol. **35**(3), 346–396 (2017).
2. V. J. Sorger, R. F. Oulton, R. M. Ma, and X. Zhang, "Toward integrated plasmonic circuits," MRS Bull. **37**(8), 728–738 (2012).
3. K. Liu, C. R. Ye, S. Khan, and V. J. Sorger, "Review and perspective on ultrafast wavelength-size electro-optic modulators," Laser Photonics Rev. **9**(2), 172–194 (2015).
4. K. Liu, N. Li, D. K. Sadana, and V. J. Sorger, "Integrated nanocavity plasmon light sources for on-chip optical interconnects," ACS Photonics **3**(2), 233–242 (2016).
5. N. G. Theofanous, M. Aillerie, M. D. Fontana, and G. E. Alexakis, "A frequency doubling electro-optic modulation system for Pockels effect measurements: application in $LiNbO_3$," Rev. Sci. Instrum. **68**(5), 2138–2143 (1997).
6. E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A review of lithium niobate modulators for fiber-optic communications systems," IEEE J. Sel. Top. Quantum Electron. **6**(1), 69–82 (2000).

#318686 https://doi.org/10.1364/OE.26.015445
Journal © 2018 Received 16 Jan 2018; revised 1 Apr 2018; accepted 14 May 2018; published 6 Jun 2018

7. I. Bar-Joseph, C. Klingshirn, D. A. B. Miller, D. S. Chemla, U. Koren, and B. I. Miller, "Quantum-confined Stark effect in InGaAs/InP quantum wells grown by organometallic vapor phase epitaxy," Appl. Phys. Lett. **50**(15), 1010–1012 (1987).
8. A. Melikyan, L. Alloatti, A. Muslija, D. Hillerkuss, P. C. Schindler, J. Li, R. Palmer, D. Korn, S. Muehlbrandt, D. Van Thourhout, B. Chen, R. Dinu, M. Sommer, C. Koos, M. Kohl, W. Freude, and J. Leuthold, "High-speed plasmonic phase modulators," Nat. Photonics **8**(3), 229–233 (2014).
9. U. Koch, C. Hoessbacher, J. Niegemann, C. Hafner, and J. Leuthold, "Digital plasmonic absorption modulator exploiting epsilon-near-zero in transparent conducting oxides," IEEE Photonics J. **8**(1), 4800813 (2016).
10. T. Stauber, N. M. R. Peres, and A. K. Geim, "Optical conductivity of graphene in the visible region of the spectrum," Phys. Rev. B **78**(8), 085432 (2008).
11. V. P. Gusynin, S. G. Sharapov, and J. P. Carbotte, "Magneto-optical conductivity in graphene," J. Phys. Condens. Matter **19**(2), 026222 (2007).
12. S. Schmitt-Rink, D. S. Chemla, and D. A. B. Miller, "Theory of transient excitonic optical nonlinearities in semiconductor quantum-well structures," Phys. Rev. B Condens. Matter **32**(10), 6601–6609 (1985).
13. B. Lee, W. Liu, C. H. Naylor, J. Park, S. C. Malek, J. S. Berger, A. T. C. Johnson, and R. Agarwal, "Electrical tuning of exciton-plasmon polariton coupling in monolayer MoS_2 integrated with plasmonic nanoantenna lattice," Nano Lett. **17**(7), 4541–4547 (2017).
14. S. M. Sze, *Physics of Semiconductor Devices*, 2nd edition (Wiley, New York, 1981).
15. S. Nashima, O. Morikawa, K. Takata, and M. Hangyo, "Measurement of optical properties of highly doped silicon by terahertz time domain reflection spectroscopy," Appl. Phys. Lett. **79**(24), 3923–3925 (2001).
16. F. Michelotti, L. Dominici, E. Descrovi, N. Danz, and F. Menchini, "Thickness dependence of surface plasmon polariton dispersion in transparent conducting oxide films at 1.55 μm," Opt. Lett. **34**(6), 839–841 (2009).
17. M. Noginov, L. Gu, J. Livenere, G. Zhu, A. Pradhan, R. Mundle, M. Bahoura, Y. A. Barnakov, and V. A. Podolskiy, "Transparent conductive oxides: plasmonic materials for telecom wavelengths," Appl. Phys. Lett. **99**(2), 021101 (2011).
18. A. P. Vasudev, J. H. Kang, J. Park, X. Liu, and M. L. Brongersma, "Electro-optical modulation of a silicon waveguide with an "epsilon-near-zero" material," Opt. Express **21**(22), 26387–26397 (2013).
19. Z. Ma, Z. Li, K. Liu, C. Ye, and V. J. Sorger, "Indium-tin-oxide for high-performance electro-optic modulation," Nanophotonics **4**(1), 198–213 (2015).
20. G. V. Naik, V. M. Shalaev, and A. Boltasseva, "Alternative plasmonic materials: beyond Gold and Silver," Adv. Mater. **25**(24), 3264–3294 (2013).
21. V. J. Sorger, R. Amin, J. B. Khurgin, Z. Ma, H. Dalir, and S. Khan, "Scaling vectors of attoJoule per bit modulators," J. Opt. **20**(1), 014012 (2018).
22. M. Liu, X. Yin, E. Ulin-Avila, B. Geng, T. Zentgraf, L. Ju, F. Wang, and X. Zhang, "A graphene-based broadband optical modulator," Nature **474**(7349), 64–67 (2011).
23. Z. Ma, M. H. Tahersima, S. Khan, and V. J. Sorger, "Two-dimensional material-based mode confinement engineering in electro-optic modulators," IEEE J. Sel. Top. Quantum Electron. **23**(1), 81–88 (2017).
24. R. F. Oulton, V. J. Sorger, D. A. Genov, D. F. P. Pile, and X. Zhang, "A hybrid plasmonic waveguide for subwavelength confinement and long-range propagation," Nat. Photonics **2**(8), 496–500 (2008).
25. V. J. Sorger, Z. Ye, R. F. Oulton, Y. Wang, G. Bartal, X. Yin, and X. Zhang, "Experimental demonstration of low-loss optical waveguiding at deep sub-wavelength scales," Nat. Commun. **2**, 331 (2011).
26. R. Amin, C. Suer, Z. Ma, I. Sarpkaya, J. B. Khurgin, R. Agarwal, and V. J. Sorger, "A deterministic guide for material and mode dependence of on-chip electro-optic modulator performance," Solid-State Electron. **136**, 92–101 (2017).
27. R. Amin, C. Suer, Z. Ma, I. Sarpkaya, J. B. Khurgin, R. Agarwal, and V. J. Sorger, "Active material, optical mode and cavity impact on nanoscale electro-optic modulation performance," Nanophotonics **7**(2), 455–472 (2017).
28. S. Sun, A. H. A. Badawy, V. Narayana, T. El-Ghazawi, and V. J. Sorger, "The case for hybrid photonic plasmonic interconnects (HyPPIs): low-latency energy-and-area-efficient on-chip interconnects," IEEE Photonics J. **7**(6), 1–14 (2015).
29. K. Liu, S. Sun, A. Majumdar, and V. J. Sorger, "Fundamental scaling laws in nanophotonics," Sci. Rep. **6**(1), 37419 (2016).
30. C. Ye, K. Liu, R. A. Soref, and V. J. Sorger, "A compact plasmonic MOS-based 2×2 electro-optic switch," Nanophotonics **4**(3), 261–268 (2015).
31. V. J. Sorger, N. D. Lanzillotti-Kimura, R. M. Ma, and X. Zhang, "Ultra-compact silicon nanophotonic modulator with broadband response," Nanophotonics **1**(1), 17–22 (2012).
32. R. M. Ma, R. F. Oulton, V. J. Sorger, G. Bartal, and X. Zhang, "Room-temperature sub-diffraction-limited plasmon laser by total internal reflection," Nat. Mater. **10**(2), 110–113 (2011).
33. Z. Ma, S. Khan, M. Tahersima, and V. J. Sorger, "Temperature dependence of a sub-wavelength compact graphene plasmon-slot modulator," https://arXiv:1709.01465 (2017).
34. V. J. Sorger and X. Zhang, "Physics. Spotlight on plasmon lasers," Science **333**(6043), 709–710 (2011).
35. V. J. Sorger, R. F. Oulton, J. Yao, G. Bartal, and X. Zhang, "Plasmonic Fabry-Pérot nanocavity," Nano Lett. **9**(10), 3489–3493 (2009).
36. N. Li, K. Liu, V. J. Sorger, and D. K. Sadana, "Monolithic III–V on silicon plasmonic nanolaser structure for optical interconnects," Sci. Rep. **5**(1), 14067 (2015).

37. C. T. Phare, Y.-H. D. Lee, J. Cardenas, and M. Lipson, "Graphene electro-optic modulator with 30 GHz bandwidth," Nat. Photonics **9**(8), 511–514 (2015).
38. M. Ayata, Y. Fedoryshyn, W. Heni, B. Baeuerle, A. Josten, M. Zahner, U. Koch, Y. Salamin, C. Hoessbacher, C. Haffner, D. L. Elder, L. R. Dalton, and J. Leuthold, "High-speed plasmonic modulator in a single metal layer," Science **358**(6363), 630–632 (2017).
39. C. Haffner, W. Heni, Y. Fedoryshyn, J. Niegemann, A. Melikyan, D. L. Elder, B. Baeuerle, Y. Salamin, A. Josten, U. Koch, C. Hoessbacher, F. Ducry, L. Juchli, A. Emboras, D. Hillerkuss, M. Kohl, L. R. Dalton, C. Hafner, and J. Leuthold, "All-plasmonic Mach–Zehnder modulator enabling optical high-speed communication at the microscale," Nat. Photonics **9**(8), 525–528 (2015).
40. E. N. Adams, "Motion of an electron in a perturbed periodic potential," Phys. Rev. **85**(1), 41–50 (1952).
41. Y. Yafet, "The *g* value in conduction electron spin resonance," Phys. Rev. **106**(4), 679–684 (1957).
42. E. O. Kane, "Band structure of indium antimonide," J. Phys. Chem. Solids **1**(4), 249–261 (1957).
43. A. Chernikov, C. Ruppert, H. M. Hill, A. F. Rigosi, and T. F. Heinz, "Population inversion and giant bandgap renormalization in atomically thin WS_2 layers," Nat. Photonics **9**(7), 466–470 (2015).
44. Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, "Probing excitonic dark states in single-layer tungsten disulphide," Nature **513**(7517), 214–218 (2014).

1. Introduction

Applications in data centers (core computing) and the looming era of sensor-driven internet-of-things (edge computing) drive demand for data bandwidth. Given the limitation of charging electrical wires, the need for optical communication is made [1]. Thus, electrical control of optical signals (electro-optic, EO) conversion remains a critical function [2]. However, the weak light–matter interaction (LMI) of Silicon or III-V photonics electro-optic modulators (EOM) requires footprints and energy-per-bit functions that are orders of magnitude higher compared to their electronic 3-terminal switching counterparts (i.e. transistors). The challenge is to obtain the highest optical modal index change with the least amount of voltage (i.e. steepest switching), while observing optical loss limitations [3,4], under a fundamental constraint given by the Kramers-Kronig relationship. While EO modulation can be enabled by either changing the real part (n) of the modal refractive index leading to phase shifting-based interferometer-like devices termed here electro-optic modulators (EOM), this work focuses solely on physical effects modulating the imaginary part (κ) leading to the class of electro-absorptive modulators (EAM). In both types, the fundamental complex index of refraction is altered electrically inside the active material, modifying the optical mode's propagation constant inside the waveguide. Assuming on-off-keying, EOMs always require an interferometric scheme to induce an amplitude modulation which fundamentally requires more real estate on-chip compared to EAMs, which, in contrast, induce optical amplitude changes directly in a linear device design. Furthermore, here we make the distinction between current-driven and voltage-driven modulators; in the former the index change is a result of injecting carriers into (or removing from) the active region thus enabling (disabling) the optical transitions and therefore adding (subtracting) to the oscillator strength inside the active material layer. We only consider current-driven (i.e. 'gated') modulation effects in this work. In contrast, in voltage-driven modulators no current flows in/out of the active region, and the change in index is induced by the energy level shifts and oscillator strength change induced by the electric field (e.g. Stark effect). The widely used lithium niobate ($LiNbO_3$) modulators based on Pockels effect and III-V semiconductor based quantum-confined Stark effect modulators are two examples of voltage driven EO modulators [5–7]. Recent current-driven modulators (e.g. state-of-the-art plasmonic modulators), on the other hand, have shown promising performance with respect to footprint, speed and power consumption due to their ability to confine the optical mode and hence increase the interaction of photons with the active material [8,9]. The spectrum of active current-driven materials and their respective modulation effects is wide and includes III-V based a) quantum dots (QDs) and b) quantum wells (QWs), free carrier based c) Silicon (Si) and d) Indium-tin-oxide (ITO), and emerging materials such as atomically-thin e) graphene and f) transition metal dichalcogenides (TMDs). While the promise of improved device performance has been made in an ad hoc manner, a systematic analysis and subsequent

performance comparison between optical mode, active material mechanism, and subsequent device performance tradeoffs is yet outstanding. Here we show such comparison for the first time where we contrast the different modulation mechanisms-material combinations, to include III-V based band-filling a) QDs and b) QWs, free carrier based c) Si and d) ITO, and novel 2D materials such as e) graphene's Pauli Blocking and f) exciton modulation in transition metal dichalcogenides (TMDs). We show that the modulation performance based on free-carriers falls short of all other mechanisms considered, while the remaining materials perform about equally. We further show that the key to improved performance lies in the waveguide design where the benefits of tight confinement must be weighed against the increased insertion loss.

2. Absorption modulator definitions

2.1 Effective area of the waveguide

A generic picture of a current-driven EAM is given in Fig. 1, defining the modulator length, L and cross-sectional width, W. The active layer (thickness, d_a) is sandwiched between two cladding layers for gating. With applied drive voltage, V_d the drive current, I_d injects charges into the active layer. Charge, –Q in the active layer and + Q in the gate, are formed by such electrostatic gating. Key to the modulator's performance is a strong LMI, which points towards requiring a small effective mode area, S_{eff}, or simply in one dimension (1D), a short effective thickness, t_{eff}.

To determine the effective area of the waveguide we consider electromagnetic wave propagating along the z-direction with the propagation constant $\mathrm{E} = \mathrm{E}(x, y)e^{i(\beta z-\omega t)} + c.c.$ The total power flow is then $P = n_{eff}E_{a0}^2 S_{eff} / 2\eta_0$ where E_{a0} is the magnitude of the transverse electric field in the middle of active layer (Fig. 1), $\eta_0 = 377\Omega$ is the impedance of free space, and the effective index has been introduced as $n_{eff} = \beta c / \omega$. The effective area of the waveguide as shown in Appendix A is

$$S_{eff} = \frac{1}{n_{eff}^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n^2(x, y)\left(E_x^2 + E_y^2\right) dxdy / E_{a0}^2 \tag{1}$$

This definition of the effective area may differ from others in the literature, but the difference is small and involves the distinction between the effective and group indices. We note that this definition includes the fact that the active layer is not necessary at the center of the waveguide, i.e. E_{a0} may not be the peak electric field.

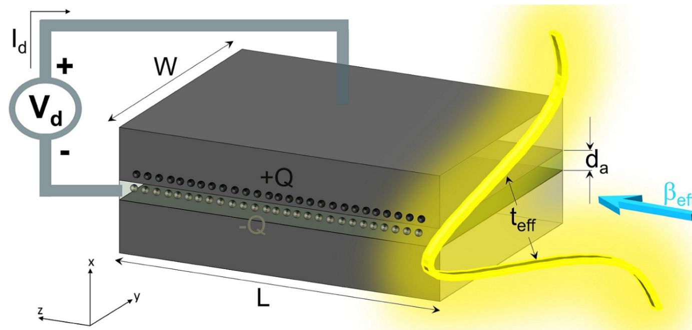

Fig. 1. Schematic of the waveguide structure with a biasing scheme to the active region (drive voltage, V_d and current (charge) flow into the active layer, I_d), the charge -Q in the active layer and + Q in the gate is induced, and propagation direction (indicated with blue arrow, β_{eff}). The associated electric field in the y-direction is shown. d_a is the width of the active region and t_{eff} is the effective thickness. The relevant coordinate system used in this work is also included to accompany the text.

Optics EXPRESS

2.2 Absorption cross-section and effective thickness

Let us now estimate the absorption change induced by the injection (or depletion of carriers). We first consider two level absorbers, which can be atoms or quantum dots (QDs) with absorption coefficient, according to Appendix B, as

$$\alpha(\omega) = \sigma_a(\omega)\frac{WF}{S_{eff}}N_{2D} = \frac{\sigma_a(\omega)}{t_{eff}}N_{2D} \tag{2}$$

where σ_a is the absorption cross-section, N_{2D} is the two-dimensional density of carriers, and F is the uniformity function (Appendix B). The effective thickness of the waveguide has been introduced as

$$t_{eff} = S_{eff} / WF = S'_{eff} / W \tag{3}$$

where $S'_{eff} = S_{eff} / F$. For the waveguides in which the field does not change much laterally, which is often the case,

$$t_{eff} \approx \frac{1}{n_{eff}^2}\int_{-\infty}^{\infty} n^2(x)\left(E_x^2 + E_y^2\right)dx / E_{a0}^2 \tag{4}$$

2.3 Switching characteristics: definitions

Next, we introduce the absorption modulation characteristics. The absorption modulation is achieved by Pauli blocking. As electrons are injected into the upper level (or holes into the lower) the transitions are blocked and absorption is reduced. Here we (arbitrarily) select the maximum (minimum) transmission to be 90% (10%), for an extinction ratio (ER) of ~10 dB modulation. Minimum transmission (maximum absorption) is achieved when no electrons are injected

$$\alpha_{\max}(\omega)L = \sigma(\omega)N_{2D}L / t_{eff} = \ln(10) \approx 2.302 \tag{5}$$

whereas maximum transmission occurs when δn_{2D} carriers are injected by applying the positive voltage to the gate,

$$\alpha_{\min}(\omega)L = \sigma(\omega)\left(N_{2D} - \delta n_{2D}\right)L / t_{eff} = -\ln(0.9) \approx 0.105 \tag{6}$$

Subtracting (6) from (5) we obtain

$$\delta n_{2D}L \approx 2.2 t_{eff} / \sigma(\omega) \tag{7}$$

and when multiplied by the waveguide width and length, we obtain the expression and using (4) to evaluate one of the key modulator characteristics – the switching charge

$$Q_{sw} = eWL\delta n_{2D} = 2.2eWt_{eff} / \sigma(\omega) = 2.2eS_{eff} / \sigma(\omega)F \tag{8}$$

We find, that the switching charge depends only on the ratio of the effective cross-section of the waveguide and the 'effective absorption cross-section' of the material $\sigma(\omega)F$. One can then determine the switching voltage as

$$V_{sw} = \frac{d_{gate}e\delta n_{2D}}{\varepsilon_0\varepsilon_{eff}} = 2.2\frac{e}{\varepsilon_0\varepsilon_{eff}}\frac{t_{eff}}{L}\frac{d_{gate}}{\sigma(\omega)} \approx \frac{e}{\varepsilon_0\varepsilon_{eff}}N_{2D}d_{gate} \tag{9}$$

where d_{gate} and ε_{eff} are the thickness and dielectric constant of the insulator between the active layer and the gate. In the last step (5) was used which implicitly indicates that minimum absorption is achieved when nearly all the absorbers are bleached. Note, the voltage depends only on the density of the QDs and thus can be adjusted within reasonable limits, say from 10^{10} to 10^{12} cm^{-2}. Similarly, according to (5) we have the freedom of choosing the length of the modulator as long as

$$L \approx 2.302 t_{eff} / N_{2D}\sigma(\omega) \tag{10}$$

Therefore, the capacitance of the modulator can be found as

$$C = \varepsilon_0 \varepsilon_{eff} WL / d_{gate} = 2.3 \frac{\varepsilon_0 \varepsilon_{eff}}{d_{gate}} \frac{W t_{eff}}{N_{2D}\sigma_a} = 2.3 \frac{\varepsilon_0 \varepsilon_{eff}}{d_{gate}} \frac{S_{eff}}{N_{2D}\sigma_a F} \tag{11}$$

The 3-dB cut-off frequency of the modulator then becomes

$$f_{3dB} = 1/2\pi RC \approx \frac{d_{gate}}{\varepsilon_0 \varepsilon_{eff}} \frac{N_{2D}\sigma_a F}{14.5 S_{eff} R} \tag{12}$$

where R is the resistance. Finally, we can determine the switching energy (per bit) as

$$U_{sw} = \frac{1}{2} Q_{SW} V_{SW} \approx \frac{e^2 d_{gate}}{\varepsilon_0 \varepsilon_{eff}} \frac{N_{2D} S_{eff}}{\sigma_a(\omega) F} \tag{13}$$

One should note that while the switching charge is rather rigidly defined by (8), in principle it is often possible to reduce the switching energy per bit $U_{sw} = \frac{1}{2} Q_{sw}^2 / C$ by simply increasing the capacitance (reducing gate oxide thickness, or simultaneously decreasing N_{2D} and increasing length), which evidently reduces the speed. Ignoring bandwidth for a moment, the switching energy itself cannot serve as a holistic performance metric for modulators. Hence, we introduce a more relevant figure of merit – the ratio of switching energy to the 3dB cut-off frequency or energy-bandwidth-ratio (EBR) as

$$EBR = U_{SW} / f_{3dB} \approx 14.5 e^2 R \frac{S_{eff}^2}{\sigma_a^2(\omega) F^2} \tag{14}$$

This expression is strikingly simple as it does not only depend on the intrinsic absorption strength of the active material, $\sigma_a(\omega)$, but also on the degree of waveguide confinement, S_{eff}, and the efficiency of coupling between the electric field in the waveguide and the active medium, F. Note that EBR is measured in units of J/Hz (or more conveniently fJ/THz) and indicates that for a constant EBR, the power consumption of the modulator increases quadratically with its speed (bandwidth). We shall see that this result leads to severe limitations of the performance. First, we shall investigate the intrinsic absorption strengths by considering four different material-based modulation mechanisms (Fig. 2).

3. Material classes for EAMs

To evaluate different modulation potentials offered by various different materials (both traditional and emerging active materials), here we chose to include four broad categories of material classes: a) two level absorbers, i.e. QDs; b) Pauli (transition) blocking based QWs and graphene; c) excitonic modulators in TMDs; and d) free-carrier schemes in Si and transparent conducting oxides, such as ITO. The changes in absorption for these broad classes of materials are schematically depicted in Fig. 2 upon active modulation. This sample space of broad categorical materials can suffice to describe most current-driven (i.e. charge) modulation mechanisms to date.

Optics EXPRESS

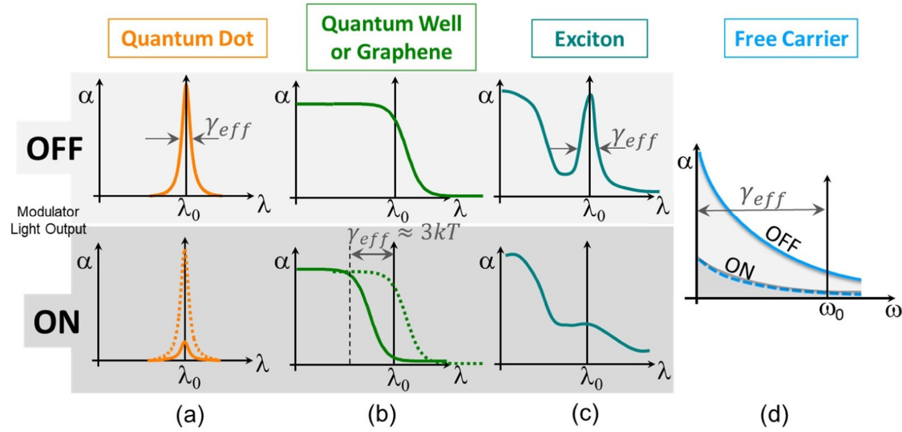

Fig. 2. Schematics of different absorption modulation mechanisms considered in this work for the OFF and ON states are for the attenuated and non-attenuated modulator output powers, respectively (a) quantum dot (QD), (b) quantum well (QW) or graphene, (c) excitons in transition metal dichalcogenides (TMDs), and (d) free carrier based modulation schemes (e.g. Si, ITO).

3.1 Modulator based on two-level absorbers

The first class of modulators we consider is two-level absorbers, a most practical example of which could be a semiconductor QD (Fig. 2(a)). The absorption cross-section in QDs is determined in Appendix B, and in Appendix C we find its value on resonance $\Delta\omega = 0$ as

$$\sigma_{QD}(\omega) = \frac{\pi\alpha_0}{n_{eff}} F_{cv} \frac{\hbar}{\gamma}\left(m_c^{-1} - m_0^{-1}\right) = \frac{\pi\alpha_0}{n_{eff}} F_{cv} \frac{\hbar^2 / m_0}{\hbar\gamma}\left(\frac{m_0}{m_c} - 1\right) = 1.75\times10^{-17} cm^2 \frac{F_{cv}}{\hbar\gamma n_{eff}}\left(\frac{m_0}{m_c} - 1\right) \quad (15)$$

where α_0 is the fine structure constant, the energy broadening $\hbar\gamma$ is given in electron volts (eV), m_c is the electron effective mass, and $F_{cv} = \left|\int \psi_c \psi_v dV\right|^2 <\sim 1$ is the overlap between the 'envelope' wave-functions of the conduction and valence bands. As one can perceive, the absorption cross-section does only depend on a few parameters; for instance, at resonance it depends only on the resonance width and the effective mass. Since the latter is essentially determined by the value of the bandgap we obtain, therefore, for a given photon energy $\hbar\omega$

$$\sigma_{QD}(\omega) = \frac{\pi\alpha_0}{n_{eff}} F_{cv} \frac{E_P}{\hbar\omega} \frac{\hbar^2 / m_0}{\hbar\gamma} \approx 3.5\times10^{-16} cm^2 \frac{F_{cv}}{(\hbar\gamma)(\hbar\omega) n_{eff}} \quad (16)$$

where $E_p \approx 16 - 24eV$ is defined in the Appendix C and all the energies are in eV. One can further show that if we introduce additional inhomogeneous broadening (possibly important in QDs, but less so for doped ions), it will simply add up as roughly $\gamma \approx \sqrt{\gamma_{Lorentz}^2 + \gamma_{inh.}^2}$ (Voigt profile), and one can still use the expressions (15) and (16). Thus in the end, the absorption cross-section depends mainly on the broadening. Finally, using (10) and (15) we obtain the required (i.e. for the an assumed 10dB extinction ratio) length of QD modulator as

$$L_{QD} \approx 2.3 \frac{n_{eff} t_{eff}}{\pi\alpha_0 N_{QD}} \frac{\gamma}{F_{cv}\hbar\left(m_c^{-1} - m_0^{-1}\right)} \quad (17)$$

3.2 Free carrier accumulation-depletion style modulator

Let us consider now a free electron gas in the conduction band of a semiconductor such as in Si, ITO, or some other material whose density is $N_e(x,y)$. The schematics of this modulator case is shown in Fig. 2(d). The expression for the absorption cross-section is derived in Appendix D as

$$\sigma_{fc}(\omega)=\frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar}{m_c}\frac{\gamma}{\omega^2+\gamma^2}=\frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar^2}{m_0(\hbar\gamma_{eff})}\frac{m_0}{m_c}\approx\frac{7.02\times10^{-17}cm^2}{(\hbar\gamma_{eff})}\times\frac{m_0}{m_c} \quad (18)$$

where the effective detuning is $\gamma_{eff}=\gamma+\omega^2/\gamma$. As one can see this expression is similar to the one for the QDs with one major difference – due to non-resonant character of the absorption the cross section for the free carriers is orders of magnitude lower than that for the resonant QDs. Also, we see that for the wavelength in the telecom range the cross-section actually scales with broadening caused by scattering γ. Therefore, as shown in the next section, ITO leads intrinsically to more modulation per unit charge than high-quality Silicon. Note, the maximum absorption is achieved at $\gamma=\omega$ when $\gamma_{eff}=2\omega$ and

$$\sigma_{fc.\max}(\omega)=\frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar}{m_c\omega}\times\frac{1}{2}=3.53\times10^{-17}cm^2\frac{1}{\hbar\omega n_{eff}}\frac{m_0}{m_c} \quad (19)$$

3.3 Two-dimensional electron gas in semiconductor QWs

Let us consider a N_{QW} number of quantum wells inside a waveguide, where each of the QWs is populated with the 2D carrier density n_{2D} that are distributed according to Fermi-Dirac distribution $f_c(\mathrm{E}_c)=\{1+\exp[(\mathrm{E}_c-\mathrm{E}_f)/kT]\}^{-1}$ (Fig. 2(b)). Since the position of the Fermi level, E_f changes with carrier density, δn_{2D} so does the absorption. The expression for the absorption coefficient is obtained in Appendix E as

$$\alpha_{QW}=\frac{1}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{WFN_{QW}}{S_{eff}}F_{cv}H(\hbar\omega-\mathrm{E}_g)[1-f_c] \quad (20)$$

where F is described by Appendix B, Eq. (53) and the effective thickness is $t_{eff}=S_{eff}/WF$, same as (3); H is a Heaviside step function and effective mass ratio factor $0<\beta_v<0.5$ has been introduced. Next, we want to evaluate the change in absorption occurring when the 2D density of carriers changes. Differentiating (20) over the electron density results in

$$\frac{d\alpha_{QW}}{dn_{2D}}=\frac{\partial\alpha_{QW}}{\partial f_c}\frac{\partial f_c}{\partial\mathrm{E}_{fc}}\frac{d\mathrm{E}_f}{dn_{2D}}=-\sigma_{QW}(\omega)t_{eff}^{-1} \quad (21)$$

where the differential absorption cross-section is

$$\sigma_{QW}(\omega)=\frac{F_{cv}}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{\pi\hbar^2}{m_ckT}\frac{\exp\left[(\mathrm{E}_c-\mathrm{E}_f)/kT\right]\left[1+\exp(-\mathrm{E}_f/kT)\right]}{\left\{1+\exp\left[(\mathrm{E}_c-\mathrm{E}_f)/kT\right]\right\}^2} \quad (22)$$

and it reaches its maximum value when both the photon energy is equal to the bandgap and the Fermi level is also at the band-edge, i.e. $E_c=E_{fc}=0$, resulting in:

$$\sigma_{QW-\max}(\omega) = \frac{F_{cv}}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{\pi\hbar^2}{2m_c kT} \approx 2.75\times10^{-17}cm^2\frac{1}{kTn_{eff}}\frac{m_0}{m_c}\frac{F_{cv}}{1+\beta_v} \quad (23)$$

The result is nearly identical to (16) – with the only difference being that kT plays the role of broadening. Note, the result does not depend on the number of QWs. Then we can proceed to estimating the change of absorption with induced charge, and the absorption becomes $\Delta\alpha_{QW} = \sigma_{QW}(\omega)\delta n_{2D}/t_{eff}$. This, however, will be an overestimation of the effect because we have linearized the dependence of absorption on density near $\mathrm{E}_f = 0$ in (21). Approaching the problem from a different angle by following discussion in section 3.1, the minimum transmission (maximum absorption) is achieved when no electrons are injected:

$$\alpha_{QW,\max}(\omega)L = \frac{F_{cv}}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{N_{QW}L}{t_{eff}}[1-f_{c,\min}] = \ln(10) \approx 2.302 \quad (24)$$

And, similar to the 2-level case before, the maximum transmission (minimum absorption) is achieved when δn_{2D} carriers are induced by the gate (injected).

$$\alpha_{QW,\min}(\omega)L = \frac{F_{cv}}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{N_{QW}L}{t_{eff}}[1-f_{c,\max}] = -\ln(0.9) \approx 0.105 \quad (25)$$

Now, let us assume that the material is intrinsic and $E_{f,\min} \approx -E_{gap}/2$, therefore $f_{c,\min}(\mathrm{E}_c = 0) = [1+\exp(-E_{f,\min}/kT)]^{-1} \approx 0$. The exact value of $f_{c,\min}$ does not influence the modulation as long as it is small. Then we obtain from (24), $\frac{F_{cv}}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{N_{QW}L}{t_{eff}} \approx 2.3$. Substituting this into (25) we obtain $f_{c,\max}(\mathrm{E}_c = 0) = [1+\exp(-\mathrm{E}_{f,\max}/kT)]^{-1} \approx 0.96$ and $\mathrm{E}_{f,\max} \approx 3.15kT$. Next, we can find $n_{2D,\max} \approx 3.2N_{QW}m_c kT/\pi\hbar^2$ while $n_{2D,\min} \approx 0$. Therefore, switching charge can be found as

$$Q_{SW} = eWL(n_{2D.\max} - n_{2D,\min}) = eW\times2.2\frac{t_{eff}}{N_Q}\frac{n_{eff}}{\pi\alpha_0}\frac{1+\beta_v}{F_{cv}}\times3.2\frac{m_c kT}{\pi\hbar^2}N_{QW} \approx 2.2e\frac{S_{eff}}{\sigma_{QW}(\omega)F} \quad (26)$$

which is exactly the same as (8), as long as we can formally re-introduce the effective differential cross-section (23) as

$$\sigma_{QW}(\omega) = \frac{F_{cv}}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{\pi\hbar^2}{m_c 3kT} \approx 5.4\times10^{-17}cm^2\frac{1}{3kTn_{eff}}\frac{m_0}{m_c}\frac{F_{cv}}{1+\beta_v} \quad (27)$$

Finally, one can combine the thermal broadening $3kT$ and the broadening γ to an effective broadening, $\hbar\gamma_{eff} = \sqrt{(\hbar\gamma)^2 + (3kT)^2}$ and modify (27) as

$$\sigma_{QW}(\omega) = \frac{1}{1+\beta_v}\frac{\pi^2\alpha_0}{n_{eff}}\frac{\hbar^2}{m_c\hbar\gamma_{eff}}F_{cv} \approx 5.4\times10^{-17}cm^2\frac{1}{\hbar\gamma_{eff}n_{eff}}\frac{m_0}{m_c}\frac{1}{1+\beta_v}F_{cv} \quad (28)$$

which has the same form as (15) producing similar results. However, when choosing the modulator length we now have less flexibility to assure that the maximum (off-state) absorption is satisfied according to

$$L_{QW} \approx 2.3 \frac{n_{eff} t_{eff}}{\pi \alpha_0 N_{QW}} \frac{1+\beta_v}{F_{cv}} \tag{29}$$

The only adjustable parameter is the number of QWs, which cannot be too large because the carriers will always tend to accumulate non-uniformly in the few QWs closest to the gate. In this respect, the Pauli-blocking type of modulator is inferior to quantum confined stark effect (QCSE) modulators in which no charge is stored. If we compare this with the result for QDs in (17), we can see that QD modulator would have the same length if the density of QDs is $N_{QD} \approx \gamma m_c / \hbar \approx 2\times10^{12} cm^{-2}$ which is a bit too high for self-organized QDs – hence the QW modulator has the advantage of requiring a shorter length, which in turn increases the attainable speed but also the required switching voltage.

3.4 Graphene

The operating principle of graphene for modulation is similar to that of the QW modulator and shown in Fig. 2(b). Even though there have been many derivations of graphene's optical absorption [10,11], it is instructive to re-derive the inter-band absorption in graphene from the first principle because it allows us to see how closely related it is to absorption by any other 2D structure whether there are any Dirac electrons involved in it. In fact, when the electrons involved in Pauli blocking are located at roughly 0.4 eV away from the Dirac point (assuming a wavelength of 1550 nm = 0.8 eV), all the peculiarity of graphene becomes irrelevant and as a result its performance as an electro-absorption modulator is conceptually no different from any other semiconductor material. The differential absorption cross-section for graphene obtained in Appendix F is

$$\sigma_{gr}(\omega) = \frac{\pi^2 \alpha_0}{n_{eff}} \times \frac{1}{kT} \frac{\exp\left[(\hbar\omega/2 - \mathrm{E}_{fc})/kT\right]}{\left\{1+\exp\left[(\hbar\omega/2 - \mathrm{E}_{fc})/kT\right]\right\}^2} \frac{\hbar^2 v_F^2}{2\mathrm{E}_f} \tag{30}$$

where $v_F \sim 10^8 cm/s$ is the Fermi velocity. This expression obviously peaks at $E_f = \hbar\omega/2$, and we obtain

$$\sigma_{gr}(\omega) = \frac{\pi^2 \alpha_0}{n_{eff}} \times \frac{1}{kT} \frac{1}{4} \frac{\hbar^2 v_F^2}{\hbar\omega} \approx 9.7\times10^{-17} cm^2 \frac{1}{n_{eff} kT} \tag{31}$$

In accordance with the previous section, in order to account for the fact that the absorption does not change linearly with the Fermi level, and also for the additional broadening, the effective broadening is introduced as $\hbar\gamma_{eff} = \sqrt{(\hbar\gamma)^2 + (3kT)^2}$, and

$$\sigma_{gr}(\omega) = \frac{\pi^2 \alpha_0}{n_a} \times \frac{1}{\hbar\gamma_{eff}} \frac{1}{4} \frac{\hbar^2 v_F^2}{\hbar\omega} \approx 9.7\times10^{-17} cm^2 \frac{1}{\hbar\gamma_{eff}} \tag{32}$$

Comparing (32) with (28) we see that for a bandgap of ~0.8 eV the effective mass is about $m_c \approx 0.05 m_0$. If we now assume that semiconductor has only one valence band and it is a light hole so that effective mass ratio parameter $\beta_v = 1$, the results are strikingly similar for graphene and QWs which is easy to understand, since graphene and many typical III-V semiconductors have roughly similar matrix transition elements, and for a given transition energy have similar joint densities of states.

3.5 Excitons in 2D semiconductors

We now turn our attention to another prospective medium for electro-absorption modulators, which has recently seen renewed interest – two dimensional (2D) excitons (Fig. 2(c)).

Optics EXPRESS

Excitons in the 2D semiconductor QWs were considered back in 1980's, mostly in the context of nonlinear optical devices [12]. But the binding energies in the III-V semiconductor QWs are comparable to room temperature thermal energy, hence excitons easily dissociate and have a broad absorption spectrum, which in the end made QW devices based on excitonic absorption impractical. Yet, in the last decade the interest has shifted to the 2D TMDs where the low dimensionality of the material increases the binding energy, mostly due to reduced screening, as compared to III-Vs. Thus, TMDs show increased absorption over QWs, and it was further suggested that these new 'robust' excitons can be used for efficient light modulation [13]. However a higher 'robustness' of the exciton should make it more difficult to change its absorption by any means, be that states saturation or screening. The effective cross section for the exciton obtained in Appendix G is

$$\sigma_{ex}(\omega) = 2\pi\alpha_0 \frac{\hbar}{m_c\gamma_{ex}} \approx 3.53\times10^{-17}\,cm^2 \frac{1}{\hbar\gamma_{ex}}\left(\frac{m_0}{m_c}-1\right) \tag{33}$$

which is comparable to the QDs. Since the effective mass in TMDs is typically larger than in III-V semiconductors, it appears that using excitons does not change the fundamental fact that each time a single electron is injected inside the active layer a single transition is being blocked.

In this section we found comparable oscillator strengths for each allowed transitions independent of switching material used. We therefore expect the change of absorption with injected charge to be somewhat material independent, although the quality of material changes the degree of broadening and thus EAM performance.

4. Comparison of modulator characteristics for different materials

All differential absorption cross-sections for the four material classes are summarized in Table 1. We find, that for all materials operating with Pauli blocking (absorption saturation) show comparable values ($\sigma \sim 10^{-14}cm^2$ at room temperature), whilst the free carriers offer a worse performance due to non-resonant character of absorption, since according to (18) for free carriers $\gamma_{eff} = \gamma + \omega^2/\gamma > 2\omega$. To verify these analytical results, obtained from essentially perturbative approach, we calculate the dependence of the absorption on the injected (induced by the gate) carriers n_{2D} for all the material classes. For the QWs we obtain according to (20),

$$\alpha_{QW} = \frac{\pi\alpha_0}{n_{eff}t_{eff}}\frac{F_{cv}}{1+\beta_v}N_{QW}\left(\left[\exp\left(\frac{\pi\hbar^2 n_{2D}}{N_{QW}m_c kT}\right)-1\right]e^{-\frac{E_c}{kT}}+1\right)^{-1} \tag{34}$$

which is plotted in Fig. 3(a) in units of $1/n_{eff}t_{eff}$ for N_{QW} = 1 and 3, $m_c = 0.06m_0$, $E_c \approx \hbar\omega - E_g = 50meV$, $F_{cv} = 0.8$, $T = 300K$, $\beta_v = 0.4$. Similarly for graphene,

$$\alpha_{gr} = \frac{\pi\alpha_0}{n_{eff}t_{eff}}\frac{1}{e^{\frac{\hbar v_F\sqrt{\pi n_{2D}}-\hbar\omega/2}{kT}}+1} \tag{35}$$

For QDs according to (15), we estimate

$$\alpha_{QD} = \frac{\pi\alpha_0}{n_{eff}t_{eff}}F_{cv}\frac{\hbar^2 N_{QD}}{m_0 kT}\frac{kT}{\hbar\gamma_{eff}}\left(\frac{m_0}{m_c}-1\right)\left(1-\frac{n_{2D}}{N_{QD}}\right) \tag{36}$$

Optics EXPRESS

The results are shown in Fig. 3(a) for the same effective mass and overlap as QWs, $\hbar\gamma = kT$, $N_{QD} = 10^{12}\,cm^{-2}$. For the excitons,

$$\alpha_{ex}(\omega) = \frac{8\alpha_0}{n_{eff}t_{eff}}\frac{\hbar^2 a_{ex}^{-2}}{m_0 kT}\frac{kT}{\hbar\gamma_{eff}}\frac{m_0}{m_c}\left(1-\pi a_{ex}^2 n_{2D}/4\right) \tag{37}$$

Table 1. Summary of the differential absorption cross sections for different material classes. For parameter definitions refer to the main text.

Material Absorption Cross-section	Expression	Approximate result
QD	$\sigma_{QD}(\omega) = \frac{\pi\alpha_0}{n_{eff}}F_{cv}\frac{\hbar^2/m_0}{\hbar\gamma}\left(\frac{m_0}{m_c}-1\right)$	$1.8\times10^{-17}\,cm^2\,\frac{F_{cv}}{\hbar\gamma n_{eff}}\left(\frac{m_0}{m_c}-1\right)$
QW	$\sigma_{QW}(\omega) = \frac{F_{cv}}{1+\beta_v}\pi^2\alpha_0\frac{\hbar^2}{m_c\hbar\gamma_{eff}}$	$5.4\times10^{-17}\,cm^2\,\frac{1}{\hbar\gamma_{eff}}\frac{m_0}{m_c}\frac{F_{cv}}{1+\beta_v}$
Graphene	$\sigma_{gr}(\omega) = \pi^2\alpha_0\times\frac{1}{\hbar\gamma_{eff}}\frac{1}{4}\frac{\hbar^2 v_F^2}{\hbar\omega}$	$9.7\times10^{-17}\,cm^2\,\frac{1}{\hbar\gamma_{eff}}$
WSe_2	$\sigma_{ex}(\omega) = 2\pi\alpha_0\frac{\hbar}{m_c\gamma_{ex}}$	$3.5\times10^{-17}\,cm^2\,\frac{1}{\hbar\gamma_{ex}}\frac{m_0}{m_c}$
Free Carriers (	$\sigma_{fc,\max}(\omega) = 4\pi\alpha_0\frac{\hbar}{m_c\gamma_{eff}}$	$7.1\times10^{-17}\,cm^2\,\frac{1}{\hbar\gamma_{eff}}\frac{m_0}{m_c}$

The results are shown in Fig. 3(a) using exciton radius, $a_{ex} = 2nm$ and the conduction effective mass, $m_c = 0.2m_0$. For the free carrier modulators, we use

$$\alpha_{fc}(\omega) = \frac{4\pi\alpha_0}{n_{eff}t_{eff}}\frac{\hbar}{m_c}\frac{\gamma}{\omega^2+\gamma^2}n_{2D} \tag{38}$$

$\gamma = 1/\tau$ is the carrier scattering rate i.e. collision frequency, electron mobility μ and τ are related by $\mu = |q|\tau/m_c$. The conductivity effective mass, m_c is taken as $0.26m_0$ [14]. μ is taken as 1100 $cm^2V^{-1}s^{-1}$ at 10^{16} cm^{-3} carrier concentration level (i.e. electrons for Silicon) [15]. Unlike the doped silicon, ITO whose chemical composition is usually given as *In_2O_3:SnO_2* can be considered an alloy as concentration of tin relative to indium can be as high as 10%. Several previous studies have calculated the permittivity of ITO using the experimentally measured reflectance and transmittance, and we choose a fitting result of Michelotti et. al whereas γ depends on the deposition conditions, in our analysis we have taken $\gamma = 1.8\times10^{14}$ rad s^{-1} [16–20].

Looking at the results (Fig. 3(a)), where the normalized absorption $\alpha' = \alpha n_{eff}t_{eff}$ is shown, we notice that while the absolute values of absorption are different for all Pauli-blocking schemes, the slopes of the curves are rather similar, as is expected from the similarity of the differential absorption cross-sections in Table 1. Since it is the slope of absorption vs. density characteristics that is important for the modulator performance (i.e. slope steepness ER/V_{bias}), one shall expect the performance of all Pauli-blocking modulating schemes to be in the same range. Next, we find the length required to achieve ~10 dB absorption in units of t_{eff} i.e. $L' = L/t_{eff}$ according to (10) (Fig. 3(d)). Furthermore, we evaluate the change of total absorption $\alpha L = \alpha' L'/n_{eff}$ as a function of the injected charge per unit waveguide cross-

Optics EXPRESS

section $Q' = Q / S'_{eff} = en_{2D} L' t_{eff} W / S'_{eff} = en_{2D} L'$ (Fig. 3(b)). For graphene, we only show the AC charge $Q' = e(n_{2D} - n_{2D,0}) L'$ where $n_{2D,0} = 0.9 \times 10^{13} cm^{-2}$ is the electron density that brings the Fermi level within *3kT* from the 0.4eV. From Fig. 3(b), we can determine the switching charge necessary to obtain 10 dB on-off ratio – the values of switching charges corresponding to the material classes for 10 dB modulation is shown in Fig. 3(e). According to our simple perturbative estimation made in Section 3, all material cases (with the exception of free carrier schemes) are expected to lie within the range of $Q' \sim 2.2e / \sigma \sim 10^{-13} - 10^{-12} C / \mu m^2$, which appears accurate. The difference between graphene, QWs, QDs and TMDs is not significant considering that the exact amount of broadening is unknown. Thus, the required switching charge and voltage depends on the effective cross-section, and the only distinction between the various material classes is how easily they may be integrated into a small mode area waveguide.

Next, we calculate the capacitance per μm^2 of the effective cross-section, $C'_g = C_g / S'_{eff} = \varepsilon_0 \varepsilon_{eff} L' / d_{gate}$ (Fig. 3(f)). This allows us to obtain the drive voltage via $V_d = Q / C_g = Q' / C'_g$ so that the absorption vs. drive voltage (Fig. 3(c)); and from there switching voltage can be obtained (Fig. 3(j)). The latter does not depend on the waveguide geometry, but only on the gate insulator thickness defining the electrostatics, similar to a transistor. In terms of switching voltage, the QD modulator appears superior, but at the expense of a larger capacitance caused by the low density of carriers in QDs. Now, one can finally determine the switching energy-per-bit per unit effective waveguide area, as $U'_{sw} = U_{sw} / S'_{eff} = \frac{1}{2} Q'_{sw} V_{sw} = \frac{1}{2} Q'^2_{sw} / C'_g$ (Fig. 3(g)).

Further, we calculate the 3dB cut off frequency $f_{3-dB} = 1 / 2\pi R C_g$ assuming $R = 50\Omega$ (Fig. 3(h)). Such a low resistance may not be realistic for photonic bulk modes, where partial and selective doping has to be used in order to keep both the carrier density and hence optical loss low. In contrast, the metal deployed in plasmonics can also serve as a low resistance contact [21]. As such, the contact resistance may vary by one order of magnitude between photonic bulk designs, and plasmonic waveguide cases while keeping minimizing the respective insertion loss in mind. The requirement of a micrometer-tight integration of optoelectronic devices has also been mentioned recently as a viable path for attojoule-per-bit efficient devices [1,21]. In this context, single layer TMDs appear competitive, because the strong excitonic absorption allows one to use a very short path length, hence low capacitance with a tradeoff in the driving voltage as it is difficult to saturate this strong absorption (Fig. 3(c)). We note, however, that modal implications, in particular, in photonic waveguides, typically result in higher contact resistances than 50Ω [22]. Plasmonics-based devices allow defining the electrical capacitor with high spatial overlap to the actual device region [21].

Optics EXPRESS

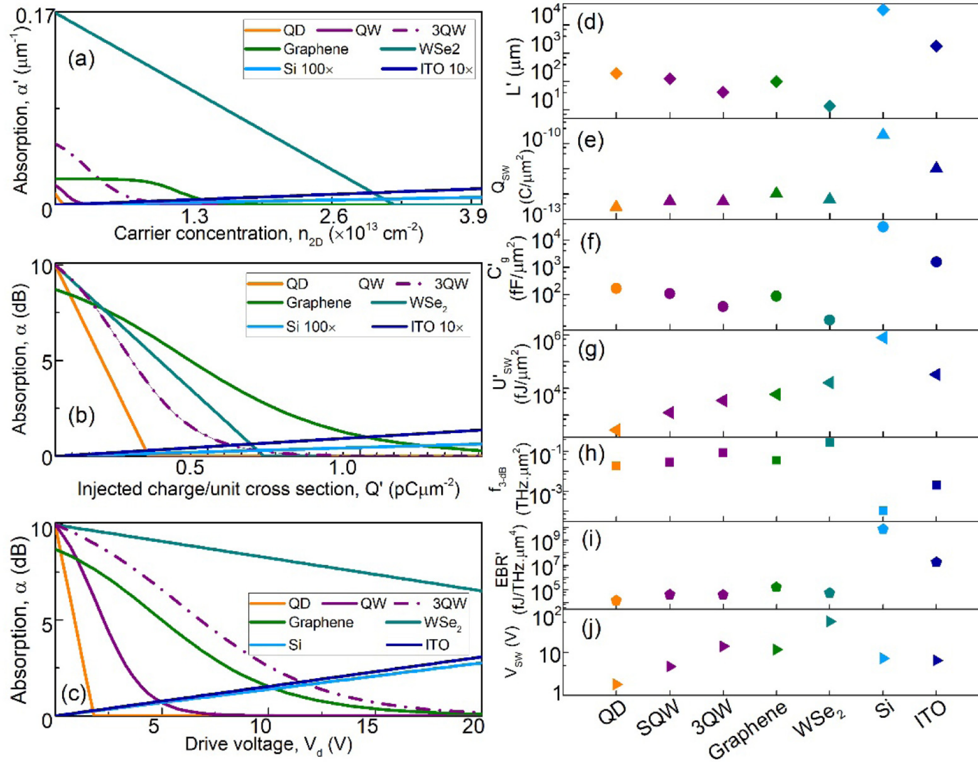

Fig. 3. Absorption modulation results for different material classes, for charge-driven active materials. (a) Normalized absorption, $\alpha' = \alpha n_{eff} t_{eff}$ as a function of carrier concentration, n_{2D} (cm^{-2}); (b-c) Optical absorption vs. (b) injected charge, Q' ($pC\mu m^{-2}$); and (c) drive voltage, V_d (Volts); respectively. (d-j) Physical device performance parameters to obtain 10dB modulation; (d) Modulator length, $L' = L/t_{eff}$; (e) Switching charge, Q_{SW} ($C/\mu m^2$); (f) Electrical device capacitance, C'_g ($fF/\mu m^2$); (g) Switching energy (energy-per-bit function), U'_{SW} ($fJ/\mu m^2$); (h) 3-dB modulation speed, $f_{3\text{-}dB}$ ($THz.\ \mu m^2$); (i) Energy-bandwidth ratio, EBR' ($fJ/(Thz.\ \mu m^4)$); and (j) Switching voltage, V_{SW} (Volts) for investigated material classes including quantum dots (QD), single and 3-layer quantum well (SQW, 3QW), Graphene, a transition metal dichalcogenide (TMD) material (WSe_2), Silicon, and Indium-Tin-Oxide (ITO). The spacing between the active layer and the gate is d_{gate} = 100 nm and ε_{eff} = 10 assumed here.

The relevant figure of merit (FOM) for modulators, is the ratio of the switching energy and cut-off frequency ('Energy Bandwidth Ratio' or *EBR*),

$$EBR = U_{SW} \,/\, f_{3dB} = \pi Q_{SW}^2 R \tag{39}$$

where evidently lower *EBR* is desired. In Fig. 3(i), we plot $EBR' = EBR / S_{eff}^{'2} = \pi Q_{SW}^{'2} R$ in units of $fJ/(THz \cdot \mu m^4)$. We find that those materials that do not rely on off-resonant absorption of free carriers, that the *EBR'* does not stray far from the same value of roughly $10^5\, fJ/(THz \cdot \mu m^4)$. This is expected since the switching charges (Fig. 3(e)) are rather close in this range. Using the switching charge in (8), we find that

$$\begin{aligned} EBR &= \pi Q_{SW}^2 R \sim \pi e^2 R S_{eff}^{'2} \times \left(\frac{m_c}{m_0}\right)^2 \left(\hbar\gamma_{eff}\right)^2 \times 2\times10^{17}\ \mu m^{-4} \\ &\approx 5\times10^4 \frac{R}{50\Omega}\frac{S_{eff}^{'2}}{\mu m^4} \times \left(\frac{m_c}{0.067 m_0}\right)^2 \left(\frac{\hbar\gamma_{eff}}{100 meV}\right)^2 fJ/THz \end{aligned} \tag{40}$$

for approximately all materials that do not rely on free carrier absorption. This result can be considered the main conclusion of all the analysis up to this point. Note, within the constraints of (40), one can increase the bandwidth by increasing the thickness or dielectric constant of the insulator while also increasing the switching voltage and energy – in the same way how one would adjust the threshold voltage and speed of the field effect transistor. Indeed, it is important to include parasitic capacitances and (40) simply presents the fundamental upper bound of the FOM. Since the effective mass in the conduction band is determined by the bandgap (larger the bandgap, the larger the effective mass roughly), the only 'intrinsic' material property affecting device performance is the effective broadening, which for most systems cannot be reduced below few kT without external cooling. This only leaves the external parameter – effective cross-section of the waveguide as the main factor for consideration, which is capable of improving the FOM. We note that the intrinsic disadvantages of 2D materials being atomically thin, may be compensated by achieving smaller cross-sections using, for instance, plasmonic modes. Overall, according to (68) one can indeed achieve a fJ-level switching energy with 1THz bandwidth if the effective waveguide cross-section can be reduced to less than 10^{-2} μm^2, and we explore this possibility in the following section.

5. Effective area for different modal structures

Here we explored a total of 14 waveguide-modes-material EAM options including photonic and plasmonic modes (Fig. 4(a-n)). Representative of the free carriers, we chose conventional Silicon (Si) and a transparent conductive oxide (TCO) emerging material, Indium Tin Oxide (ITO). We have included graphene as a unique 2D material, WSe_2 representing TMDs, and InGaAs representing conventional III-V materials. In order to understand the impact on the effective area, we diversify photonic and plasmonic waveguide mode designs to include different mode structures, such as bulk, slot [23], and hybrid-photon-plasmon (HPP) [24,25], for the chosen materials – Si, ITO, graphene and WSe_2. The Si, ITO and graphene mode structures are chosen similar to our previous work [26,27]. The WSe_2 structures are the same as graphene ones with just the active material changed to WSe_2. In principle, III-V QWs or QDs can also be placed inside the slot structure, but such designs are not considered in this work. As expected, bulk waveguide designs do not allow for small effective cross-sections. However, the actual improvement is not more than 100 × except for the Silicon slot case where the mode is highly squeezed due to the high index Si being below the slot air gap realizing the high E-field concentration. However such a miniscule effective area in this mode also presents tradeoff in terms of high insertion loss (IL) of ~27 dB. We note that the waveguide dimensions assumed in Fig. 4 are exemplary, but smaller effective cross-section are obtainable as well. Yet, a balance between insertion loss and shrinking the mode size should be considered from a photonic interconnect integration point of view [28]. Employing plasmonic structures in this regard can help the cause as plasmonic modes enable a few orders of magnitude smaller effective modal cross-section compared to photonic modes [29–36].

Optics EXPRESS

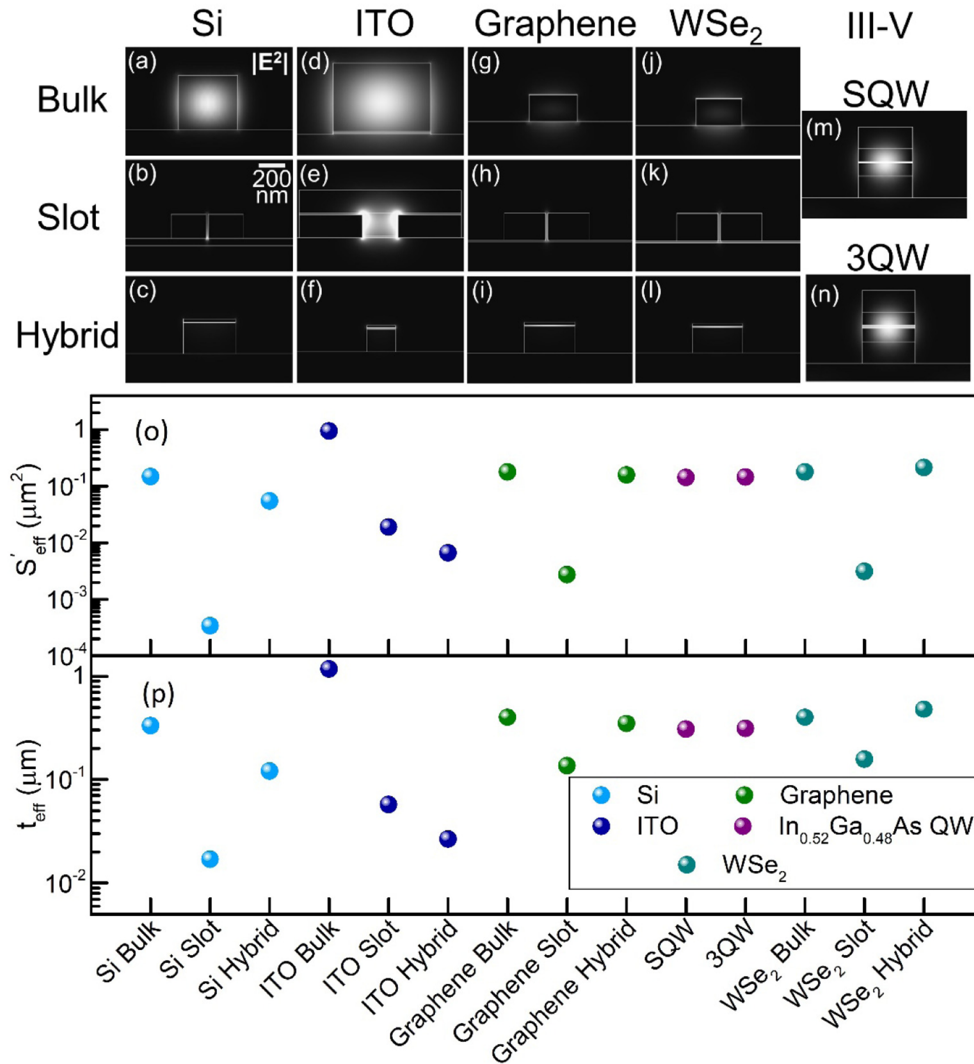

Fig. 4. **(a-n)** Cross-sectional mode profiles for different structures and material classes using FEM analyses, normalized electric field intensity, $|E|^2$ is shown. **(o)** Corresponding effective cross-sectional modal area, S'_{eff}; and **(p)** effective thickness, t_{eff}. The Si, ITO and graphene structures are chosen form our previous work [26,27]. WSe_2 structures are the same in dimensions as graphene ones just changing the active material to WSe_2, $In_{0.52}Ga_{0.48}As$ QWs are chosen having 5 nm thickness and 470 nm width, with GaAs separate confinement heterostructure (SCH) and barrier layers.

6. Energy-per-bit function, speed and material broadening

The free carrier modulators operate far from resonance and are therefore not competitive with the other material structures. In fact, reducing the broadening will actually worsens their performance as an EAM, which is not the case for the electro-optic modulators (i.e. phase-shifting modulators). But the other transition blocking-based material systems do use resonant absorption, hence can benefit from reduced broadening as the transitions are sharper and a higher ER is possible in the same footprint. In case of QDs and TMDs the absorption linewidth contains homogeneous broadening, which is at least partially temperature-dependent (phonon scattering) and an inhomogeneous part, due to material variations (especially in QDs). In graphene and QWs, in addition to scattering, the main cause of

broadening is associated with the Fermi-function spread on the scale of *3kT*. For these materials the effective linewidth is $\gamma_{eff} = \left(\gamma_{homo}^2 + \min\{3kT \quad or \quad \gamma_{inh}\}^2\right)^{1/2}$, where γ_{homo} is the material or homogeneous broadening and γ_{inh} is the inhomogeneous broadening. All the EAM performance parameters with respect to the effective broadening, effective thickness and effective cross-sectional area are calculated (Fig. 5).

Next, we discuss different material classes to demonstrate the effect of effective material broadening and effective modal area variations on EAM performance (Fig. 5). Pauli blocking based graphene, band filling based QW modulators, narrow resonant QDs, and excitonic modulation based WSe_2 are chosen to investigate effects of changing the effective broadening. Different free carrier based materials, namely Si and ITO, are also chosen to signify variation of effective material broadening in free carrier based schemes. Our results show monotonic modulator improvements with both reduced broadening and reduced effective modal area. The latter points to polaritonic modes where the field is squeezed into small (i.e. sub-diffraction limited) effective modal areas. Regarding broadening, almost 1-2 orders of magnitude improvements can be attained by operating at cryogenic temperatures, i.e. $\gamma_{eff} << 0.1$eV. We used a nominal 77 K for our calculations corresponding to cryogenic temperature. However, beyond a broadening corresponding to room temperature, the effect declines. To study the dependence on the effective broadening, γ_{eff} for QWs and graphene, we change those device parameters that rely on operating temperature via a scaling factor, $G = T/300\ K$. As such, the switching charge, Q_{SW} scales as G, while the capacitance, C, stays constant, and so does the 3-dB modulation bandwidth. The modulator length, L is unchanged from the variation in broadening; the switching voltage, V_{SW} scales as G; switching energy, i.e. the energy-per-bit function, U_{SW} and energy-bandwidth ratio, *EBR* both scale as G^2. The situation is different, however, for resonant narrow line emitters like QDs and WSe_2. Since we used $\gamma = kT$ in above discussion, we introduce the scaling factor, $G = \gamma / kT$ to observe variations in changing γ_{eff}. Incidentally, a broader γ_{eff} means a longer device length needed to absorb. Therefore, switching charge Q_{SW}, capacitance C, modulator length L scales all with G; the switching voltage, V_{SW} is constant ($\sim Q_{SW}/C$); switching energy, U_{SW} scales as G; and finally, energy-bandwidth ratio, *EBR* scales as G^2.

As before, we select a resistance, $R = 50\Omega$ and subsequently calculate the capacitance, C using $C = \varepsilon_0 \varepsilon_{eff} WL / d_{gate} = \varepsilon_0 \varepsilon_{eff} W t_{eff} L / d_{gate} t_{eff} = (\varepsilon_0 \varepsilon_r / d_{gate}) S_{eff}^{'} L'$. However, we note that photonic-modes are sensitive to insertion losses, and require selective doping of the contact regions in order to reduce the resistance, setting a trade-off between added losses from doping versus lowering the contact resistance. This bottleneck is circumvented in plasmonic modulators, where the metal can be synergistically used as a contact ensuring 10's of Ω low resistance. Our results indicate footprint reduction (i.e. scaling) of about one order of magnitude for 10 dB modulation in QDs and TMDs upon thermal cooling. Similar improvements can be noticed in terms of lowering the electrical capacitance and thence, increasing the 3-dB modulation bandwidth as well for QDs and TMDs, for reduced effective broadening. Also thermal cooling lowers the switching charge by about one order of magnitude for all the material classes considered, i.e. QWs, QDs, WSe_2 and graphene. The energy– per–bit function, or switching energy, U_{SW}, can be further lowered by ~10 × – 15 × by improving the effective material broadening. Resonant narrow line emitters, i.e. QDs and WSe_2 in this work, observe about one order of magnitude reduction in switching energy; whereas Pauli blocking enabled graphene and band filling based modulators using QWs results in about 15 × lower energy-per-

Optics EXPRESS

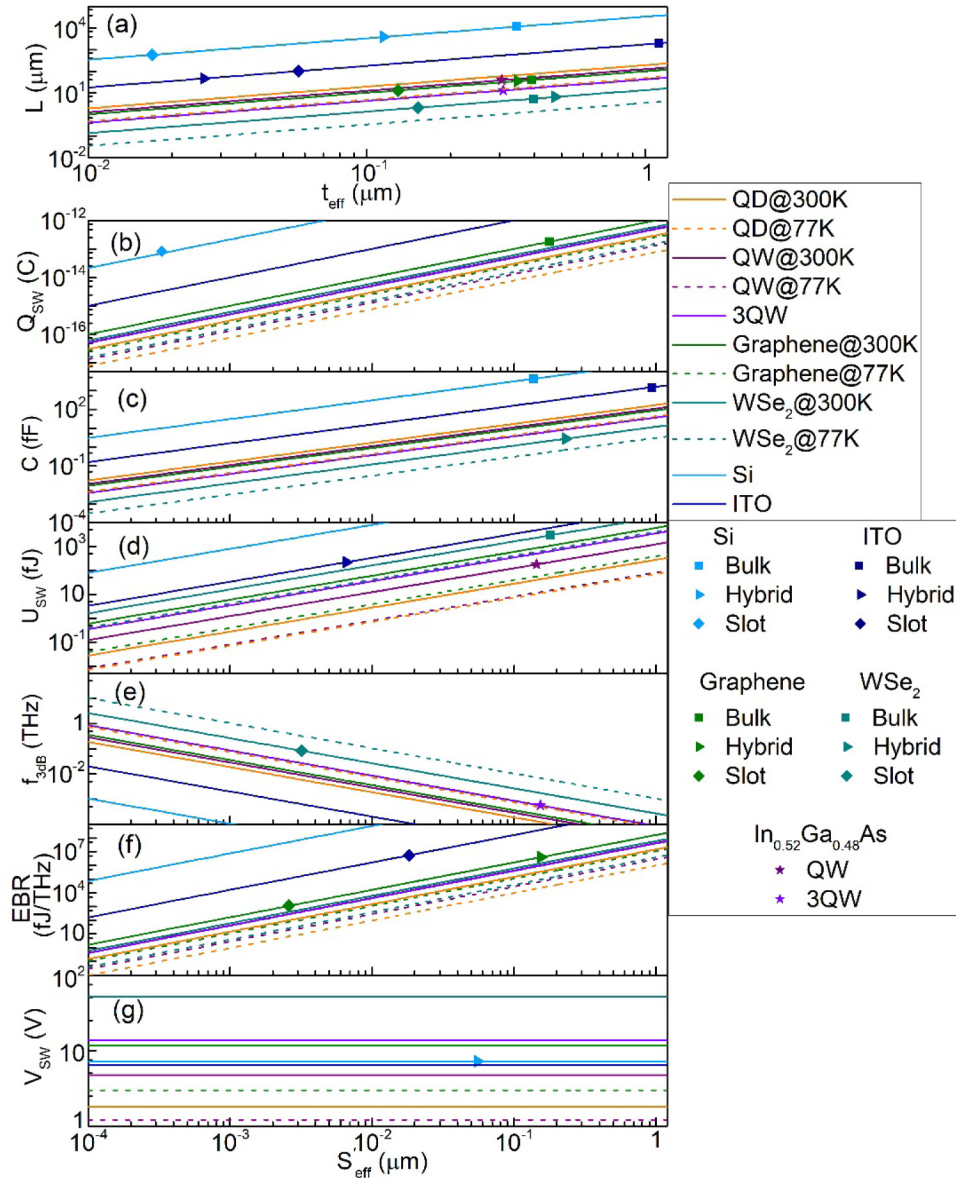

Fig. 5. **(a)** 10 dB absorption modulator length, L vs. effective thickness, t_{eff} for all the comparable material classes with different broadening. **(b-g)** Relevant device parameters vs. effective mode areas, S'_{eff} for different effective material broadening, γ_{eff} (eV) for all the comparable material classes; **(b)** Switching charge, Q_{SW} (C); **(c)** Capacitance, C (fF); **(d)** Switching energy (E/bit function), U_{SW} (fJ); **(e)** 3-dB modulation bandwidth, $f_{3\text{-}dB}$ (THz); **(f)** Energy-bandwidth ratio, EBR (fJ/THz); and **(g)** Switching voltage, V_{SW} (V). The varying amount of broadening is implicit by the materials shown in the legend corresponding to room temperature, i.e. 300 K, and thermal cooling down to 77 K. The corresponding t_{eff} and S'_{eff} for the different modes from the previous section are also marked to aid comparing device performances. d_{gate} = 100 nm. Lowering the gate oxide thickness will increase the energy efficiency through improved electrostatics, but also reduce the modulation speed via increased capacitance. All the different modes considered are marked in (a), whereas the different mode performances are scattered across in (b-g) as the markings correspond to the same x-value in all of them. Best EBR is found for graphene and TMD plasmonic slot waveguide-based modulators.

Optics EXPRESS

-bit for cryogenic operation. The switching voltage also can be decreased in graphene and QWs by employing cryogenic cooling by almost an order of magnitude. As some of the materials enable higher modulation speeds and some lower energy-per-bit functions upon reduction in effective broadening, the intended purpose and applications should be kept in mind when designing modulators and choosing the active material. Our deterministic energy-bandwidth ratio (*EBR*), in a way, helps to level the field as almost one order of magnitude improvement can be noticed for all of the material classes considered for variations in effective broadening.

As the 3-dB speed and energy-per-bit behave inversely with the effective modal area, our deterministic figure of merit, *EBR* decreases by several orders of magnitude; thus selecting a plasmonic slot waveguide with highest optical confinement can achieve high performing modulators withstanding higher IL. Exemplary of attainable *EBR* values from experimentally manifested devices include the notable graphene ring modulator, exhibiting an *EBR* of ~2.7×10^4 fJ/THz with 30 GHz bandwidth and ~800 fJ of switching energy [37]; all metallic plasmonic high-speed modulator attaining an *EBR* of 1.6×10^3 fJ/THz experimentally demonstrated with projected values down to ~500 fJ/THz [38]; and an all plasmonic modulator with $\chi^{(2)}$ polymer reported values of ~350 fJ/THz [39]. We note that, the performance metrics reported in our analysis are for linear devices and can further be enhanced by factors of a few (~1 order of magnitude) employing cavity feedback with low-to-medium (100-1000) quality factor cavities that do not suppress the signal completely inside the cavity. Also, the gate oxide is somewhat large here (d_{gate} = 100 nm), which worsens the energy per bit, but allows for high-speed devices via a low capacitance. For realistic switching speeds in the 10's to maybe in the future 100's GHz range, current modulators should be designed with a thin oxide to improve the energy consumption.

Highly confined modes are inherently accompanied by higher IL and tradeoffs must be made for performance benefits. For example, the slot graphene mode exhibits ~27 dB IL, and a larger (~57% higher t_{eff}) similar slot mode exhibits IL of ~18 dB trading in footprint (~61% larger L). WSe_2 slot and hybrid modes exhibit IL of ~3 and ~14 dB, respectively. Note, WSe_2 shows lower loss because the waveguide is shorter due to the maximum absorption being high. The required switching voltage is also higher. Furthermore, note that if multiple (say 5) monolayers of graphene is used instead of 1, the *EBR* and Q_{SW} will remain same, but L will be reduced by 5 and IL will also be reduced by a factor of 5. Of course, the switching voltage will increase but the bandwidth is wider as well.

The intricate interplay between *EBR* and IL will be the subject of our future work, but a certain conclusion can be made here; in general, for all the plasmon-assisted waveguides IL is the dominant consideration, hence the path to increasing *EBR* lies in reducing the length by increasing the absorption, using excitonic absorption in TMDs, multiple graphene layers, multiple QWs or high concentration of QDs. Unfortunately, that would also lead to a higher switching voltage and energy combined with the speeds that are too high to be practical (after all, what is the advantage of 2THz modulator in the absence of the electronic devices capable of driving it at that speed?). Therefore, given a fixed value of *EBR* for a given effective area of waveguide it is desirable to reduce the switching energy (and voltage) while also reducing the speed to whatever is necessary (say 100 GHz) for a particular application, which can be achieved by increasing the length, but would also result in the increased IL. That leaves us with only one feasible approach – increasing the gate capacitance, possibly using high-K dielectrics. It is remarkable that the path to miniaturization of EAMs follows exactly the path to miniaturization of field effect transistors! Then again both EAM and FET are charge control devices, and, as we have shown in this work, EAM borrows many FET characteristics such as accumulated charge, threshold voltage and gate capacitance, hence the route to EAM optimization is expected to follow the one blazed by FET development.

Prior to concluding, we should mention that while we only analyze electro-absorption modulators here, the main conclusions are also relevant to the electro-optic modulators; from

the Kramers-Kronig relations between the real and imaginary parts of the refractive index it follows that the change in the refractive index is related to the change in absorption coefficient by about $\Delta n_{eff}(\omega) \sim (\gamma_{eff}/\Delta\omega)\Delta\alpha\lambda/4\pi$, where $\Delta\omega$ is detuning from the frequency of maximum absorption change. Therefore the switching requirement for an electro-optic modulator operating in push-pull regime $\Delta n_{eff} L = \lambda/4$ amounts to $\Delta\alpha L \sim \pi\Delta\omega/\gamma_{eff}$, which is the same as the EAM switching condition in (13) scaled by $\Delta\omega/\gamma_{eff}$. In order to avoid spurious absorption modulation one must maintain large detuning $\Delta\omega >> \gamma_{eff}$ which explains why the electro-optic modulators have larger footprints than EAMs. However, other than this fact, all the relevant characteristics of EOMs follow the same dependences on the material differential absorption cross-section and waveguide effective area. We shall address performance metrics of EOMs in greater detail in our upcoming work.

7. Conclusions

We have carried out a holistic physical device analysis for waveguide-based electro-absorption modulators. In our analysis we have evaluated various performance characteristics, such as switching charge, voltage, energy-per-bit, and bandwidth, and also introduced the performance metric – the ratio of switching energy to the bandwidth. We have shown that other than insertion loss, the modulator performance is limited by just one fundamental material characteristic – differential absorption cross-section and one fundamental waveguide characteristic – its effective modal cross-section. We analyzed a variety of actively researched switching materials to include two-level absorbers such as quantum dots, free carrier accumulation or depletion modulators such as based on ITO or Silicon, Pauli blocking in graphene and III-V semiconductor quantum wells, and excitons in two-dimensional (2D) atomic layered materials found in transition metal dichalcogendies. Apart from free carrier modulators, all material classes have essentially the same absorption cross section, on the scale of 10^{-14} cm^2; as the scale of the momentum matrix element scales inversely with the chemical bond length which does not change much from one material to another for the allowed transitions (this fact can be summarized as the oscillator sum rule). The only way to increase the absorption cross-section is to minimize broadening which is challenging as quantum dots typically exhibit broadening in excess of 100 meV, and in graphene and quantum wells, thermal broadening alone is about 75 meV at room temperature. Reduction of exciton broadening in 2D materials may bring some benefits but not by more than a factor of ~5. Beyond broadening, we considered waveguide designs with smaller effective cross-sections, and find that the effective modal cross-section is always accompanied by the increased insertion loss and hence increased power dissipation that may nullify the benefits of reduced switching energy. However, important is that, in principle, up to a few fJ switching energies (thin waveguide and larger gate oxide) are attainable at 100's of Gbps speed with the known technologies, provided that the numerous fabrication-related issues, such as excessive insertion loss, poor coupling, parasitic capacitance and contact resistance, are solved. Taken together, the highest performing electro-absorptive modulators can be built by choosing multiple graphene layers or multiple QWs or a 2D material and interface it with a plasmonic slot waveguide incorporating high-K dielectric. With this, we hope this work will provide guidance to the community in developing next-generation high-performance modulators incorporating emerging materials.

Appendix A: Power flow in the waveguide

To determine the effective area of the waveguide we consider electromagnetic wave propagating along the direction z with the propagation constant $\mathrm{E} = \mathrm{E}(x,y)e^{i(\beta z-\omega t)} + c.c.$ The total power flow is then $P = n_{eff} E_{a0}^2 S_{eff} / 2\eta_0$ where E_{a0} is the magnitude of the transverse electric

Optics EXPRESS

field in the middle of active layer (Fig. 1) and $\eta_0 = 377\Omega$ is the impedance of free space. The effective index has been introduced as $n_{eff} = \beta c / \omega$. To find the effective crosss-section, we fist determine the longitudinal component of the Poynting vector $S = E \times H$ which can be found as $S_z(x,y) = E_y H_x - E_x H_y$. According to Maxwell equations $\frac{\partial}{\partial z} H_x = \beta H_x = \omega n^2(x)\varepsilon_0 E_x$ and $\frac{\partial}{\partial z} H_y = \beta H_x = -\omega n^2(x)\varepsilon_0 E_x$. Thus, the time-averaged Poynting vector magnitude is $\bar{S}_z(x,y) = \frac{\omega n^2(x,y)\varepsilon_0 (E_y^2 + E_x^2)}{2\beta} = \frac{n^2(x,y)(E_y^2 + E_x^2)}{2 n_{eff} \eta_0}$,. The total power flow is then

$$P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \bar{S}_z(x,y)dxdy = \frac{1}{2 n_{eff}\eta_0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n^2(x,y)\left(E_x^2 + E_y^2\right)dxdy = \frac{n_{eff} E_{a0}^2}{2\eta_0} S_{eff} \tag{41}$$

where E_{a0} is the magnitude of the transverse electric field in the middle of active layer (Fig. 1) from which Eq. (1) from the main text follows.

Appendix B: Effective area of active region and absorption cross-section

Let us now estimate the absorption change induced by the injection (or depletion of carriers). We first consider two level absorbers, which can be atoms or quantum dots (QDs) having resonant frequency ω_0 and density $N_a(x,y)$. The absorbers have Lorentzian line shape with linewidth (broadening) γ, characterized by the broadening function

$$g(\hbar\omega) = \frac{1}{\pi\hbar}\frac{\gamma}{\Delta\omega^2 + \gamma^2} \tag{42}$$

where $\Delta\omega = \omega - \omega_0$ is the detuning from the resonance. According to Fermi's Golden rule the density of photons changes as

$$\frac{dN_{ph}(x,y)}{dt} = \frac{2\pi}{\hbar}\frac{e^2}{4} E_{eff}^2 r_{12}^2 g(\hbar\omega) N_a(x,y) \tag{43}$$

where r_{12} is the transition matrix element of dipole, E_{eff} as the component of the electric field that is actually being absorbed (for many 2D structures it is the in-plane or TE component), and the factor of ¼ comes from the fact that only positive frequencies cause the transition. Then, the change of the energy density due to absorption in the active layer becomes

$$\frac{dU(x,y)}{dt} = \hbar\omega \frac{dN_{ph}(x,y)}{dt} = e^2 r_{12}^2 \frac{E_{eff}^2(x,y)}{2}\frac{\omega}{\hbar\gamma}\frac{\gamma^2}{\Delta\omega^2 + \gamma^2} N_a(x,y) \tag{44}$$

Integrating over the cross-section we obtain

$$\frac{dP}{dz} = -\int_{active}\int_{-W/2}^{W/2}\frac{dU}{dt}dxdy = \frac{-e^2 r_{12}^2}{2}\frac{\omega}{\hbar\gamma}\frac{\gamma^2}{\Delta\omega^2 + \gamma^2}\int_{active}\int_{-W/2}^{W/2} N_a(x,y) E_{eff}^2(x,y)dxdy \tag{45}$$

We now divide and multiply (45) by $E_{a0}^2 N_{a0}$ where N_{a0} is the density of absorbing entities in the middle of the active region and use $P = n_{eff} E_{a0}^2 S_{eff} / 2\eta_0$ to obtain

Optics EXPRESS

$$\frac{dP}{dz}=\frac{-e^2 r_{12}^2}{2}\frac{\omega}{\hbar\gamma}\frac{\gamma^2}{\Delta\omega^2+\gamma^2}\frac{\int\limits_{active}\int\limits_{-W/2}^{W/2} N_a(x,y)E_{eff}^2(x,y)dxdy}{N_{a0}E_{a0}^2}N_{a0}2\eta_0 P/S_{eff}n_{eff} \quad (46)$$

Now, we can introduce the effective area of the active region as

$$S_{a,eff}=\int\limits_{active}\int\limits_{-W/2}^{W/2} N_a E_{eff}^2 dxdy / E_{a0}^2 N_{a0} \quad (47)$$

where N_{a0} is the density of the absorbers in the middle of the active region. We have defined E_{eff} as the component of the electric field that is actually being absorbed (for many 2D structures it is the in-plane or TE component). The power flow over the cross section becomes

$$\frac{dP}{dz}=-\frac{e^2 r_{12}^2\eta_0}{n_{eff}}\frac{\omega}{\hbar\gamma}\frac{\gamma^2}{\Delta\omega^2+\gamma^2}\frac{S_{a,eff}}{S_{eff}}N_{a0}P=-\sigma_a(\omega)\frac{S_{a,eff}}{S_{eff}}N_{a0}P \quad (48)$$

where r_{12} is the matrix element of the dipole transition between two levels and the absorption cross-section has been introduced as

$$\sigma_a(\omega)=\frac{e^2\eta_0}{\hbar n_{eff}}r_{12}^2\frac{\omega}{\gamma}\frac{\gamma^2}{\Delta\omega^2+\gamma^2}=\frac{4\pi\alpha_0}{n_{eff}}r_{12}^2\frac{\omega}{\gamma}L(\omega) \quad (49)$$

Here, α_0 is the fine structure constant and $L(\omega)=\gamma^2/(\Delta\omega^2+\gamma^2)$ represents a Lorentzian line shape that has maximum value at resonance equal to unity. The absorption coefficient is therefore

$$\alpha(\omega)=\sigma_a(\omega)\frac{S_{a,eff}}{S_{eff}}N_{a0} \quad (50)$$

The significance of this expression is that the absorption is clearly separated into two factors – intrinsic (or material) the absorption cross-section, and the geometrical (or confinement) factor, $\Gamma = S_{a,eff}/S_{eff}$.

Now, often the active layer is essentially two-dimensional, with a two-dimensional density of carriers, which is independent on the lateral direction, i.e. $N_{2D}=\int_{active}N_a(x)dx$. Then we can find effective active area (47) as

$$S_{a,eff}=\frac{N_{2D}}{N_{a0}}W\mathcal{F} \quad (51)$$

where the 'uniformity function' is

$$\mathcal{F}=\int\limits_{active}\frac{N_a(x)}{N_{a0}}\int\limits_{-W/2}^{W/2}\frac{E_{eff}^2(x,y)}{E_{a0}^2}dydx\Big/W\int\limits_{active}\frac{N_a(x)}{N_{a0}}dx \quad (52)$$

Clearly, when the active layer is essentially a delta function, $N_a(x)=N_{2D}\delta(x-x_a)$

$$\mathcal{F}=W^{-1}\int\limits_{-W/2}^{W/2}\frac{E_{eff}^2(x_a,y)}{E_{a0}^2}dy \quad (53)$$

Optics EXPRESS

and if the field is uniform in the lateral direction, then $F = E_{eff}^{2} / E_{a}^{2} = \cos^{2}\theta_{eff}$, where θ_{eff} is the angle between the dipole and electric field. Now we can write (50) as

$$\alpha(\omega) = \sigma_a(\omega) \frac{WF}{S_{eff}} N_{2D} = \frac{\sigma_a(\omega)}{t_{eff}} N_{2D} \tag{54}$$

The effective thickness of the waveguide has been introduced as

$$t_{eff} = S_{eff} / WF = S_{eff}^{'} / W \tag{55}$$

where $S_{eff}^{'} = S_{eff} / F$. For the waveguides in which the field does not change much laterally, which is often the case,

$$t_{eff} \approx \frac{1}{n_{eff}^{2}} \int_{-\infty}^{\infty} n^2(x)\left(E_x^2 + E_y^2\right) dx / E_{a0}^{2} \tag{56}$$

Appendix C: Evaluating absorption cross-section for two-level absorbers

Here we evaluate the absorption cross-section (49) for the two-level absorbers (QDs) using the relation between the dipole matrix element, r_{12} and the matrix element of the momentum for the interband transition, P_{cv} we get [40,41]

$$r_{12} = \sqrt{\frac{F_{cv}}{2}} \frac{P_{cv}}{m_0 \omega} \tag{57}$$

where the first term includes the overlap between the 'envelope' wave-functions of the conduction and valence bands $F_{cv} = \left|\int \psi_c \psi_v dV\right|^2 <\sim 1$ and the factor of ½ is associated with specifics of the heavy hole wave-function involved in the transition. Furthermore, one can use the relation between the above momentum matrix element and the electron effective mass, m_c as [42]

$$\frac{m_0}{m_c} = 1 + \frac{2P_{cv}^2}{m_0 E_g} \approx \frac{E_p}{E_g} \tag{58}$$

where for a wide range of direct bandgap semiconductors $E_p = 2P_{cv}^2 / m_0 \approx 16 - 24eV$ and $E_g \approx \hbar\omega$ is the bandgap energy. Then we obtain the expression for the absorption cross-section on resonance

$$\sigma_{QD}(\omega) = \frac{\pi\alpha_0}{n_{eff}} F_{cv} \frac{\hbar}{\gamma}\left(m_c^{-1} - m_0^{-1}\right) = \frac{\pi\alpha_0}{n_{eff}} F_{cv} \frac{\hbar^2 / m_0}{\hbar\gamma}\left(\frac{m_0}{m_c} - 1\right) = 1.75\times10^{-17} cm^2 \frac{F_{cv}}{\hbar\gamma n_{eff}}\left(\frac{m_0}{m_c} - 1\right) \tag{59}$$

where the energy broadening $\hbar\gamma$ is given in electron volts (eV). This is Eq. (15) from the main text.

Appendix D: Free carrier absorption cross-section

Here we derive the expression for the absorption cross-section of free electrons. Dielectric constant of free electron plasma in Drude approximation can be written as

$$\varepsilon_r(\omega, x, y) = \varepsilon_\infty - \frac{N_e(x,y)e^2}{\varepsilon_0 m_c} \frac{1}{\omega^2 + i\omega\gamma} \tag{60}$$

where ε_∞ is the dielectric constant of the undoped semiconductor. The rate of absorption of electro-magnetic energy is proportional to the imaginary part of the dielectric constant, $\varepsilon_{r,im}$ as

$$\frac{dU(x,y)}{dt}=\omega\varepsilon_0\varepsilon_{im}(\omega,x)\frac{E_{eff}^2(x)}{2}=\frac{N_e(x,y)e^2}{m_c}\frac{\gamma}{\omega^2+\gamma^2}\frac{E^2(x,y)}{2} \tag{61}$$

where effective electric field, E_{eff} is the same as the total field, E. This expression is similar to (44), thence we follow the previous steps leading to (50) with the effective absorption area described by (47) with N_e in place of N_a, and absorption cross-section now can be defined as

$$\sigma_{fc}(\omega)=\frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar}{m_c}\frac{\gamma}{\omega^2+\gamma^2}=\frac{4\pi\alpha_0}{n_{eff}}\frac{\hbar^2}{m_0(\hbar\gamma_{eff})}\frac{m_0}{m_c}\approx\frac{7.02\times10^{-17}cm^2}{(\hbar\gamma_{eff})}\times\frac{m_0}{m_c} \tag{62}$$

where the effective detuning is $\gamma_{eff}=\gamma+\omega^2/\gamma$. This expression is Eq. (18) from the main text.

Appendix E: Absorption coefficient in semiconductor QWs

Considering N_{QW} number of quantum wells inside a waveguide, where each of the QWs is populated with the 2D carrier density n_{2D} distributed according to Fermi-Dirac distribution $f_c(\mathrm{E}_c)=\{1+\exp[(\mathrm{E}_c-\mathrm{E}_f)/kT]\}^{-1}$. The Fermi energy (relative to the bottom of conduction band) can be found as

$$\mathrm{E}_f=kT\ln\left[\exp\left(\frac{\pi\hbar^2 n_{2D}}{N_{QW}m_c kT}\right)-1\right] \tag{63}$$

which can be differentiated to obtain

$$\frac{dn_{2D}}{d\mathrm{E}_f}=\frac{N_{QW}m_c}{\pi\hbar^2}\frac{1}{1+\exp(-\mathrm{E}_f/kT)} \tag{64}$$

Let us now evaluate the absorption rate of the carriers; following Fermi's Golden rule for that we introduce the rate of absorption per unit area, analogous to (43),

$$\frac{dn_{2D}(y)}{dt}=\frac{2\pi}{\hbar}\frac{e^2}{4}E_{eff}^2(x_{QW},y)r_{cv}^2\rho_{2D}(\hbar\omega)[1-f_c(\mathrm{E}_c)] \tag{65}$$

Here, the joint density of states per unit of energy per unit of area is $\rho_{2D}(\hbar\omega)=N_{QW}\frac{m_r}{\pi\hbar^2}H(\hbar\omega-\mathrm{E}_g)$, where H is the step function, and the reduced mass m_r is found from $m_r^{-1}=m_c^{-1}+m_v^{-1}$, and the value of energy in the conduction band (CB) is $\mathrm{E}_c=(\hbar\omega-\mathrm{E}_g)m_v/m_r$. Following (57) switching from the dipole to momentum representation, we obtain

$$\frac{dn_{2D}(y)}{dt}=\frac{2\pi}{\hbar}\frac{e^2}{4}E_{eff}^2\frac{1}{2}\frac{P_{cv}^2}{m_0^2\omega^2}F_{cv}N_{QW}\frac{m_r}{\pi\hbar^2}H(\hbar\omega-\mathrm{E}_g)[1-f_c] \tag{66}$$

Next, in QWs the effective mass of the electron is determined by (57) while the in-plane effective mass of hole is typically about 2-3 times larger, then one can write for the reduced

Optics EXPRESS

mass $m_r^{-1} = (1+\beta_v)\times 2P_{cv}^2 / m_0^2\hbar\omega$ where the effective mass ratio factor $0 < \beta_v < 0.5$. Upon substituting it into (66), we obtain

$$\frac{dn_{2D}(y)}{dt} = \frac{1}{1+\beta_v}\frac{e^2}{4\hbar}\frac{E_{eff}^2}{2\hbar\omega}F_{cv}N_{QW}H(\hbar\omega - \mathrm{E}_g)[1-f_c] \quad (67)$$

Next, following (48) – (50), the expression for the absorption coefficient can be obtained as

$$\alpha_{QW} = \frac{1}{1+\beta_v}\frac{\pi\alpha_0}{n_{eff}}\frac{WFN_{QW}}{S_{eff}}F_{cv}H(\hbar\omega - \mathrm{E}_g)[1-f_c] \quad (68)$$

which is the Eq. (20) from the main text.

Appendix F: Graphene absorption cross-section

In graphene, the matrix element of momentum between the valence and conduction band is $P_{cv,gr} = m_0 v_F$, where $v_F \sim 10^8\, cm/s$ is the Fermi velocity. The matrix element of Hamiltonian in the $\mathrm{p}\cdot\mathrm{A}$ gauge then becomes $H_{cv} = \frac{1}{2}e\mathrm{v_F}\cdot\mathrm{E}/\omega$. The Joint density of states can be found next, as

$$\rho_{2D}(\hbar\omega) = \frac{1}{2\pi}\frac{\omega}{\hbar v_F^2} \quad (69)$$

where we have used the dispersion relation for the transition frequency, $\omega = 2kv_F$ and the fact that, in graphene, one deals with both spin and valley (*K*, *K'*) degeneracies. Substituting it all into Fermi's golden rule we obtain:

$$\frac{dn_{2D}(y)}{dt} = \frac{2\pi}{\hbar}\frac{e^2}{4}E_{eff}^2\frac{1}{2}\frac{v_F^2}{\omega^2}\frac{1}{2\pi}\frac{\omega}{\hbar v_F^2}[1-f_c] = \frac{e^2E_{eff}^2}{8\hbar^2\omega}[1-f_c] \quad (70)$$

This expression is rather similar to (67), and thus following all the steps for QWs that have lead to (68) we obtain the expression for the graphene interband absorption in the waveguide:

$$\alpha_{gr} = \frac{\pi\alpha_0}{n_{eff}t_{eff}}[1-f_c] \quad (71)$$

Next, we consider the changes in graphene absorption with density of electrons; since the most change in absorption occurs when the Fermi level approaches $E_f = \hbar\omega/2 >> kT$, the relation between density of electrons and Fermi level is $n_{2D} \approx \mathrm{E}_f^2/\pi\hbar^2 v_F^2$ and differentiating it, we obtain $dn_{2D}/d\mathrm{E}_f \approx 2\mathrm{E}_f/\pi\hbar^2 v_F^2$. Now, following (21)

$$\frac{d\alpha_{gr}}{dn_{2D}} = \frac{\partial\alpha_{QW}}{\partial f_c}\frac{\partial f_c}{\partial\mathrm{E}_{fc}}\frac{d\mathrm{E}_f}{dn_{2D}} = -\sigma_{gr}(\omega)t_{eff}^{-1} \quad (72)$$

where the differential absorption cross-section is

$$\sigma_{gr}(\omega) = \frac{\pi^2\alpha_0}{n_{eff}}\times\frac{1}{kT}\frac{\exp\left[(\hbar\omega/2-\mathrm{E}_{fc})/kT\right]}{\left\{1+\exp\left[(\hbar\omega/2-\mathrm{E}_{fc})/kT\right]\right\}^2}\frac{\hbar^2 v_F^2}{2\mathrm{E}_f} \quad (73)$$

This is Eq. (30) from the main text.

Optics EXPRESS

Appendix G: Effective absorption cross-section for excitons in 2D semiconductors

The 2D material exciton is characterized by its 2D Bohr radius

$$a_{ex} = \frac{2\pi\varepsilon_{eff}\varepsilon_0\hbar^2}{e^2 m_r} \tag{74}$$

which is half as large as 3D exciton radius. The effective dielectric constant ε_{eff} in 2D materials approaches unity while the effective mass $m_r \sim 0.25m_0$ is somewhat larger than in III-V semiconductors. Thus, the exciton Bohr radius in TMDs is on the order of only a few nanometers long [43], exhibiting exciton binding energy of $E_{ex} = \hbar^2 / 2m_r a_{ex}^2 \sim 0.5eV$ [44]. The absorption of the exciton can be easily obtained by using the value of the exciton envelope wave function at the origin, i.e. probability of finding electron and hole in the same spatial location $2|\Phi_{ex}(0)|^2 = 4/\pi a_{ex}^2$ in place of two 2D density of atoms or QDs in (54) to obtain

$$\alpha_{ex}(\omega) = 4\sigma_{ex}(\omega) / \pi a_{ex}^2 t'_{eff} \tag{75}$$

where according to (49) and (15), operating on the excitonic resonance

$$\sigma_{ex}(\omega) = 4\pi\alpha_0 r_{12}^2 \frac{\omega}{\gamma} \approx 2\pi\alpha_0 \frac{\hbar}{\gamma}(m_c^{-1} - m_0^{-1}) \tag{76}$$

where we have used $r_{12}^2 = P_{cv}^2 / m_0^2\omega^2 = \hbar / 2m_c\omega$. The absorption is quite large (Fig. 3(a)) and therefore the length of the modulator, according to (10) can be small, $L_{ex} \approx 2.302 t_{eff} a_{ex}^2 / 4\pi\sigma_{ex}(\omega) \sim 10 t_{eff}$. Next, we inject free carriers into the TMDs; where three processes take place simultaneously: (i) state filling, (ii) bandgap renormalization and (iii) screening. Thus, excitons bleach as recently observed by Heinz who attributed it to Mott transition [43]. Mott transitions occur in 3D semiconductors when the screening radius becomes comparable to the exciton radius, i.e. when $n_{2D} \sim a_{ex}^{-2}$. Strictly speaking the screening in 2D system saturates. Therefore, exciton bleaching most likely takes place because of state filling [12]. The exciton wave function can be considered a coherent superposition of the electron-hole-pair states with the wave vectors between 0 and roughly $1/a_{ex}$. The density of these states (with spin and valley degeneracy) is roughly $n_{bl} \approx 4/\pi a_{ex}^2$. The exact dependence of excitonic absorption on the density of injected carriers may be quite complicated, however, for our order-of-magnitude analysis it can be linearized as

$$\alpha(\omega, n_{2D}) = \alpha(\omega,0)\times(1 - n_{2D}/n_{bl}) = \frac{8\alpha_0}{a_{ex}^2 t'_{eff}} \frac{\hbar}{m_c\gamma}\left(1 - \pi a_{ex}^2 n_{2D}/4\right) = \frac{8\alpha_0}{a_{ex}^2 t_{eff}} \frac{\hbar}{m_c\gamma} - \sigma'_{ex} n_{2D} \tag{77}$$

As one can see, the exciton radius and binding energy do not play a role in the determination of the switching charge, which is still determined by (8) with the effective cross-section for the exciton as

$$\sigma_{ex}(\omega) = 2\pi\alpha_0 \frac{\hbar}{m_c\gamma_{ex}} \approx 3.53\times10^{-17} cm^2 \frac{1}{\hbar\gamma_{ex}}\left(\frac{m_0}{m_c} - 1\right) \tag{78}$$

which is Eq. (33) from the main text.

Funding

Army Research Office (ARO) (W911NF-16-2-0194); Air Force Office of Scientific Research (AFOSR) (FA9550-17-1-0377).

131.94.11.215 {ts '2018-05-22 10:37:02'}