at a frequency range of 25 to 425 MHz, with an energy consumption of >1000 frames per joule, making it suitable for applications that include a self-driving car.

A limitation of Modha *et al.*'s approach is its relatively high power requirements when scaled to large AI models. This is partly due to the data architecture (how data are collected, stored, and flowed through the system) that permits low-latency inference by making data transfers possible from memory located both near and far across the chip. Increased power and size demands also arise from elaborate peripheral chips (such as field-programmable gate arrays coprocessors) and memory that are needed to power up the compute-memory integrated chip from a cold start. From this perspective, an edge device using NorthPole must have access to a reasonably high-power source and incur any associated increase in its size. Several applications of AI inference chips, however, include functioning within sufficiently powerful machines such as automobiles, aircraft, and military vehicles, which would benefit from the chip design of Modha *et al.*

"...analog approaches may ultimately bring real-time AI to hand-held edge devices..."

In the quest for lower-power AI edge inference accelerators, a somewhat different architecture called compute-in-memory (CIM) has been explored (*2*). CIM brings memory (weights) and compute (multipliers) units physically close together, at the level of the multiplier units used in MACs (recall that MACs are the primary operations required for inference). CIM can be either digital or analog. In digital CIMs, the weights are stored as bits using static random-access memory (which retains data as long as power is being supplied) and the multiplication is finite precision binary arithmetic, similar to the operation used by Modha *et al.* An analog CIM is more elegant in that it directly uses Ohm's law to perform multiplication and uses a capacitor as an adder. It has potentially much lower power needs compared to digital MACs. The weights are stored as the conductance of a conductive element, and current flowing through the element is proportional to the product of input (applied voltage) and the conductance (weights). The currents through several such elements are added to provide the primary inference functionality. In principle, this MAC operation is very fast, and the power for inference is attractively low.

The challenge, especially in the analog CIM approach, lies in the choice of the conductive element and the resolution of its conductance, which determines the resolution of the weights. In one example, a combination of several phase-change memory elements is used to store the weights (*3*). However, such analog CIM systems require the integration of new materials, which necessitates the addition of specialized steps to the industry-standard manufacturing process for complementary metal oxide semiconductors. An approach that addresses this obstacle requires instead a standard semiconductor device that traps charge (called the charge trap transistor) and uses it to tune its resistance and achieve an estimated 8 equivalent bits of resolution in the resulting weights (*4*).

A disadvantage of analog techniques is that programming the weights is cumbersome, time consuming, and energy intensive. Thus, more suitable for edge applications would be AI models that are updated only occasionally. In both analog CIM and digital approaches, precision of the weights and computations play a key role in accelerating inference. Thus, more work is necessary on designing training algorithms to yield AI models that are robust to both reduction in precision and systemic biases in analog devices. Recent progress reported for analog (*2*) and digital (*5*) platforms is encouraging.

Innovative chip architecture and digital approaches such as those reported by Modha *et al.* will play an important role in the near-term development of edge-based inference with moderate-size platforms. However, analog approaches may ultimately bring real-time AI to hand-held edge devices, especially as the enabling technologies become more mainstream and easily manufacturable. ■

REFERENCES AND NOTES

1. D. S. Modha, F. Akopyan, A. Andreopoulos, R. Appuswamy, J. V. Arthur, A. S. Cassidy, P. Datta, M. V. DeBole, S. K. Esser, C. Ortega Otero, J. Sawada, B. Taba, A. Amir, D. Bablani, P. J. Carlson, M. D. Flickner, R. Gandhasri, G. J. Garreau, M. Ito, J. L. Klamo, J. A. Kusnitz, N. J. McClatchey, J. L. McKinstry, Y. Nakamura, T. K. Nayak, W. P. Risk, K. Schleupen, B. Shaw, J. Sivagnaname, D. F. Smith, I. Terrizzano, T. Ueda, *Science* **382**, 329 (2023).
2. Z. Wan, T. Wang, Y. Zhou, S. S. Iyer, V. P. Roychowdhury, *ACM J. Emerg. Technol. Comput. Syst.* **18**, 1 (2022).
3. M. Le Gallo *et al.*, *Nat. Electron.* **6**, 680 (2023).
4. X. Gu, Z. Wan, S. S. Iyer, *IEEE Trans. Electron. Devices* **66**, 10, 4183 (2019)
5. T. Dettmers, M. Lewis, S. Shleifer, L. Zettlemoyer, "8-bit optimizers via block-wise quantization," 9th International Conference on Learning Representations, ICLR 2022, 25 to 29 April 2022.

10.1126/science.adk6874

PHOTONICS

Tracking light-induced charge transport

Precise charge dynamics could help to improve the operation of solar cells and sensors

By **Rachel E. Bangle** *and* **Maiken H. Mikkelsen**

Key to developing photocatalytic and solar cells is understanding charge dynamics. Light absorption in metals can initiate charge transport by exciting electrons to form high-energy "hot" charges, but this energy is rapidly lost as heat (*1*). To capture the extra energy that hot charges (or carriers) possess before they decay and use it to generate an electrical current, the carriers must be transported across an interface called a transport junction. The movement of hot charges determines the efficiency of these junctions. However, hot-carrier dynamics are notoriously difficult to measure owing to their ultrashort lifetimes (*2*–*4*). On page 299 of this issue, Taghinejad *et al.* (*5*) show that ejection of hot carriers at a transport junction produce terahertz (THz, 10^{12} Hz) radiation that provides information exclusively on hot-carrier dynamics. This could provide a reliable way to characterize hot carriers and thereby aid the design of sensors and solar cells.

Hot carriers are formed when collectively oscillating electrons, called plasmons, transfer energy within the metal to form excited electron and hole (positive charge) states on timescales of <100 fs (1 fs = 10^{-15} s). This initial state is very energetic and thus outside of the usual thermal equilibrium of electrons in metal. However, nonequilibrium carriers rapidly (<1 ps, 1 ps = 10^{-12} s) spread that energy among a larger electron population through scattering collisions. This raises the average electron temperature, and the electrons quickly (<100 ps) dissipate the excess energy to the metal lattice as heat. Hot-carrier extraction can occur during

Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA.
Email: m.mikkelsen@duke.edu

this process if the electrons have enough energy to overcome the energetic barrier at a metal-dielectric interface (the transport junction). It is unknown, however, whether electrons in thermal equilibrium and excited through scattering can participate in charge transport along with nonequilibrium carriers (*6*).

Experiments directed at measuring hot-carrier dynamics are obscured by the multiple hot-carrier decay pathways that occur on similar timescales in the same locations (*7*). Traditional charge transport measurements have reported numbers of photoejected electrons (*8*) and have quantified the energetic distribution of hot carriers (*9*). These measurements, however, have not provided insights into hot-carrier dynamics. Ultrafast optical measurements have shown electron ejection to occur within 50 fs but have been unable to distinguish at which stage in the energetic decay hot carriers undergo ejection (*10*). Similar measurements have elucidated subpicosecond plasmon decay timescales but have not revealed ejection timescales (*11*). Further, theoretical descriptions of hot-carrier transport are complicated by the multiple distance- and time-scales that must be considered (*12*). Without a comprehensive model, discrepancies between measured charge transport efficiencies cannot be explained (*1*), and the contributions of hot carriers in each of the early decay stages remain unclear (*6*). Debate continues as to whether effects attributed to hot-carrier transport might instead be due to heating (*13*).

Taghinejad *et al.* report a measurement technique for hot-carrier transport dynamics in a nanoscale metal-dielectric transport junction composed of gold and titanium dioxide. The technique excites the metallic nanostructure with pulsed near-infrared light, which generates a current that oscillates on picosecond and subpicosecond timescales (see the figure). The oscillating current then results in terahertz radiation, which is detected with an ultrafast electro-optic detector. Although indirect, this terahertz radiation comes exclusively from hot carriers expelled from the metal, effectively disentangling electron transport from alternative hot-carrier decay pathways with similar timescales.

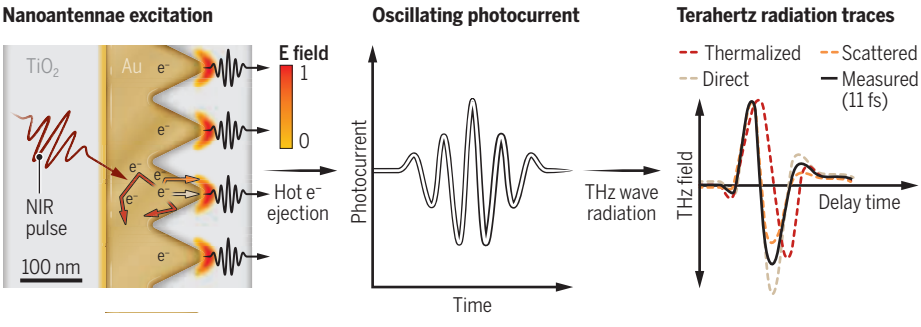
The magnitude of terahertz radiation substantially increased because the authors used asymmetric metal-dielectric junctions to generate the hot carriers. These junctions consisted of thin, gold nanostripes embedded in titanium dioxide, where one side of the stripe was flat, whereas the other had a sawtooth pattern. When excited with near-infrared light, these structures acted as nanoantennae, concentrating electric fields at the tips of the triangular gold teeth. This caused hot carriers to be preferentially created and ejected on the sawtooth side of the stripe. As such, the currents produced at the two nanostripe sides, which progress in opposite directions, were asymmetric and did not counteract each other. With this setup, the photocurrents (and thus the terahertz radiation) were amplified.

The terahertz radiation was modeled as a sum of different contributions to create a comprehensive mechanistic picture. Using this technique, Taghinejad *et al.* differentiated transport of hot carriers at each stage in their energetic decay. They determined that hot-electron ejection dynamics were influenced by both electrons immediately after direct excitation and electrons that had partially decayed through at least one electron-electron collision (scattering). Importantly, however, the terahertz radiation could not be adequately modeled with elevated electron temperatures, suggesting that electrons in thermal equilibrium do not meaningfully contribute to hot-carrier ejection.

By extending the technique reported by Taghinejad *et al.* to metal-dielectric junctions made of other materials and with different sizes, shapes, and environments, transport junction properties can be linked to hot-carrier dynamics. This will lead to predictive design of hot carrier–based sensors, solar cells, and photochemical reactions with energetic structures that capture the maximum number of hot carriers and with materials and dimensions that extend hot-carrier lifetimes. In addition, these measurements provide the opportunity to explore the trajectory of hot-carrier ejection depending on the shapes and orientations of metallic nanostructures (*14*, *15*). This could allow control of the direction or location of hot-carrier transport—for instance, to vary the reaction site or product selectivity within a photocatalytic cell. The approach might also enable the creation of nanojunctions that are sensitive to a variety of light properties such as polarization, orbital angular momentum, or phase, which could allow the development of switchable sensors. Implementation of faster excitation sources may further expand this technique not only to time-resolve hot-carrier transport, but also to serve as an ultrabroadband terahertz source that is useful for imaging and free-space communications. ■

A strategy to follow the mechanism of hot-carrier ejection

Pulsed near-infrared (NIR) light excites an asymmetric gold stripe (nanoantennae) within a metal (Au)–dielectric (TiO_2) transport junction, which results in electric field concentration and asymmetric hot-electron (e^-) ejection (left). This generates a fast-oscillating photocurrent (middle) and a measurable radiated terahertz field that comes exclusively from hot-electron ejection (right; the graph is illustrative). Modeling the terahertz field shows that both direct and scattered electrons contribute to the terahertz trace, but thermalized electrons do not (right).



GRAPHIC: A. FISHER/*SCIENCE*

REFERENCES AND NOTES

1. J. B. Khurgin, *Nanophotonics* **9**, 453 (2019).
2. M. L. Brongersma, N. J. Halas, P. Nordlander, *Nat. Nanotechnol.* **10**, 25 (2015).
3. E. Cortéz *et al.*, *Nat. Commun.* **8**, 14880 (2017).
4. A. O. Govorov, H. H. Richardson, *Nano Today* **2**, 30 (2007).
5. M. Taghinejad *et al.*, *Science* **382**, 299 (2023).
6. P. Christopher, M. Maskovits, *Annu. Rev. Phys. Chem.* **68**, 379 (2017).
7. A. Manjavacas, J. G. Liu, V. Kulkarni, P. Nordlander, *ACS Nano* **8**, 7630 (2014).
8. C. Clavero, *Nat. Photonics* **8**, 95 (2014).
9. H. Reddy *et al.*, *Science* **369**, 423 (2020).
10. L. Du, A. Furube, K. Hara, R. Katoh, M. Tachiya, *J. Photochem. Photobiol. C Photochem. Rev.* **15**, 21 (2013).
11. R. H. M. Groeneveld, R. Sprik, A. Lagendijk, *Phys. Rev. B Condens. Matter* **51**, 11433 (1995).
12. P. Narang, R. Sundararaman, H. A. Atwater, *Nanophotonics* **5**, 96 (2016).
13. B. Y. Zheng *et al.*, *Nat. Commun.* **6**, 7797 (2015).
14. M. Song *et al.*, *Appl. Phys. Rev.* **6**, 041308 (2019).
15. J. Baumberg, J. Aizpurua, M. H. Mikkelsen, D. R. Smith, *Nat. Mater.* **18**, 668 (2019).

ACKNOWLEDGMENTS

The authors are supported by the Air Force Office of Scientific Research (award FA9550-21-1-0312).

10.1126/science.adk6862

Tracking light-induced charge transport

Rachel E. Bangle and Maiken H. Mikkelsen

Science **382** (6668), . DOI: 10.1126/science.adk6862