ICSNN 2014, Savannah, USA, August 3-8, 2014
SAND2014-15662PE

Quantum Transport Simulation and Optimization of Below-6-nm Si FinFETs with HfSiON/SiO2 Gate Dielectrics

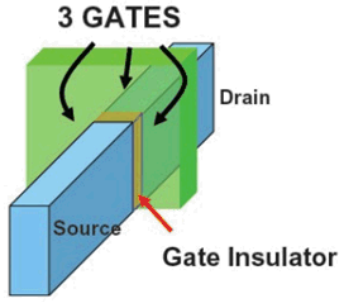
Xujiao (Suzey) Gao and Denis Mamaluy
Sandia National Laboratories, Albuquerque, NM

xngao@sandia.gov

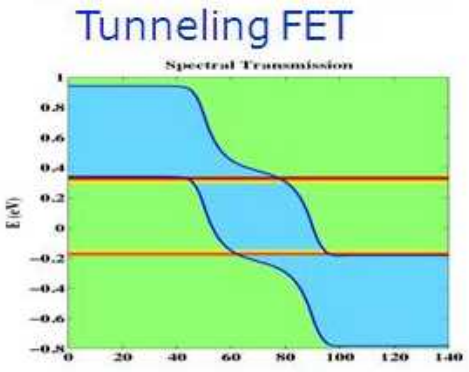
Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

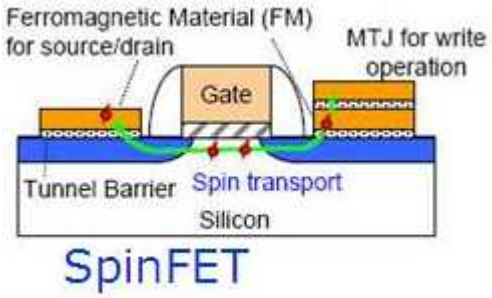
Motivation

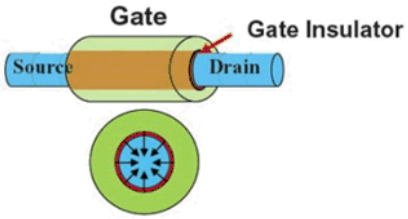
Why FinFETs ? State-of-the-art FinFETs still show the best performance for logic computing compared to the emerging beyond-CMOS devices.



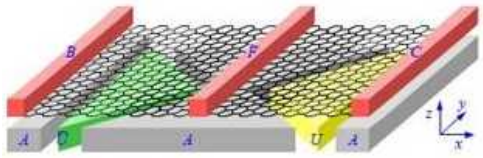
FinFET (Tri-gate)



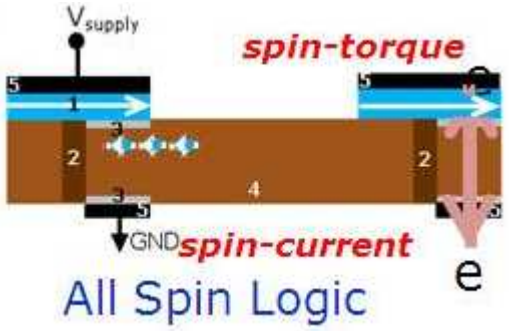




nanowire



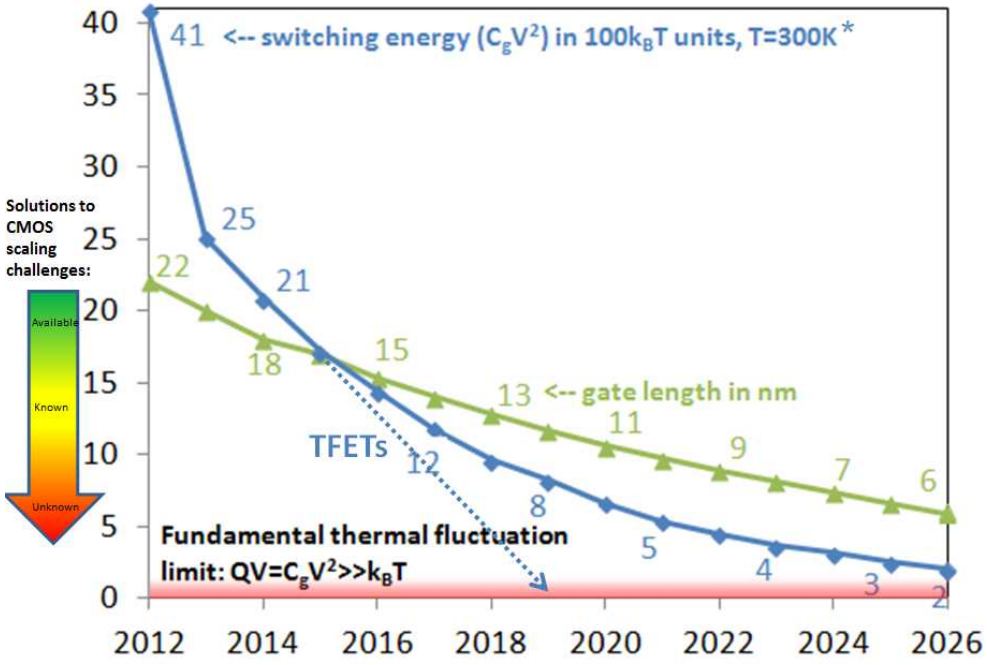
Graphene pn Junction



Motivation

Why below 6 nm ? ITRS projection to 6-nm gate length, but the gate switching energy (C_gV^2) is fast reduced with decreasing gate length.

At which gate lengths would the switching energy reach the thermal fluctuation limit (e.g., $C_gV^2 < 100\ k_BT$[1]) ?



Switching energy of TFETs would reach the thermal fluctuation limit even faster due to the lower voltage.

*Calculated from ITRS 2011/12 editions data for HP logic devices (PIDS2 tables)

[1] M.P. Frank, "The physical limits of computing," *Computing in Science & Engineering*, vol. 4, 2002, pp. 16-26.

Quantum Transport

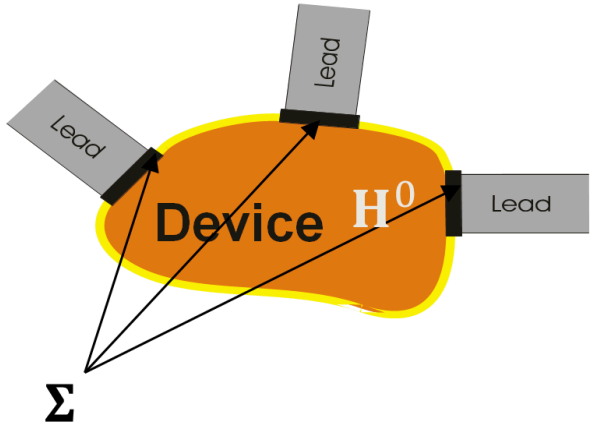
Why quantum ? As devices are shrinking towards 'true' nm scale (~10 nm feature size)[2], quantum mechanical effects[3] dominate.

Nonequilibrium Green's Function (NEGF) Formalism

- Describe quantum transport in open nanodevices
- The retarded Green's function (G^R) determines the device terminal properties (e.g., transmission, current, etc.)

$$\mathbf{G}^R = \left[E\mathbf{1} - \mathbf{H}^0 - \mathbf{\Sigma}\left(E\right)\right]^{-1}$$

- Determining G^R needs to invert a huge matrix, requiring $N_E \times O(N^3)$ cost



[2] International Technology Roadmap for Semiconductors (ITRS), 2013.
[3] G. Timp *et al.*, "The Ballistic Nano-transistor", *IEDM Tech. Dig.*, pp. 55-58,1999.

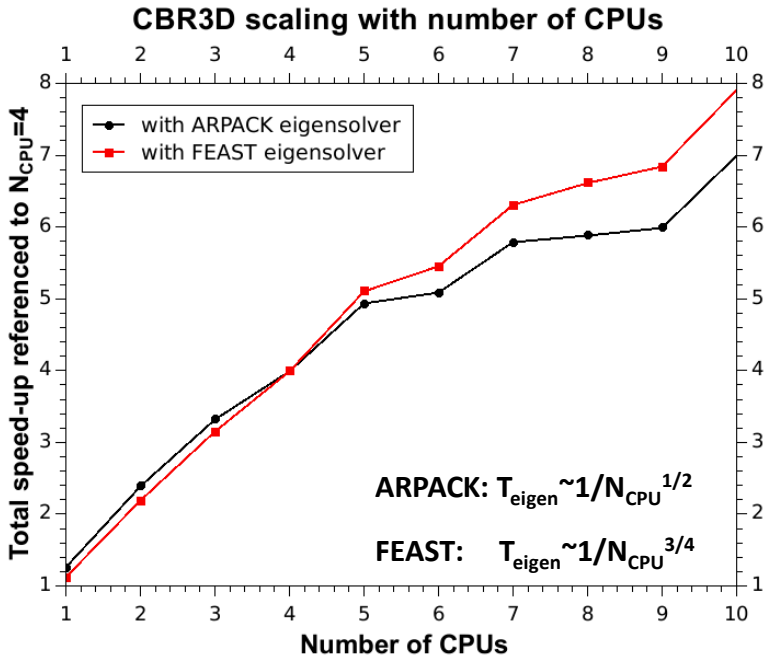
Contact Block Reduction (CBR) Method

CBR[4] **–** A very efficient approach to determine G^R

- Computational cost about $N_E \times O(N)$
- Applicable to nanodevices with any number of contacts

Highlights in the CBR Simulator:

- Fully charge self-consistent quantum transport in 2D & 3D
- Include dominant scatterings (e.g., impurity, roughness, e-e, phonon)
- Nearly linear speed-up w/ # of cores
- Significantly faster[5] than other quantum transport simulators



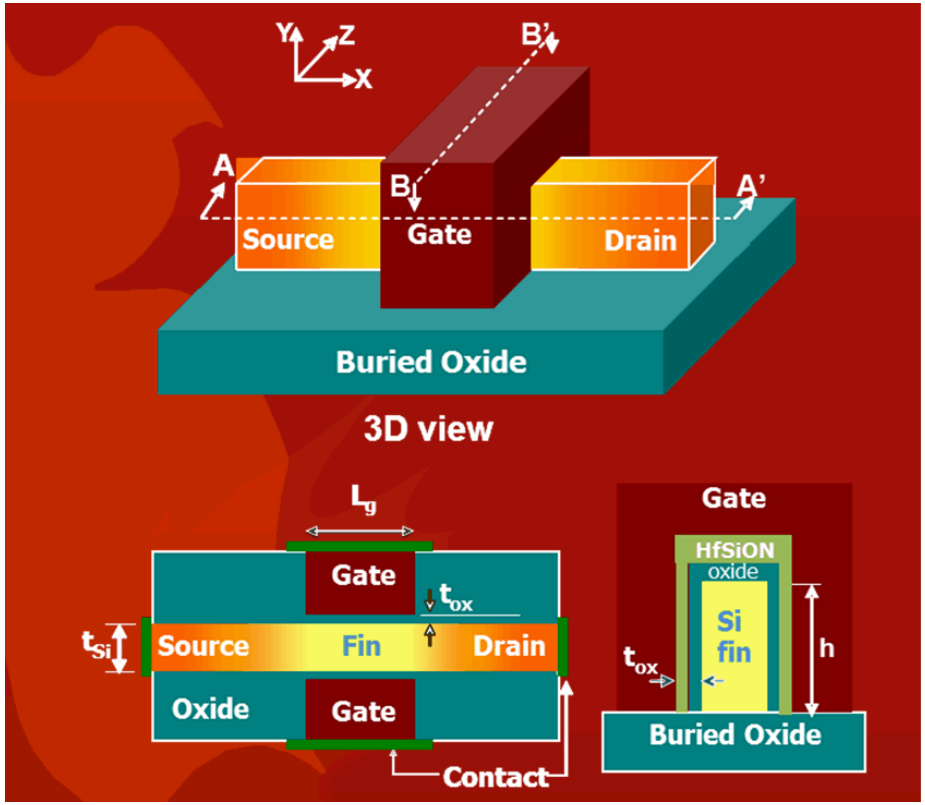
[4] D. Mamaluy, et. al., J. App. Phys. **93**, 4628 (2003); Phys. Rev. B **71**, 245321 (2005); IEEE T-ED **54**, pp. 784-796 (2007).

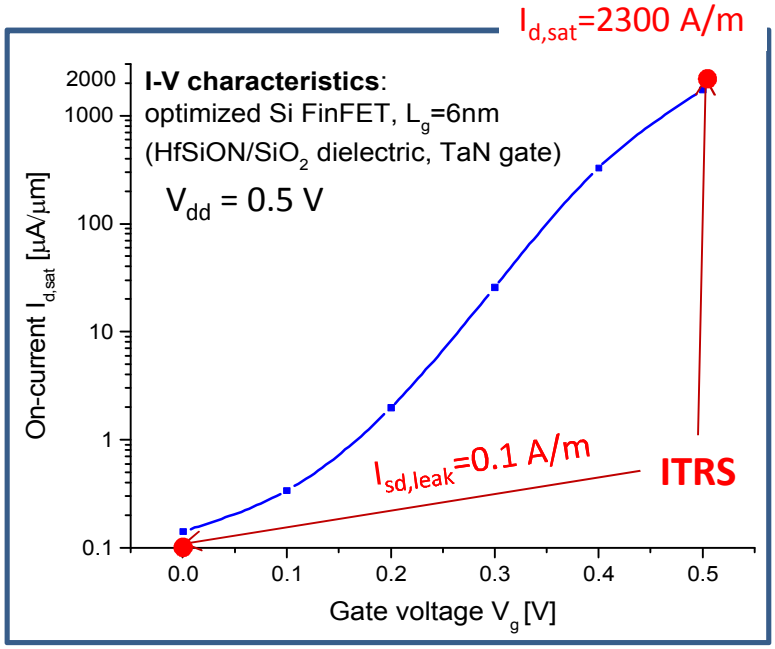
[5] Y. M. Sabry et al., Int. J. Numer. Model. (2010) DOI: 10.1002/jnm.780.

Optimized 6-nm Si FinFET

Device – HfSiON/SiO_2 gate dielectric, TaN metal gate (explicitly simulated)

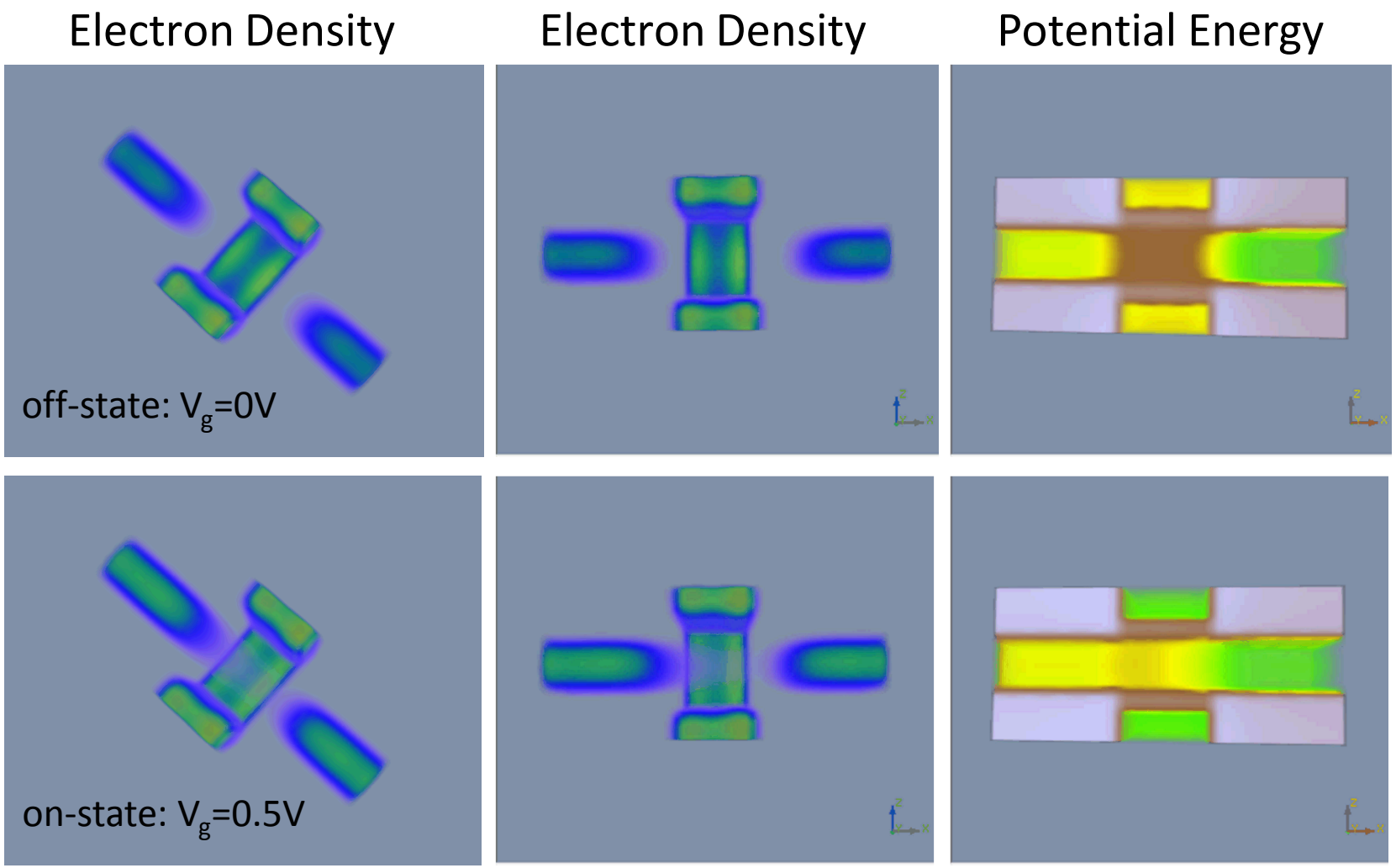
Optimization – optimize geometry and doping profile and select the I-V characteristics satisfy the ITRS projections





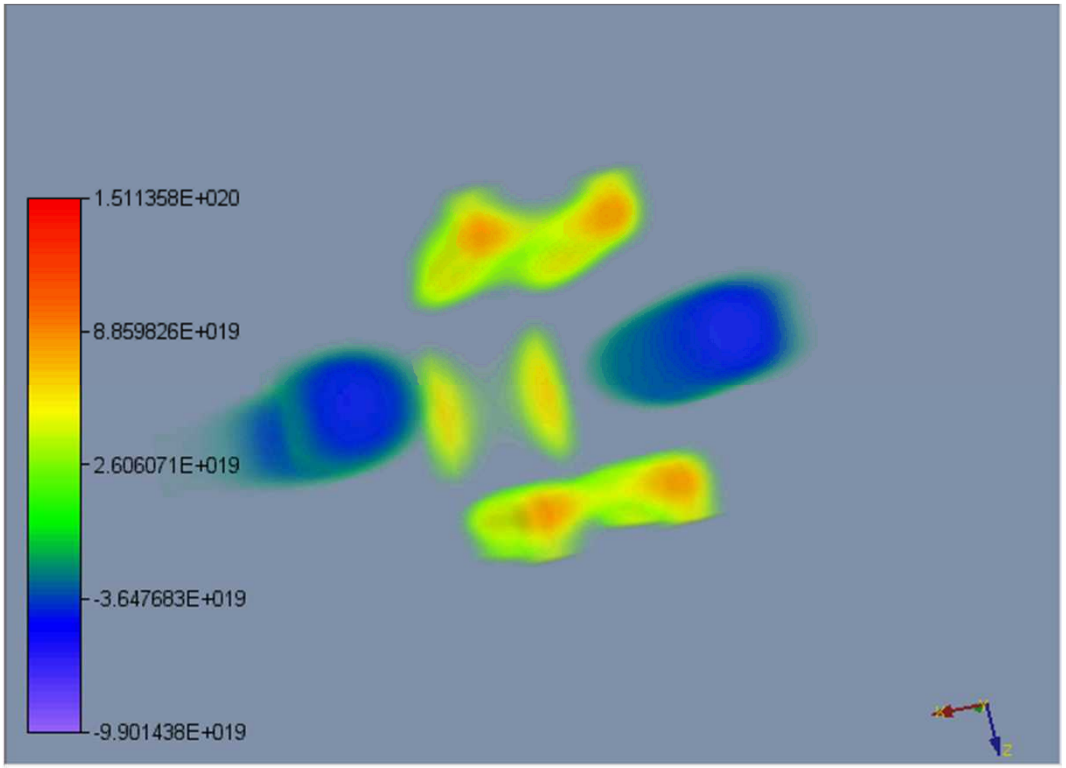
Sandia National Laboratories

Electron Density & Potential Energy



Red – high, blue – low, linear scale

Induced Capacitive Charge Distribution



Induced charge

$$c(r) = \frac{\Delta Q}{\Delta V_g} = \frac{-q\Delta n(r)}{\Delta V_g}$$

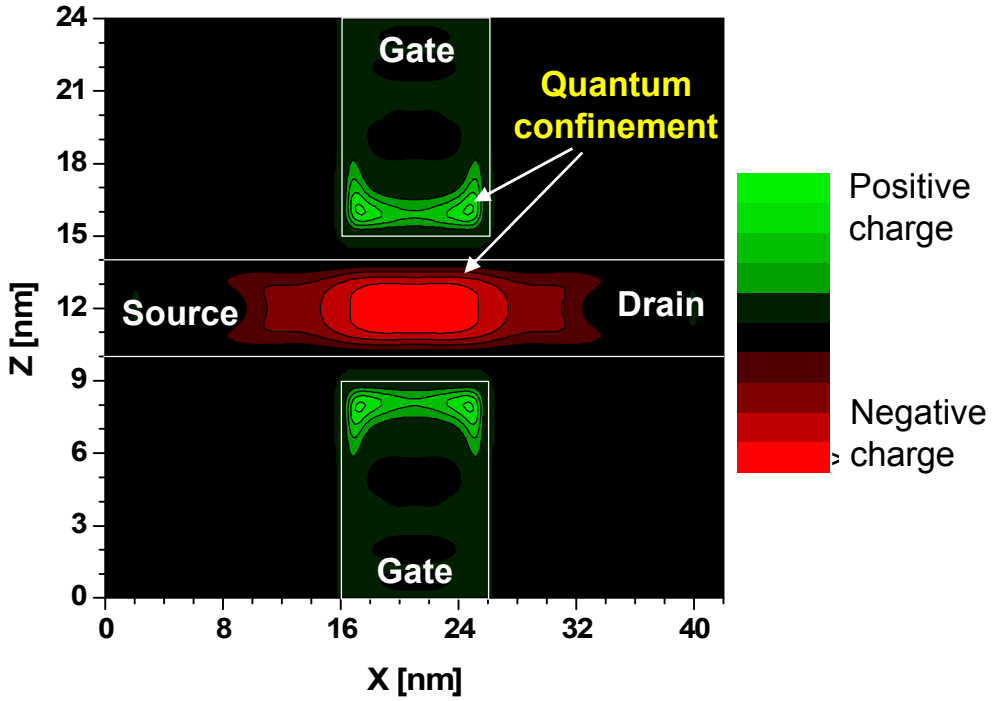
Gate capacitance

$$C_g = \frac{\int \Delta Q \, dr}{\Delta V_g}$$

- The optimized 6-nm FinFET is dominated by the fringing capacitance (the induced charge shows peak near the S/D regions).
- The induced charge distribution in the gate shows a complex profile, capturing quantum confinement and the gate-channel/S/D charge interactions.

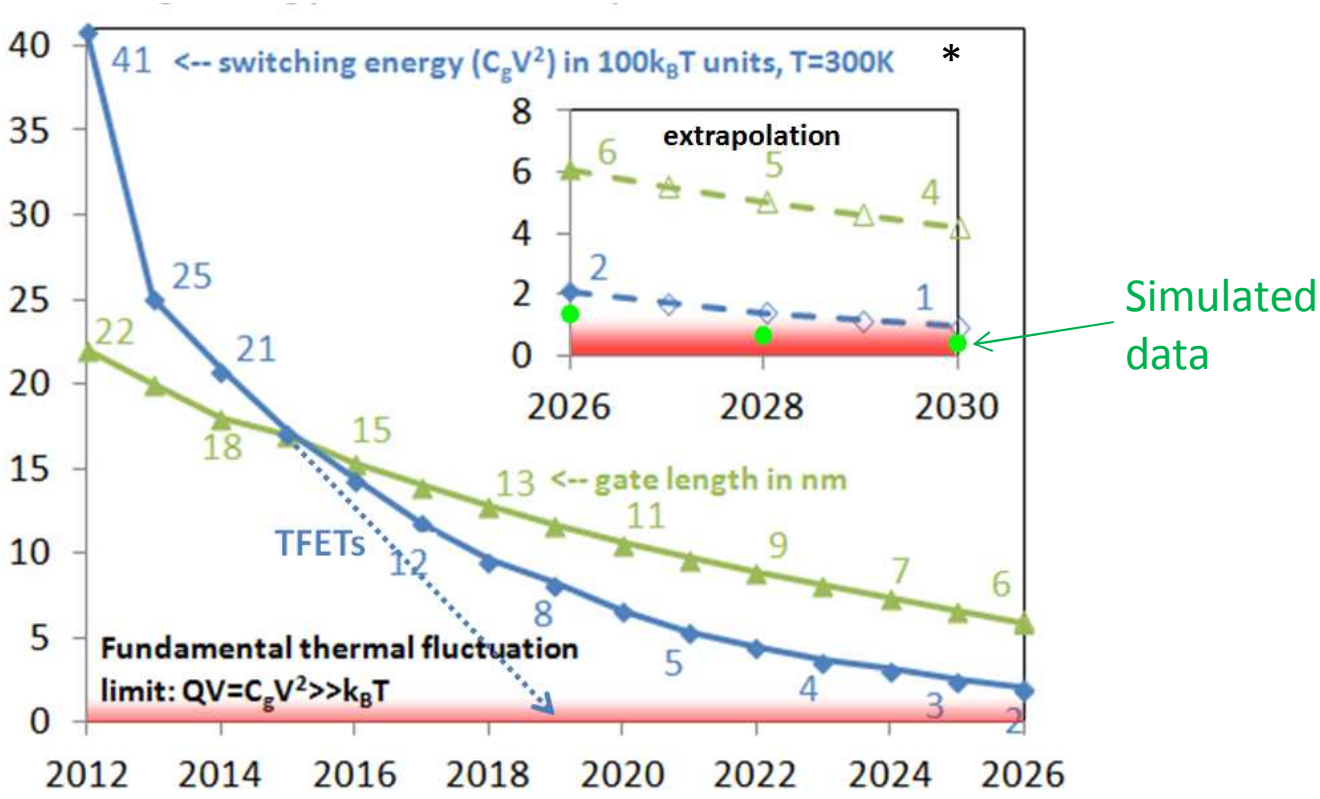
Induced Capacitive Charge Distribution

In contrast, the induced charge in an optimized 10-nm FinFET[6] is peaked in the channel.



[6] H. Khan, D. Mamaluy, D. Vasileska, IEEE T-ED 55, pp. 743-753 (2008).

Switching Energy Comparison



- Our QM simulated data are about 10% lower than ITRS, due to quantum confinement in the gate.
- Switching energy approaches the thermal fluctuation limit (< 100 k_BT) at gate length < 5 nm, which holds true for all charge-based FETs.

*Calculated from ITRS 2011/12 editions data for HP logic devices (PIDS2 tables)

Conclusion

Our findings:

- The end of FET scaling is near!
- Utilization of TFET-like power saving structures will bring the end to Moore law even sooner.
- Alternative (III-V, carbon, etc.) channel materials will **not** fix the situation since the gate capacitance is mainly determined by the node geometry and dielectric material.

Possibilities *after* the thermal fluctuation limit is reached:

1) Accept the end of Moore's law and concentrate on power dissipation reduction.
2) FET alternatives (memristors, super-conductive logic, spintronics, etc).
3) Continue Moore's law with… single-electron transistors!
   $E_{switch}=C_gV^2=Q_gV=[Q_g=q]=q^2/C_g$
   SETs have the switching energy vs gate capacitance trend opposite to all other FETs!
   Thus, SET scaling below 5nm gate length may be possible.