Thirty Years of Monte Carlo Simulations of Electronic Transport in Semiconductors:
Their Relevance to Science and Mainstream VLSI Technology

M. Fischetti, S. Laux, P. Solomon, and A. Kumar
IBM Semiconductor Research and Development Center
IBM Research Division, T. J. Watson Research Center
Yorktown Heights, NY 10598

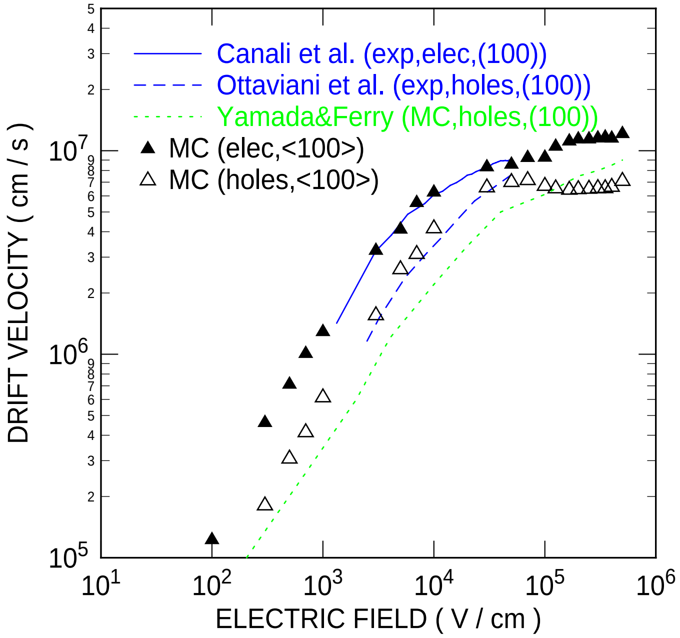
IWCE 10, October 2004

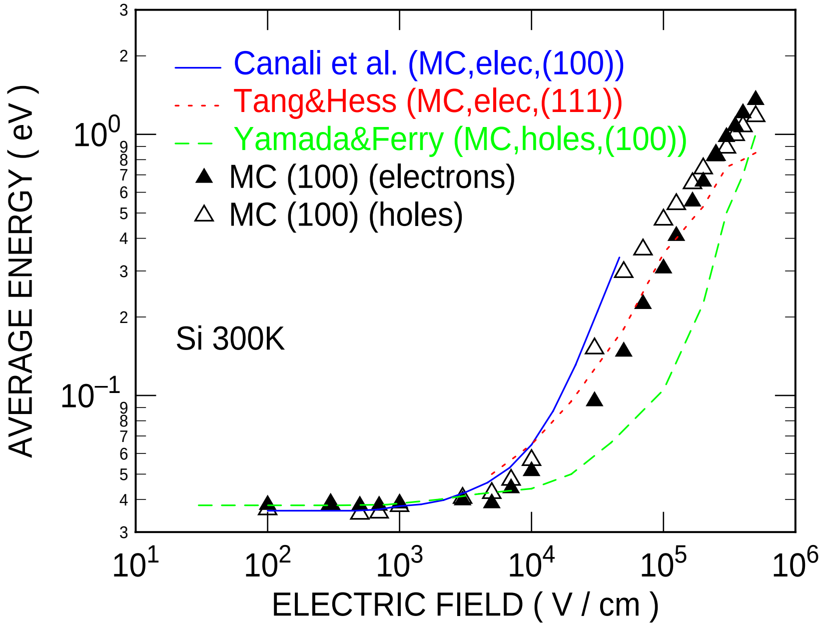
Outline

- The 'early days (*i.e.*, when we used to 'think'):
  Science, not much Technology
  - The basics of 'warm electron' transport: The Modena 'standard model'
  - The (oversold?) challenge of 'hot carriers': The 'new standard model'
  - Coulomb interactions
  - Technology? Just calibration of moments methods...
- The future days of the 'end of scaling' (*i.e.*, compute-and-do-not-think):
  Technology, not much Science
  - A little bit of Science: More Coulomb interactions
  - New devices (PD, FD and UTB SOI; Double-gate FETs, ...)
  - New materials (strained Si, Ge, III-Vs,...)
  - Old materials from a new angle ('new' crystal orientation)
- Basic (scientific?) questions at the end of the road:
  - Is ballistic transport a 'pipedream'?
  - Is the low-field mobility meaningful?
- Quantum transport: Science or fashion? Not for me to address...

1966: The dawn of Monte Carlo

- Moments of the Boltzmann Transport Equation (DD, maybe Energy Transport) more than enough to explain what little is needed about transport
- All device designers need to know is how to turn off the device
  Performance-gain comes from scaling, no matter whether we understand or not
- Non-thermal, strong off-equilibrium transport a ‘scientific ’curiosity
- HCIS 1966: Monte Carlo (Kurosawa, imported from the A-bomb) and iterative methods (Budd)
- III-Vs main target (Malvern group): Small mass, obvious heating effects (*e.g.*, Gunn effect), negative differential mobility hard to model within DD
- Silicon: The Modena ‘standard model’ (1970’s)

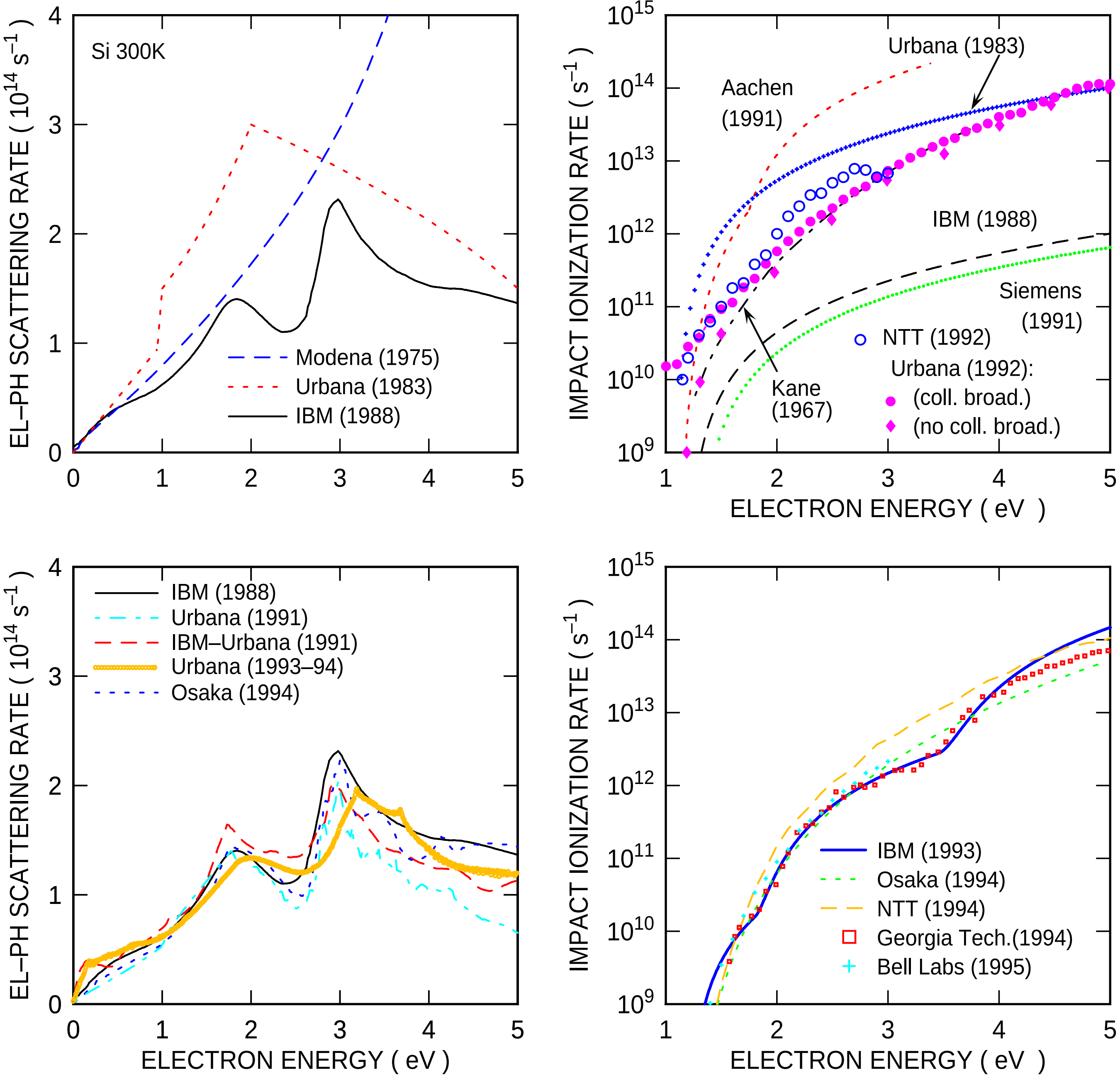




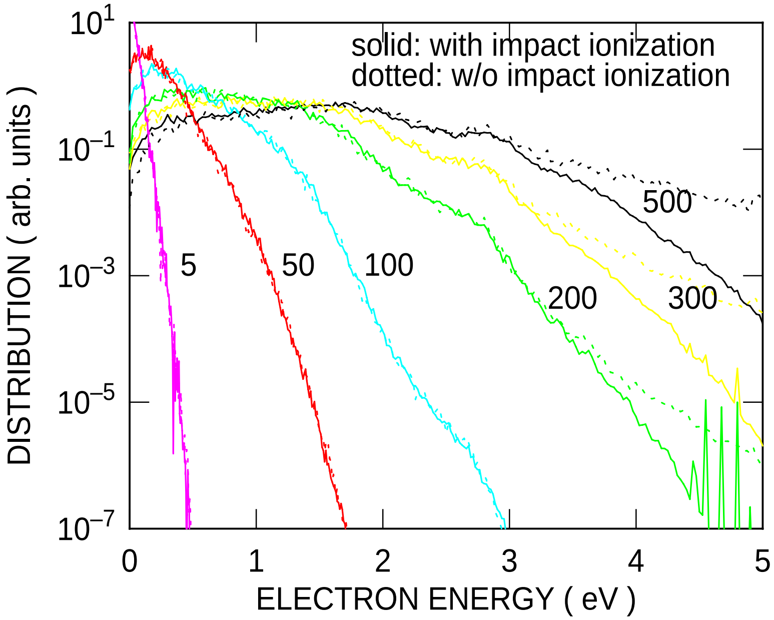
Hot carriers and the search for better models

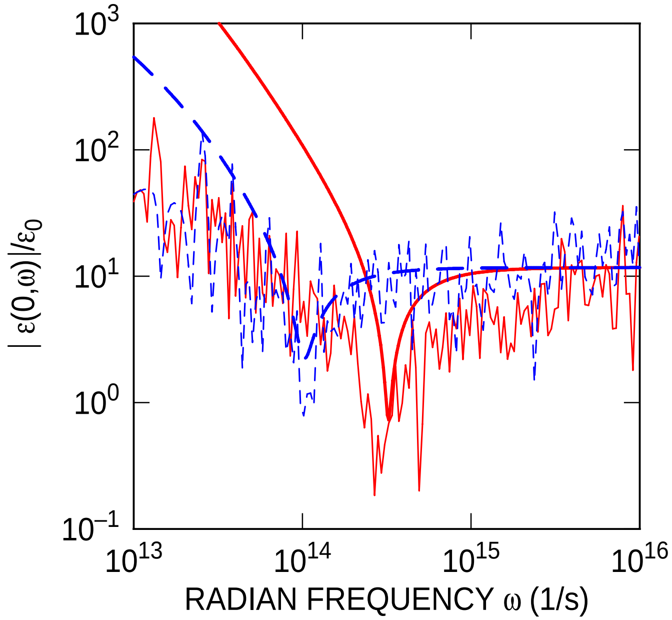
- Impact-ionization and injection into SiO_2 ‘practical’ problems (1980’s)
- Urbana: Full-band model, GaAs first, Si later
- The under-determined ‘rates’ problem:
  - Assume electron-phonon rates proportional to final DOS (Hess)...
  - Drift-velocity-*vs.*-field and ionization coefficients given...
  - No unique solutions: High-el-ph-rate AND high-ii-rate equally valid as low-el-ph-rate AND low-ii-rate
  - Theory (Urbana, IBM, Osaka, NTT) and experiments (IBM for ii) to the rescue
  - A passing fad: Band-structure fudged models

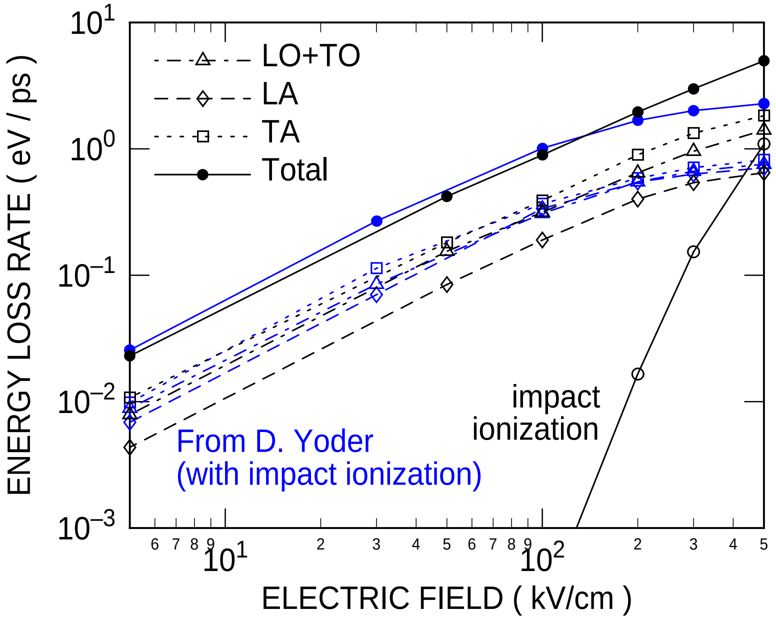
THE major achievement of MC methods: Old and new transport models

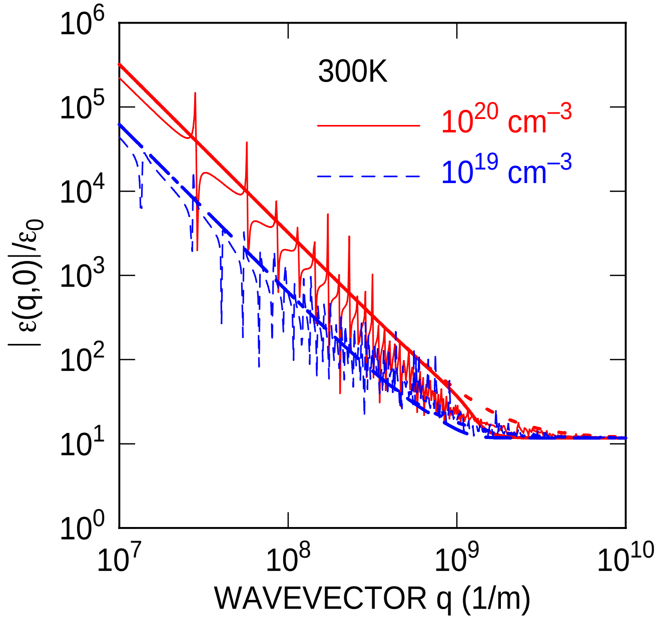


We even had time to think!



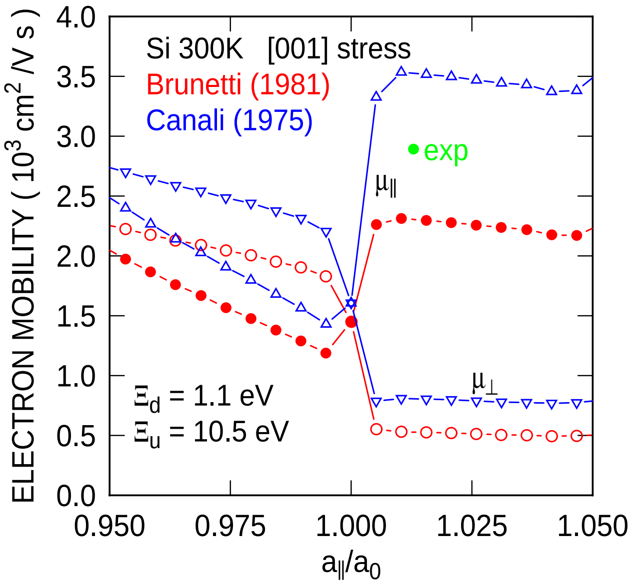


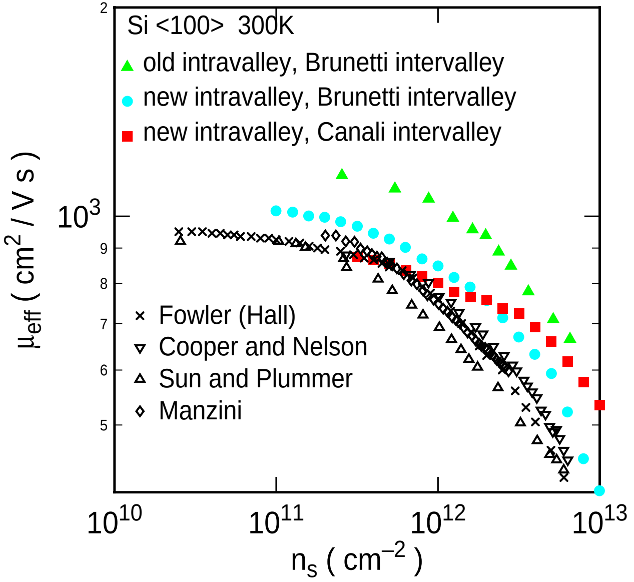




An example of ‘philosophy of Science’:
CB deformation potentials in Si and phonon scattering

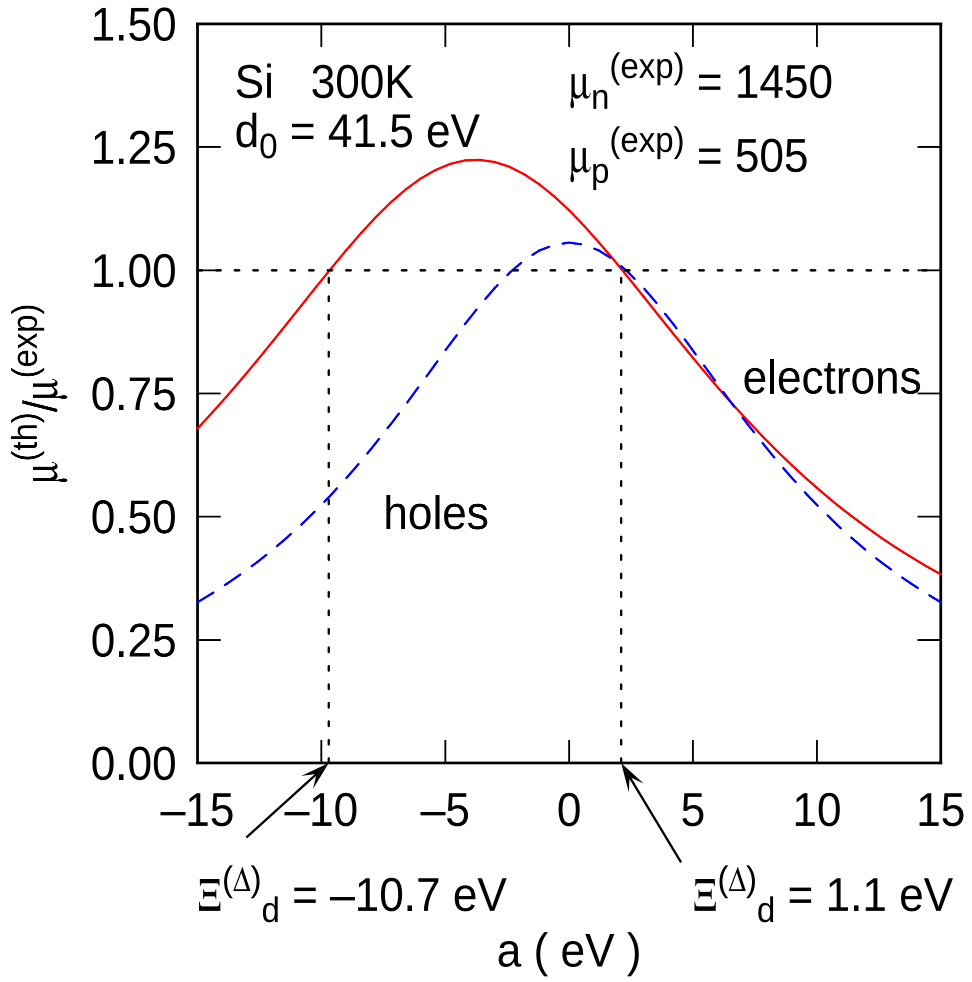
- 1993: 30% overestimation of the correct electron mobility in inversion layers with MC calculations using $\Xi_u = 9.0$ eV, $\Xi_d =$ - 11.7 eV from ‘selected’ experimental data (angle- and LA/TA-averaged $\Xi_{ave} \approx 10$ eV)
- Published proposal to use $\Xi_{ave} \approx 12$ eV (duh!)
- 1996: Compute $\Xi_u = 10.5$ eV (bulk strained Si) and determine $\Xi_d = 1.1$ eV from elec and hole mobility
- Revisit intervalley deformation potentials (from Brunetti ’79 back to Canali ’75)
- Explain both inversion-layer mobility (SR-scattering remains the issue) *and* mobility in bulk strained Si





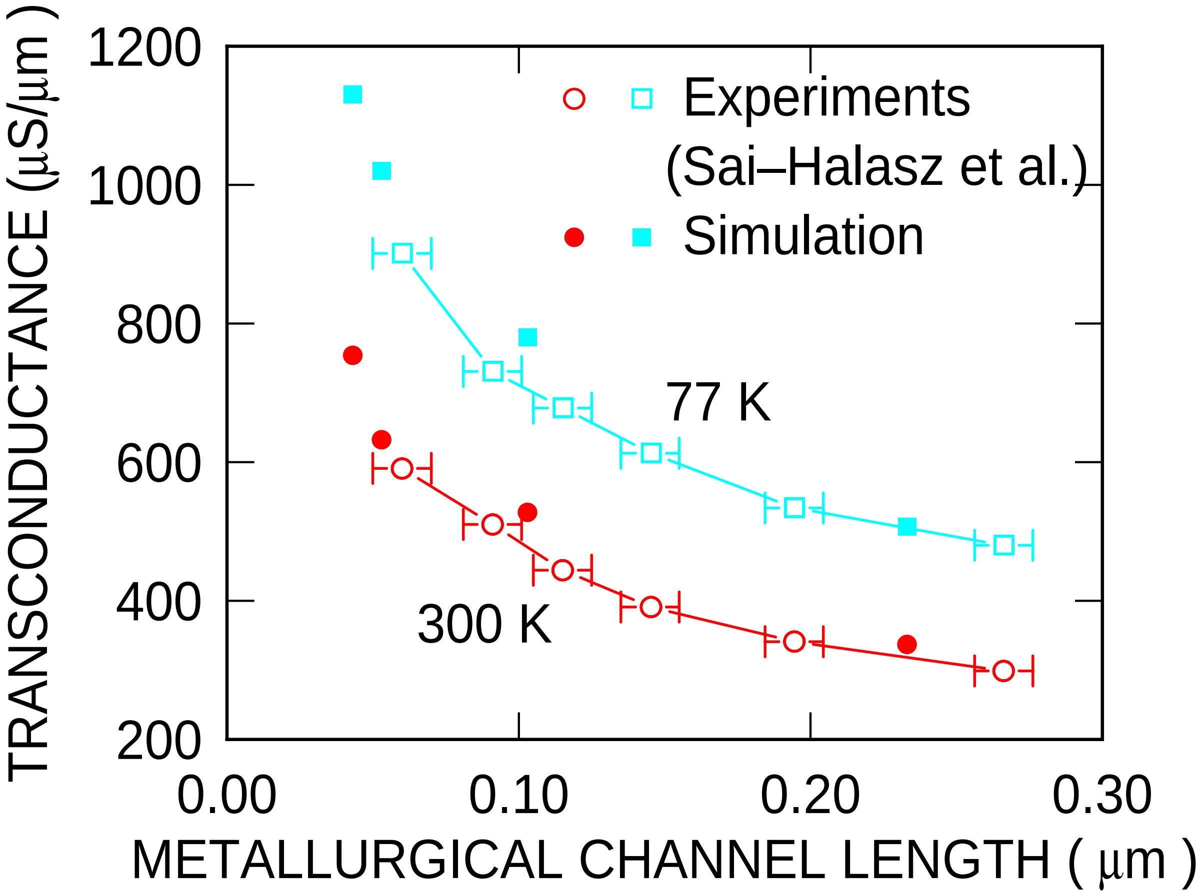
Determining Ξ_d

- Follow Herring and Vogt
- Fit simultaneously bulk electron *and* hole mobility



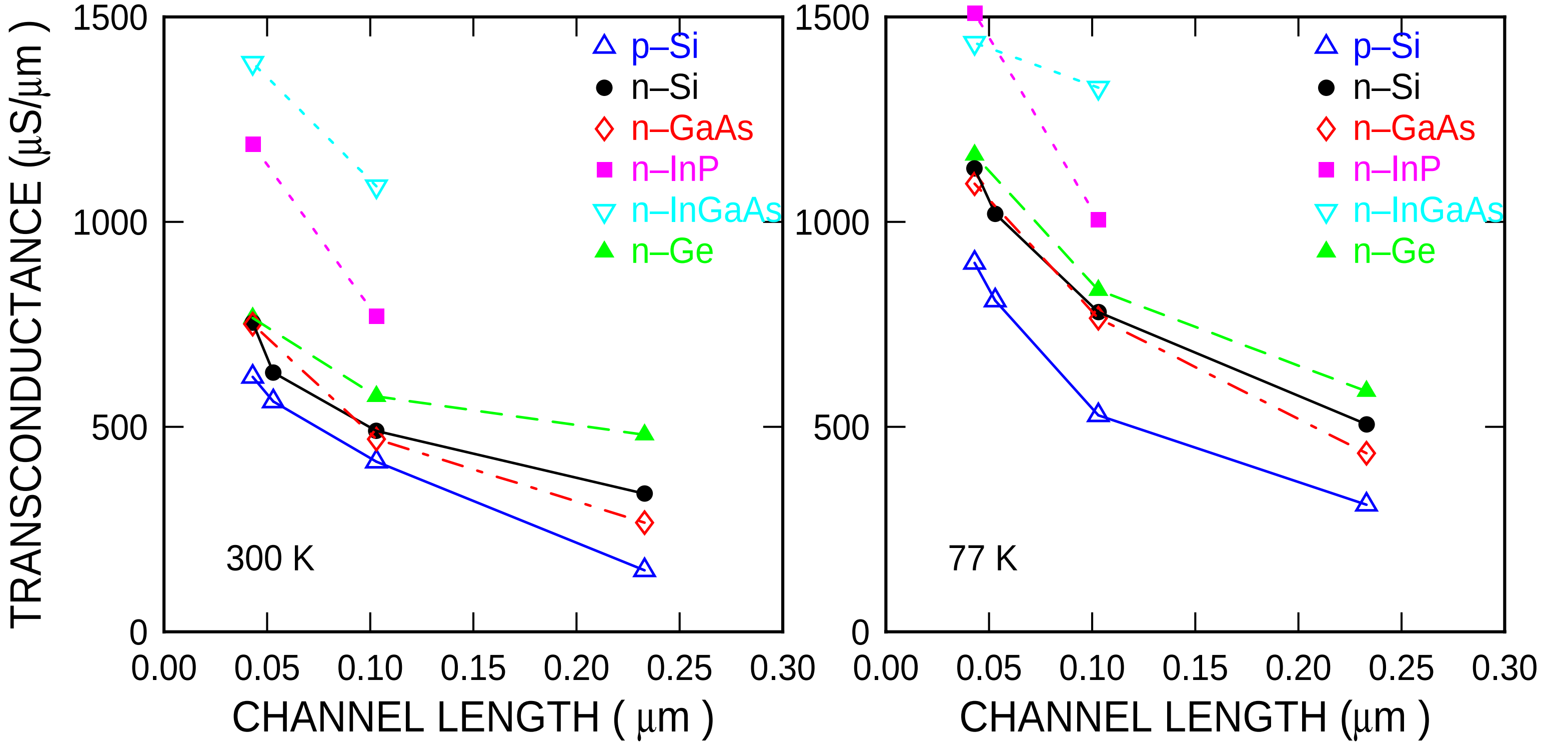
Sub-0.1μm FETs and velocity overshoot

- First 0.1 μm nFETs in 1987/88
- Good news back then: Unlimited performance gain at smaller dimensions... A ‘pipedream’?
- Evidence for velocity overshoot ‘scant’: The source fixes the current in ‘long’ devices



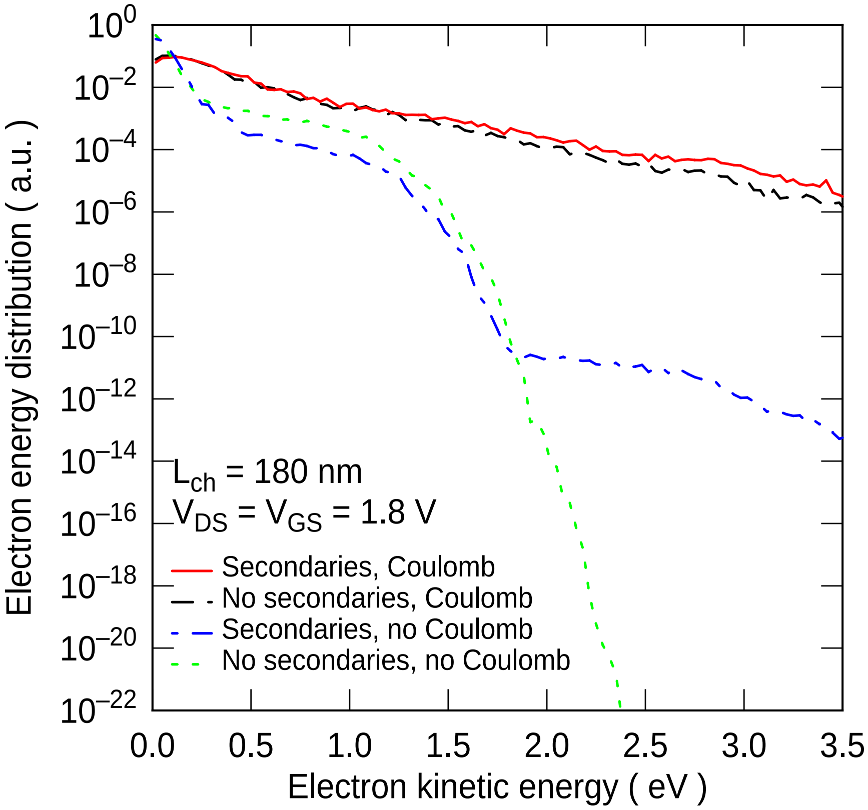
A first take on a controversial question: Is mobility important?

- Most semiconductors exhibit the same (calculated) performance
- Caused by:
  - Low DOS → loss of conduction channels → loss of transconductance (that is, low inversion capacitance)
  - Similar DOS (and so, scattering rates) for hot carriers (1 eV or so)
  - In-based materials an exception but at even shorter channel lengths...



Coulomb interactions I: Effect on energy-distribution

- Some things happen 'below threshold':
  - Substrate currents for V$_{DS}$ $\leq$ 1.1 V
  - Gate currents for V$_{DS}$ $\leq$ 3.2 V
- Strong thermalization caused by short-range electron-electron scattering
  - Strong high-energy tails above applied bias (in addition to the famous 'thermal tails')
  - Even stronger than 'ionization feedback' (Bude)



Towards the ‘end of scaling’. Coulomb interactions II: Effect on performance?

- Back to the present: Poor performance of aggressively-scaled devices
  - Scanning the literature: Poor performance of ‘record-braking’ devices
  - Off-line comments by Takagi-san (Toshiba, now at Univ. Tokyo)
  - Discussed in 2000 (IBM)
  - Emphasized by MIT (Lochtefeld and Antoniadis, IEEE EDL **22**, 95 (2001))

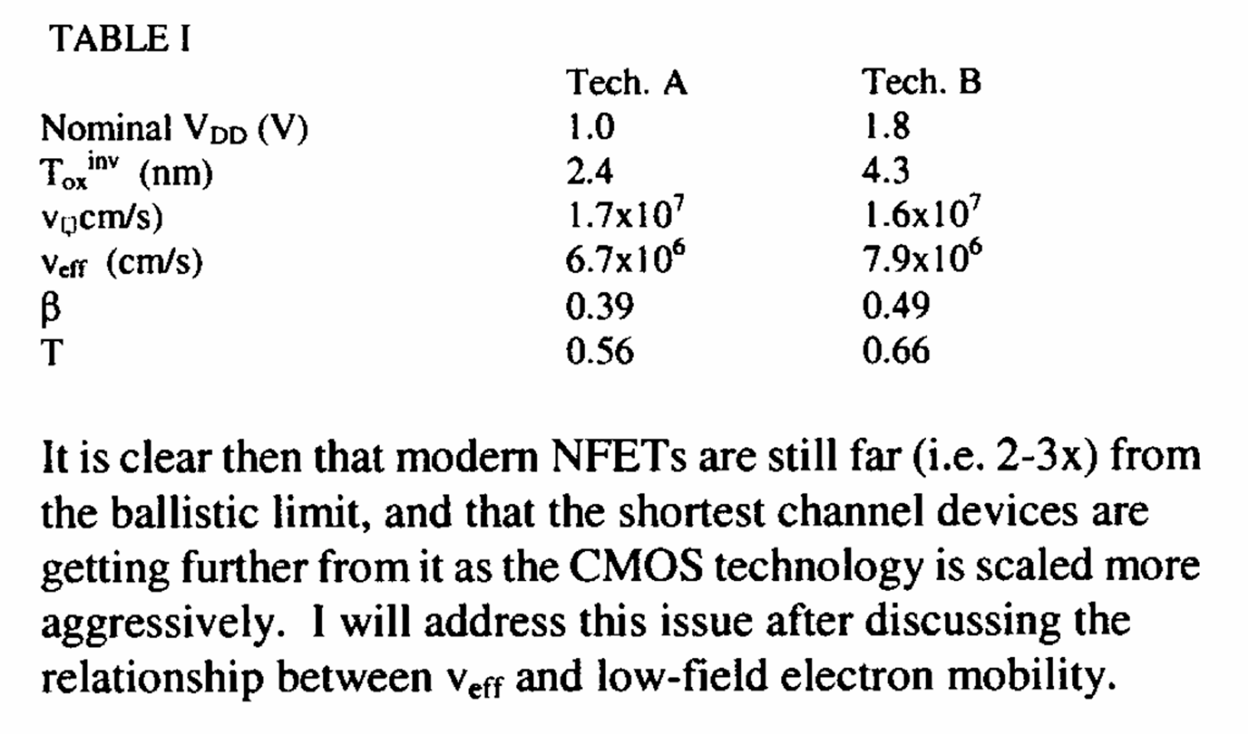


TABLE I

	Tech. A	Tech. B
Nominal V_{DD} (V)	1.0	1.8
T_{ox}^{inv} (nm)	2.4	4.3
$v_{\square}$cm/s)	1.7×10^7	1.6×10^7
v_{eff} (cm/s)	6.7×10^6	7.9×10^6
β	0.39	0.49
T	0.56	0.66

It is clear then that modern NFETs are still far (i.e. 2-3x) from the ballistic limit, and that the shortest channel devices are getting further from it as the CMOS technology is scaled more aggressively. I will address this issue after discussing the relationship between v_{eff} and low-field electron mobility.

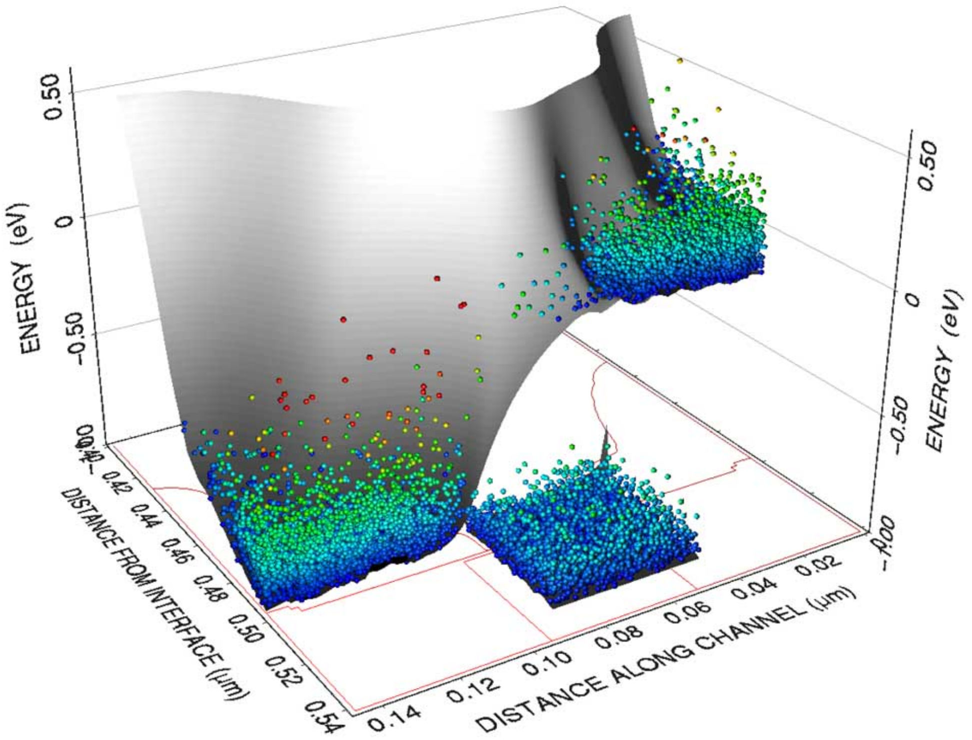
D. A. Antoniadis, 2002 VLSI Symp, p. 2

The ‘new scaling’

- Whatever the reason, we cannot scale forever and ‘scaling’ is now a different concept:
  - New device designs: SOI, ground-plane, Double-gate, FINFETs,...
  - New gate-insulators: HfO_2, $HfSiO_4$, rare-earth oxides, perovskites,...
  - New semiconductors: Strained Si, Ge, maybe III-V...
  - New contacts: metal gates, raised S/D, copper interconnects,...
  - New schemes for on-chip operation: Dual (or multiple) threshold, dual (or multiple) supply voltage,...
- The Physics: Less ‘elegant’, more challenging
- A new ‘culture’ (dictated by panic?): Lots of devices, no basic experiments

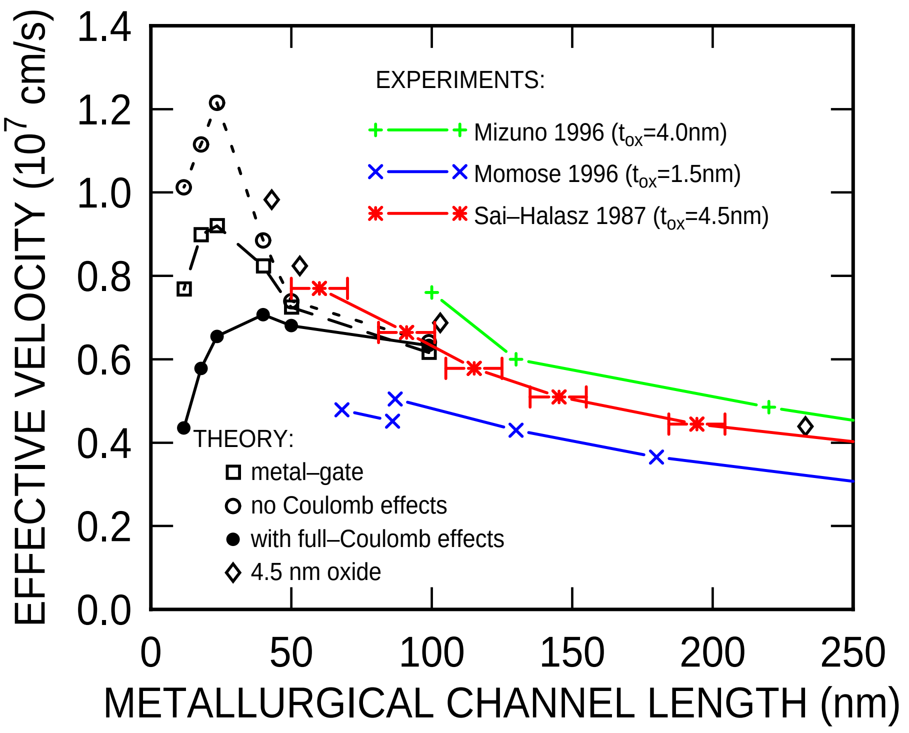
Long-range Coulomb interactions in small MOSFETs

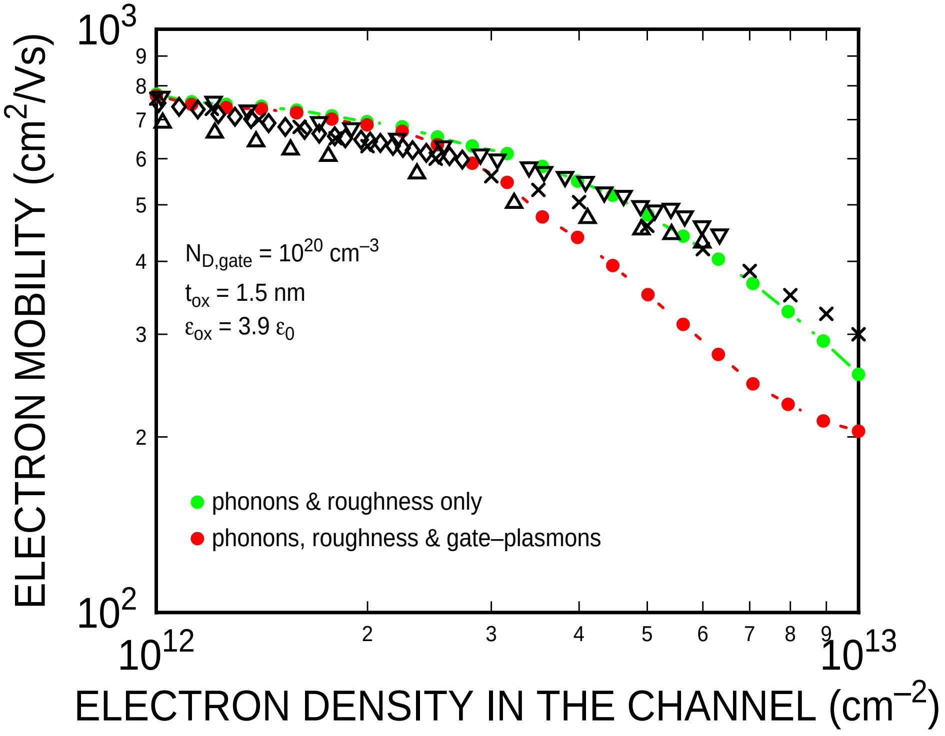
- Source, drain, and gate regions are high-density electron gases
- S/D separation (*i.e.*, channel length) is shrinking below the Debye length of the channel
- Gate needs to be 1 nm (or less!) away from the channel
- Collective 'fluctuations' in S/D perturb electrons in the channel (electron/bulk-plasmon interactions)
- Collective fluctuations in gate (interface plasmons) cause Coulomb drag



Coulomb interactions and device speed: Theory

- S/D interactions thermalize carriers, build high-energy tails, increase momentum-loss *indirectly*
- Gate-induced Coulomb-drag subtracts momentum *directly*
- Lower transconductance, lower mobility

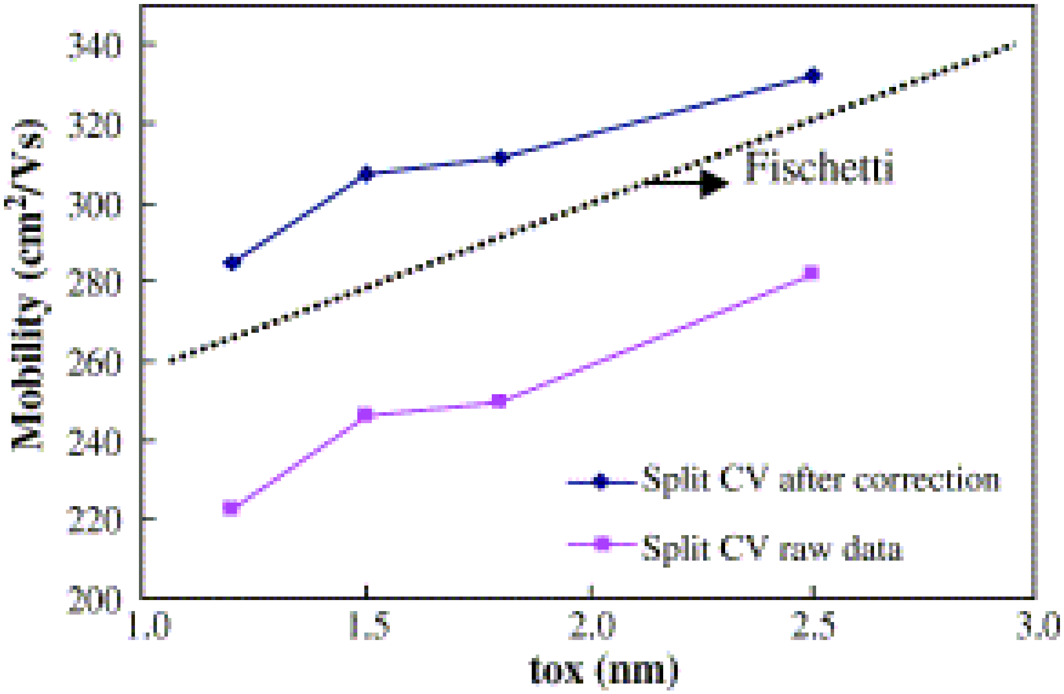




All results from full-band Monte Carlo simulation – DAMOCLES

Coulomb interactions and device speed:
Is it true?

- High-energy tails inferred from substrate currents at low energy (recently, Anil *et al.*, Solid-State Electron **47**, 995 (2003))
- Mobility degradation seen experimentally (Toshiba, Udine, Lucent, recently, Lime *et al.*)
- If true, ballistic transport is unattainable



Lime *et al.*, Solid-State Electron **47**, 1147 (2003)

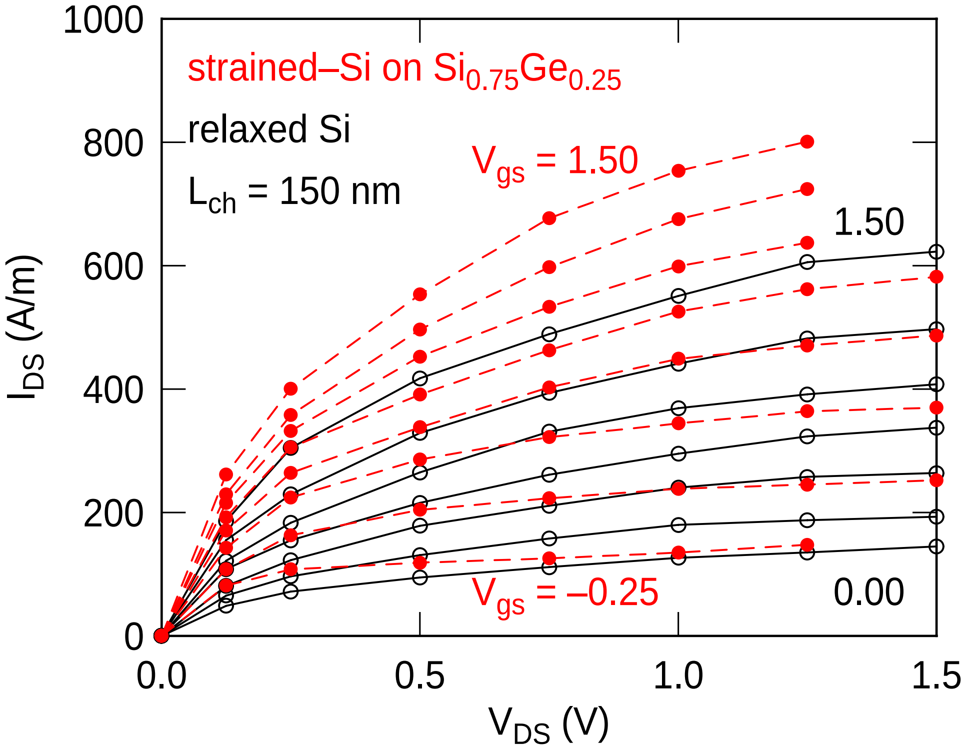
Recent drag experiments (Solomon) inconsistent with mobility-degradation

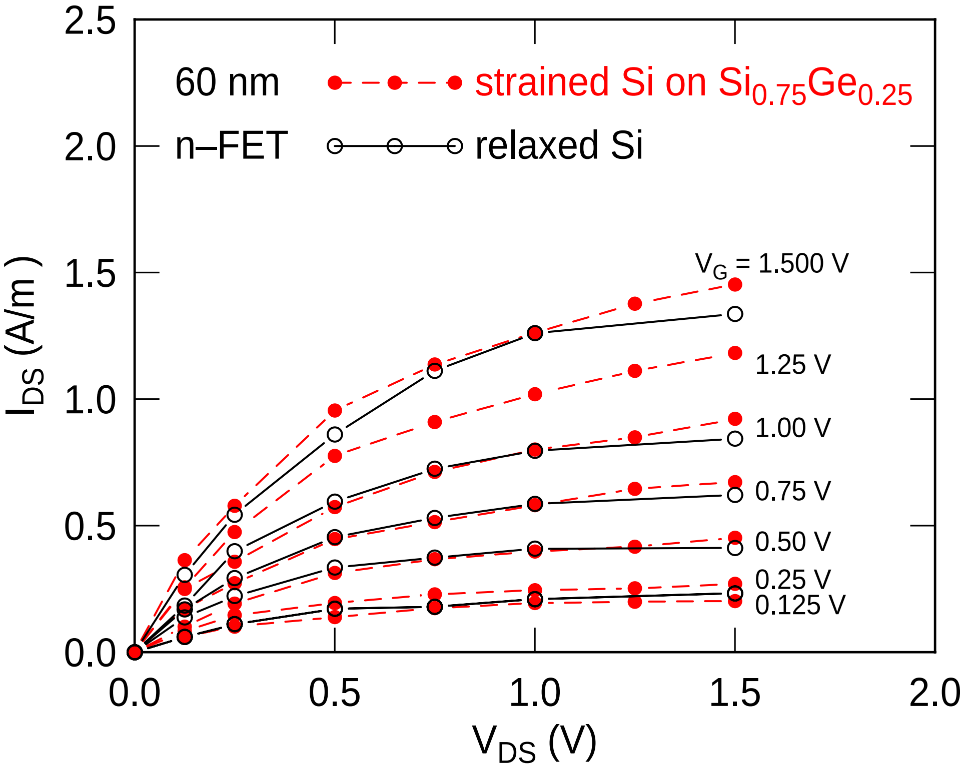
Present ‘revival’ of MC simulations

- ‘New scaling’ forces us to look for ‘revolutionary’ (as opposite to ‘evolutionary’) paths
- Too many and to expensive to try them all:
  - Strained-Si devices
  - Double-gate devices and (electrostatic) scaling limits
  - Ge MOSFETs
  - III-V compound semiconductors and the ballistic limit
- Monte Carlo (as bearer of Physics) to the rescue

Strained Si devices

- Not quite what promised by the mobility-boost, but still an advantage to SS...
- Recent DAMOCLES simulations (Kumar) show $\approx$ 30% I_{on}-boost persists down to 20 nm (Cai *et al.*, IEDM 2004)

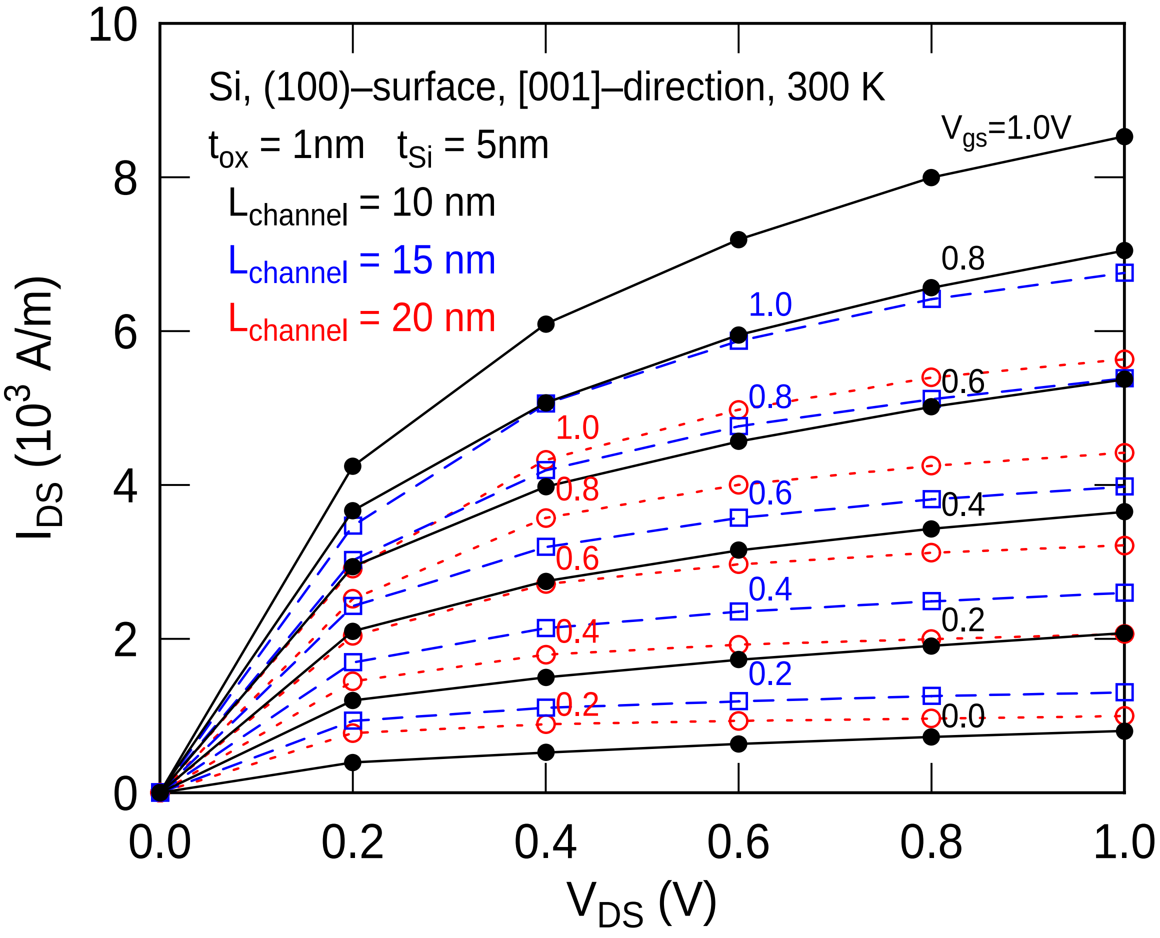




Bulk devices

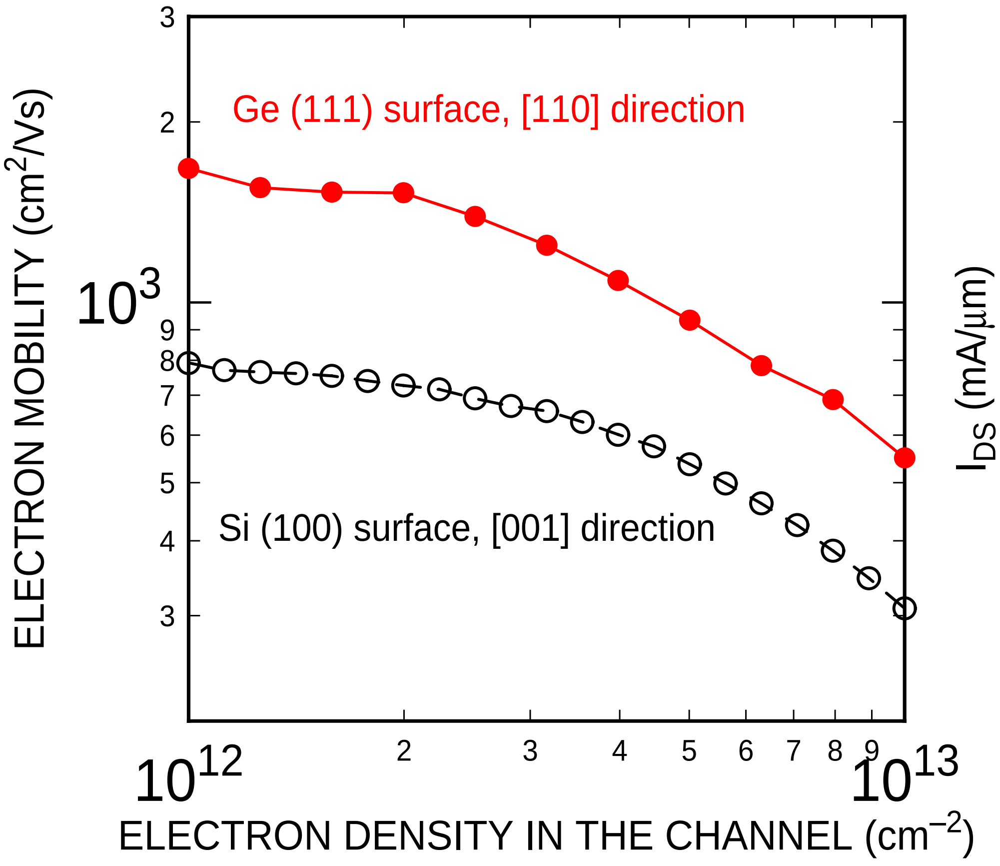
Double-gate devices: Scaling Si

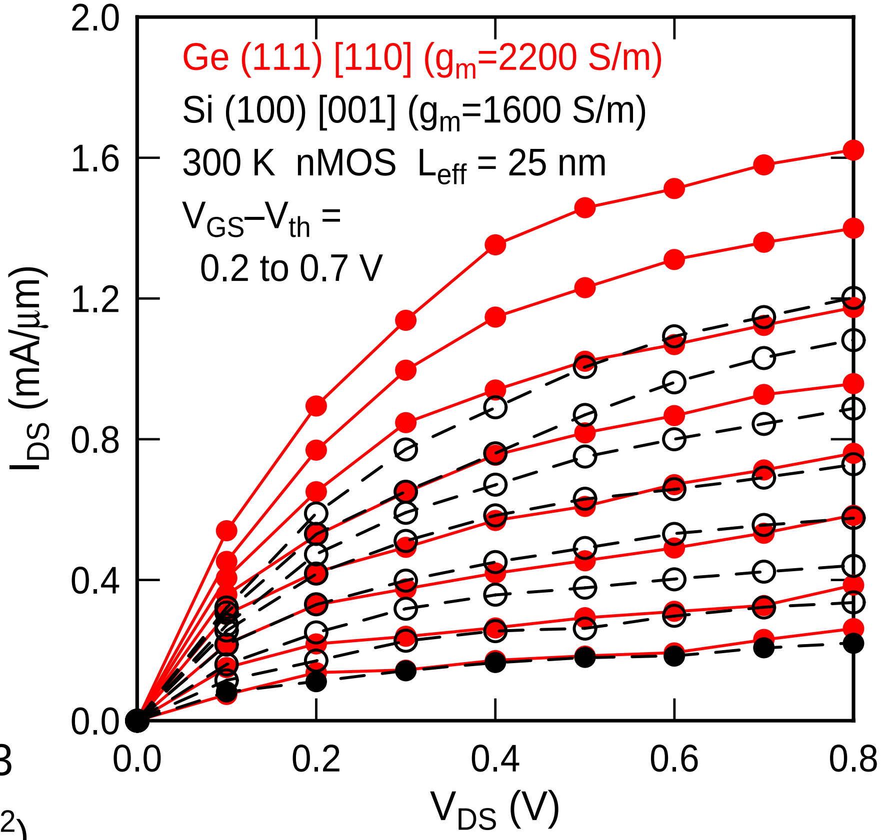
- Good electrostatic behavior down to 10 nm
- Surface-roughness/transport in thin Si are issues
- Quantum effects a concern (hard to model as well...)



Ge-based FETs

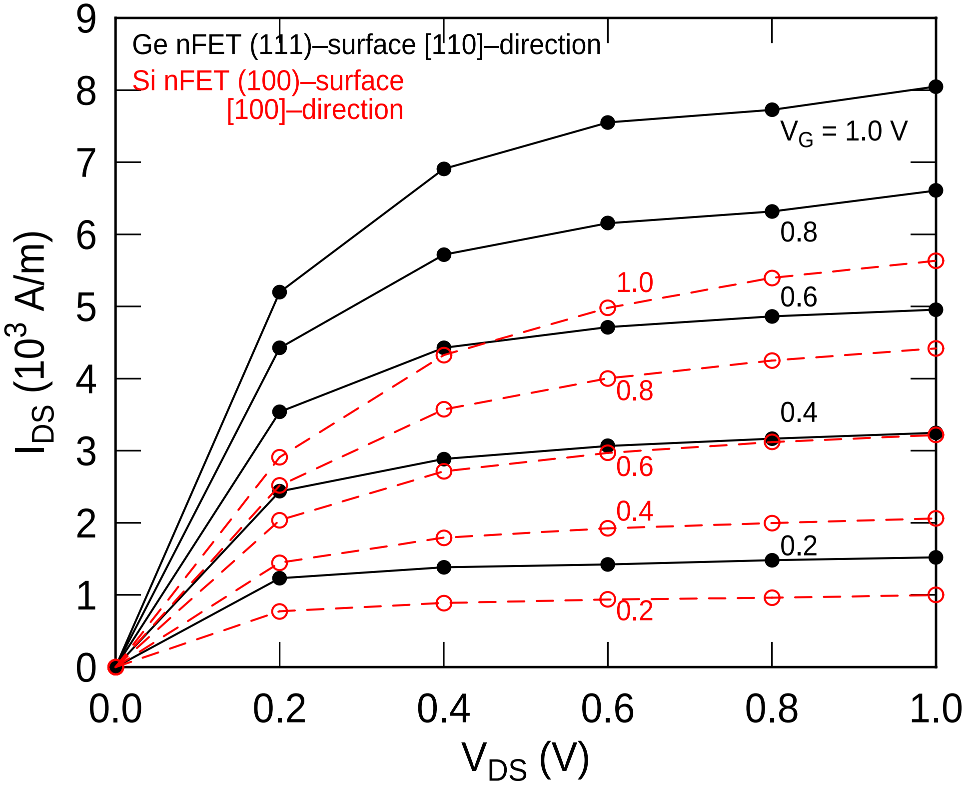
- Not quite what promised by the mobility-boost, but still an advantage to (111) Ge...
- Leakage due to band-to-band tunneling a potentially lethal problem

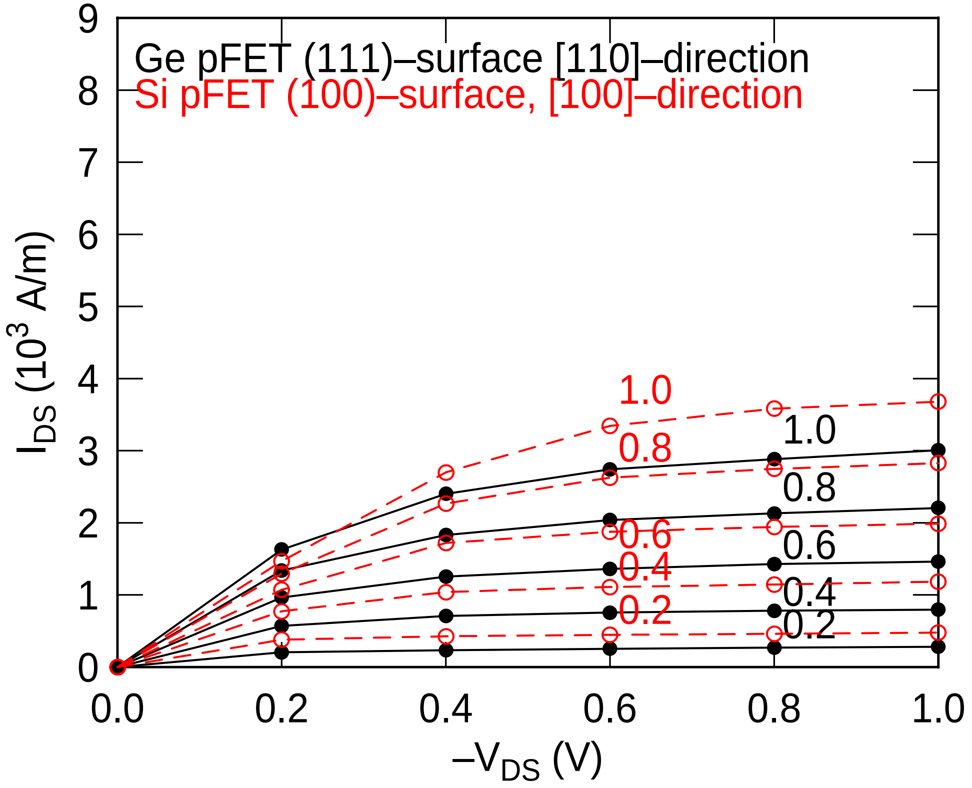




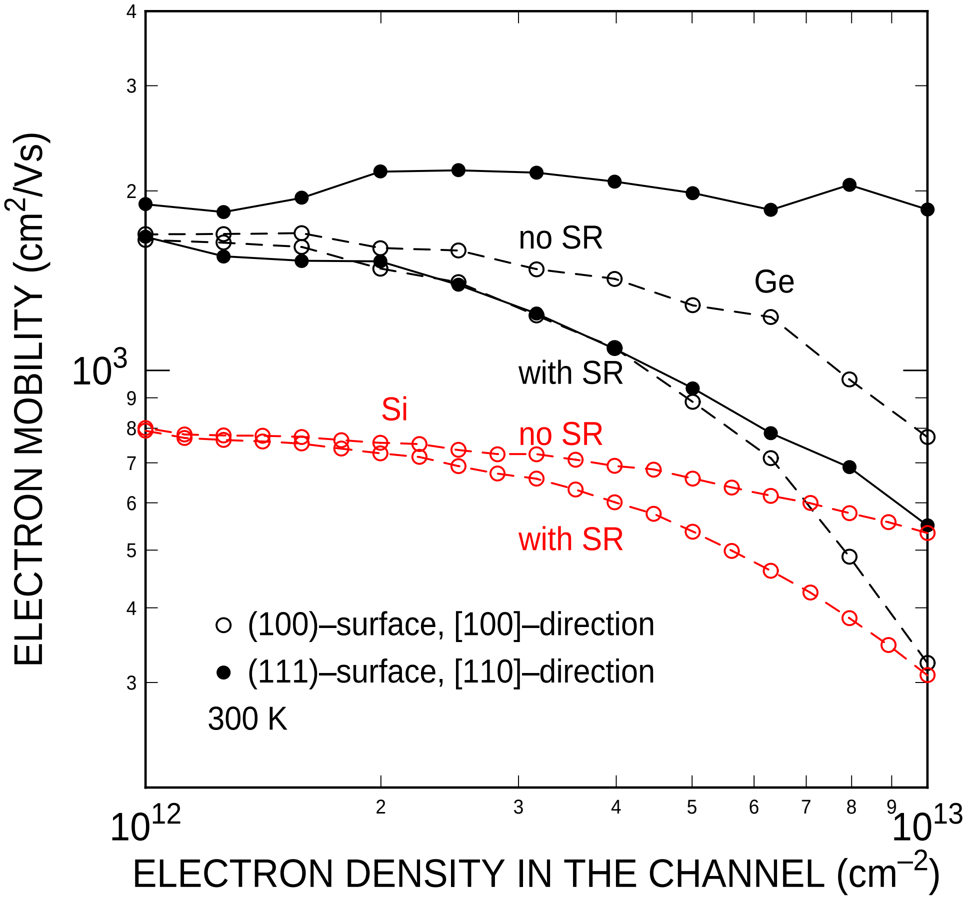
Double-gate devices: Si *vs.* Ge

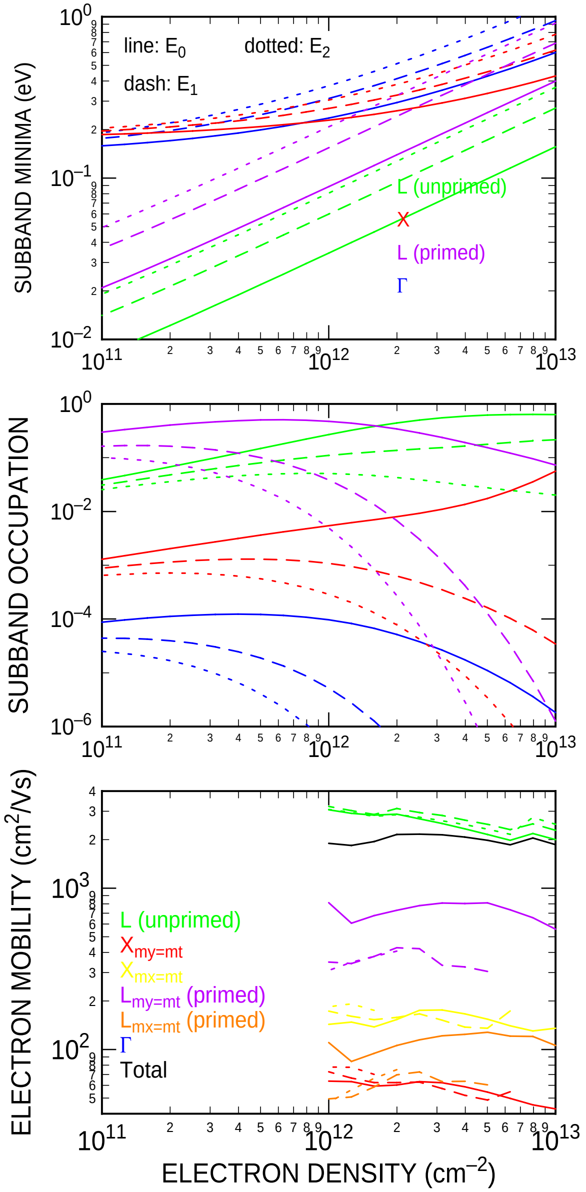
- Not quite what promised by the mobility-boost, but still an advantage to (111) Ge...





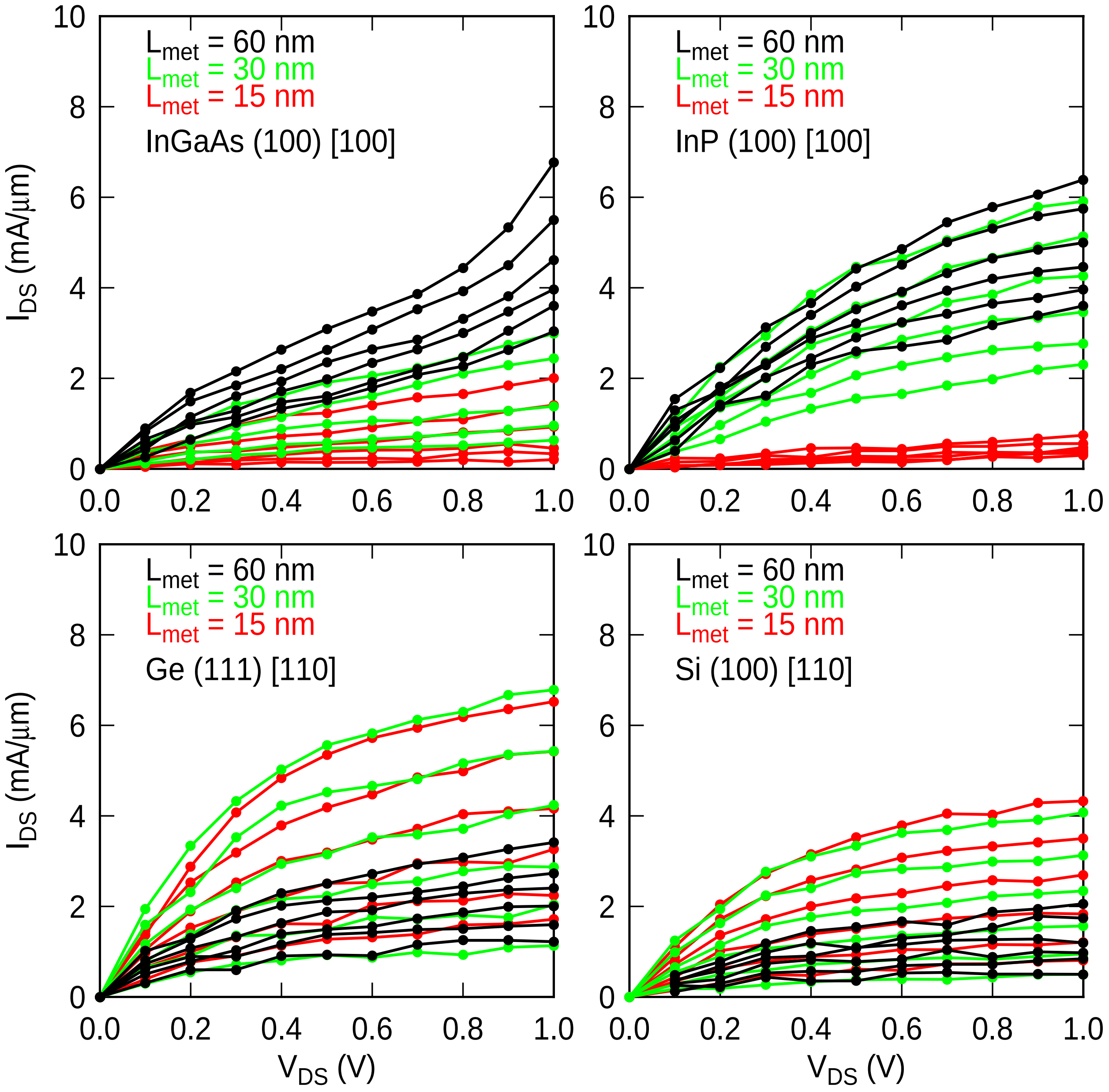
Effective mobility in nFET: Si *vs.* (111) Ge



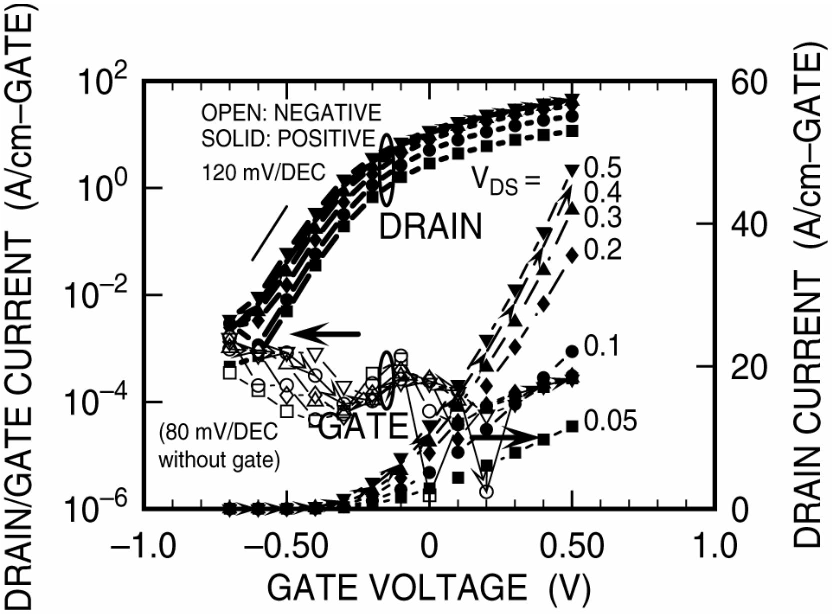


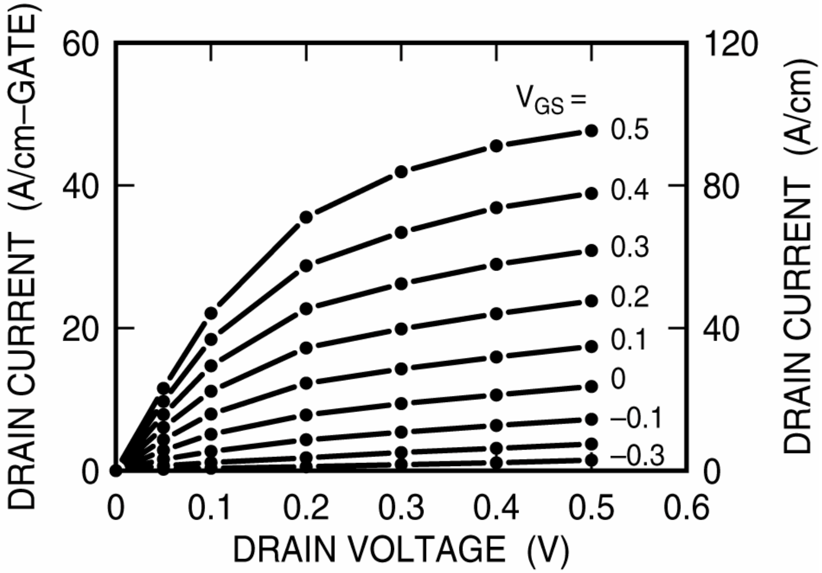
‘Fast’ materials and ballistic transport

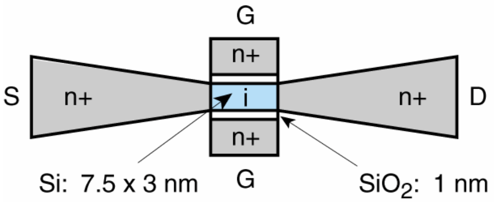
- Simulated small nFETs on various: materials: As in 1991, but shorter
- III-V semiconductors ‘choke’
- Confirmed by ballistic 2D quantum simulations (QDAME)

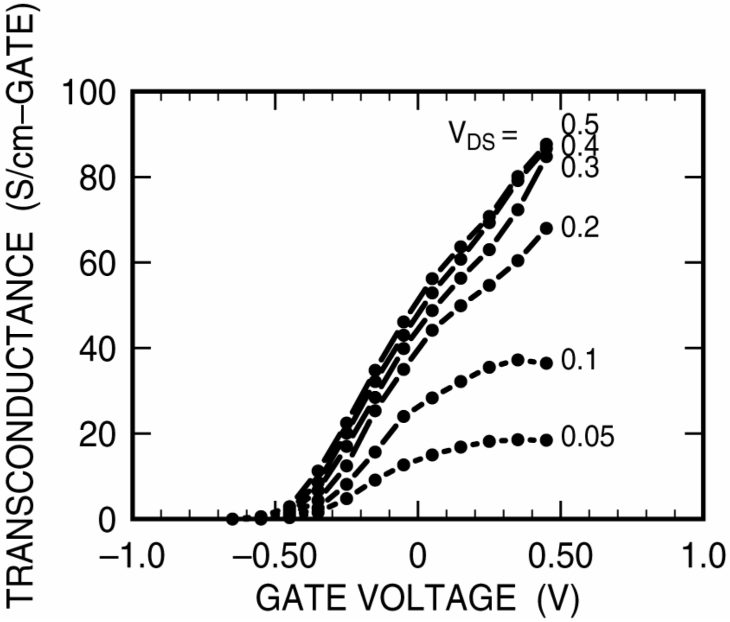


Si 7.5 nm DGFET with QDAME (DGFET)



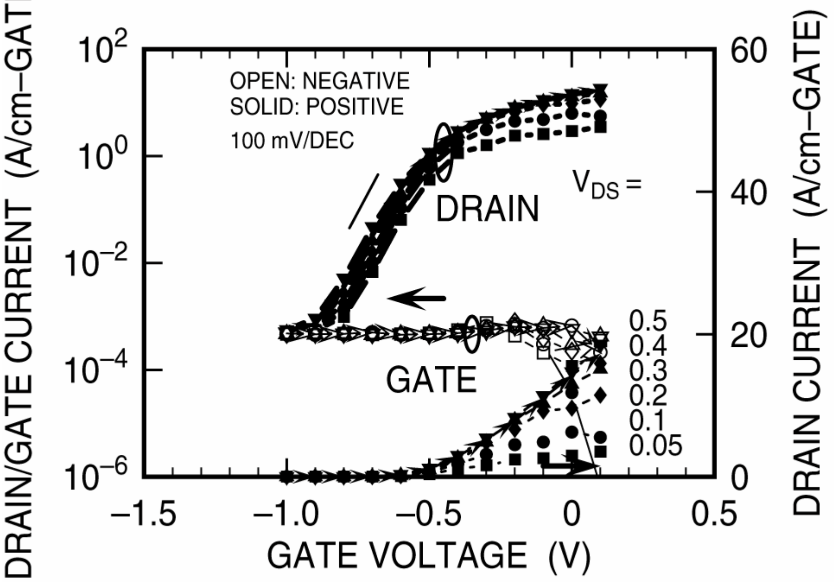


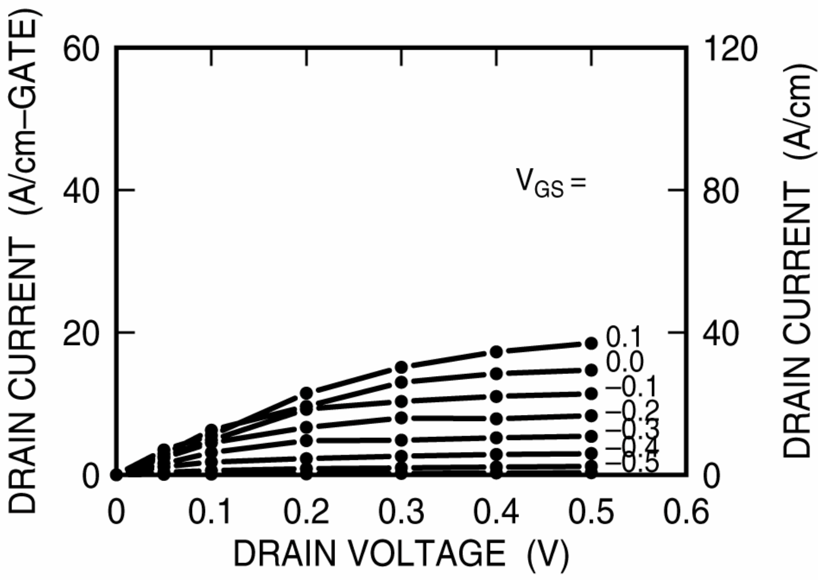


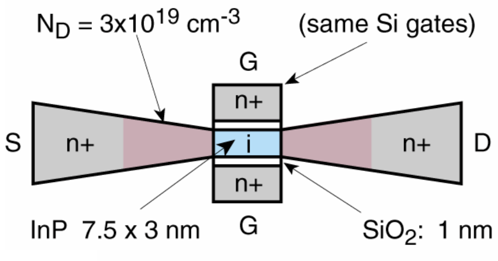


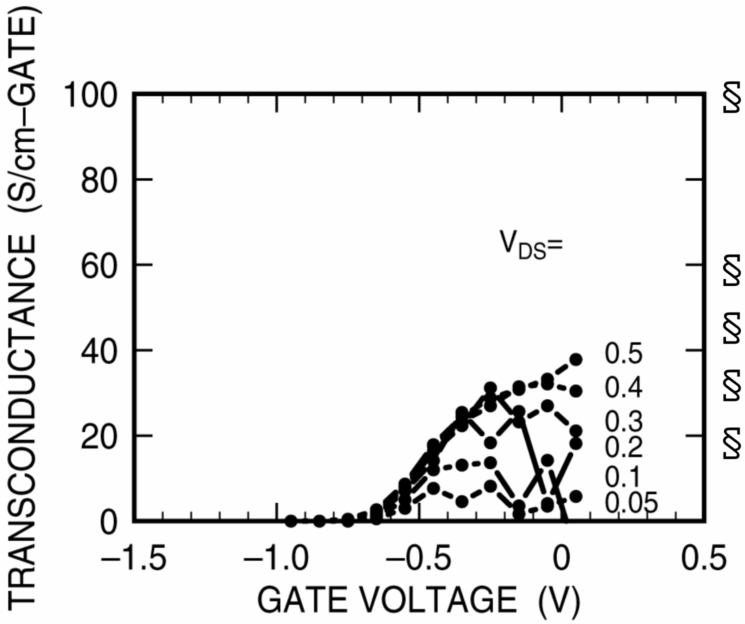
- gate current undermines SS; could have been 80 mV/DEC
- TINV = 1.87 nm
- CV/I = 0.105 psec
- $V_{dd} = 0.5$ V
- $V_T = -0.30\text{ V} \xrightarrow{\phi_{ms}} +0.15\text{ V}$

InP 7.5 nm DGFET with QDAME (DGFET-InP)









- poor current drive, but also reduced capacitance (DOS) yields same (!) CV/I
- TINV = 2.82 nm
- CV/I = 0.100 psec
- $V_{dd} = 0.5$ V
- $V_T = -0.55\text{ V} \xrightarrow{\phi_{ms}} +0.15\text{ V}$

‘Fast’ materials and ballistic transport

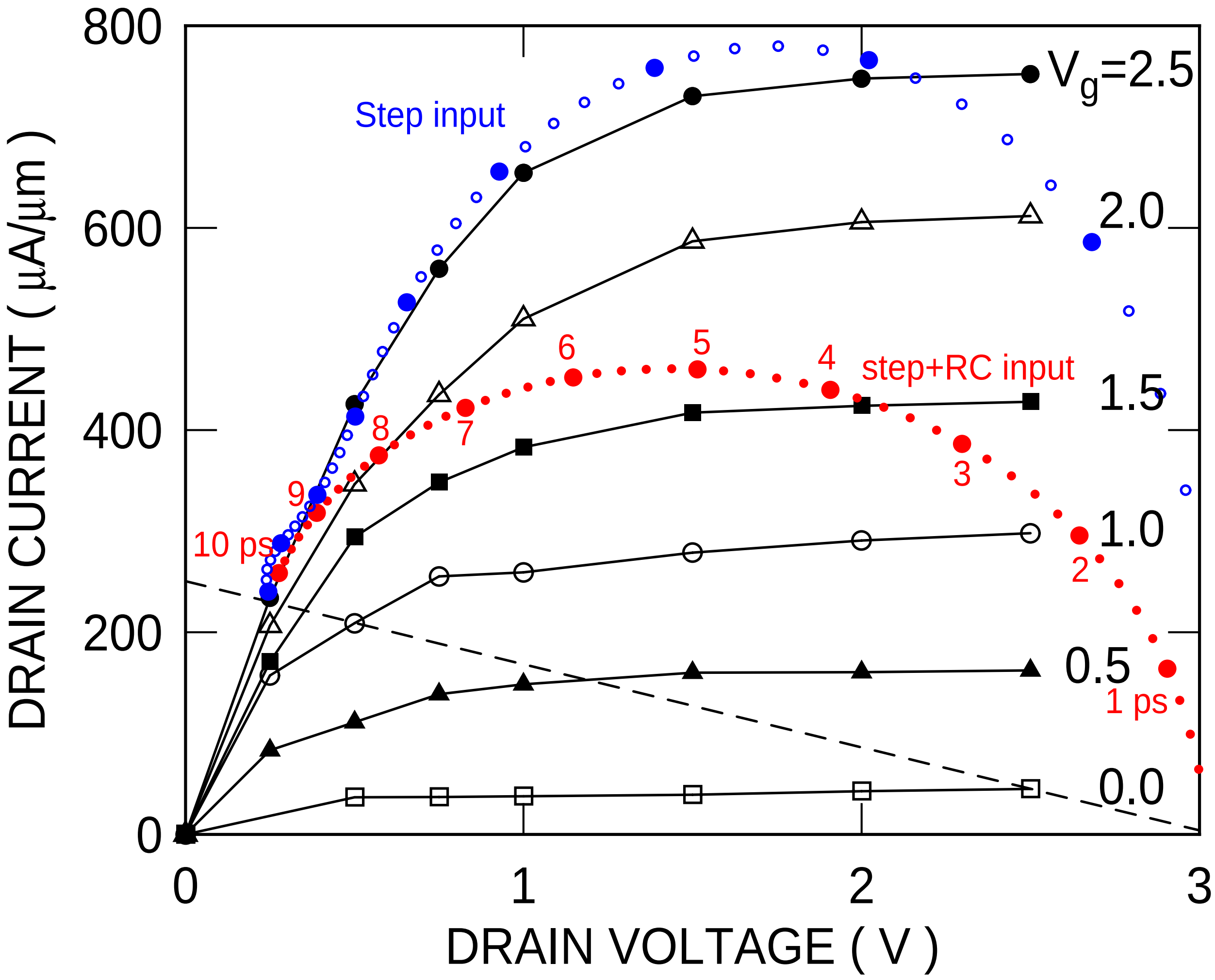
- Scattering-dominated regime:
  - Need small conductivity mass (large velocity)
  - Need small DOS mass (weak scattering)
- Approaching the ballistic limit:
  - Need small conductivity mass (large velocity)
  - Need large DOS mass (many conduction channels)
  - Ge – (110) better than (111) in the ballistic limit – a good compromise:
    - small conductivity masses
    - many quasi-degenerate valleys to boost DOS mass
    - if only the small gap weren’t a problem!
- Self-consistency transport-Poisson of utmost importance!

Basic questions on semiclassical transport

- Is ballistic transport achievable?
  Probably not: Coulomb interactions always present. Maybe gate-screening could help...
- Does the low-field mobility matter?
  Probably not: In small devices a large DOS mass may help.

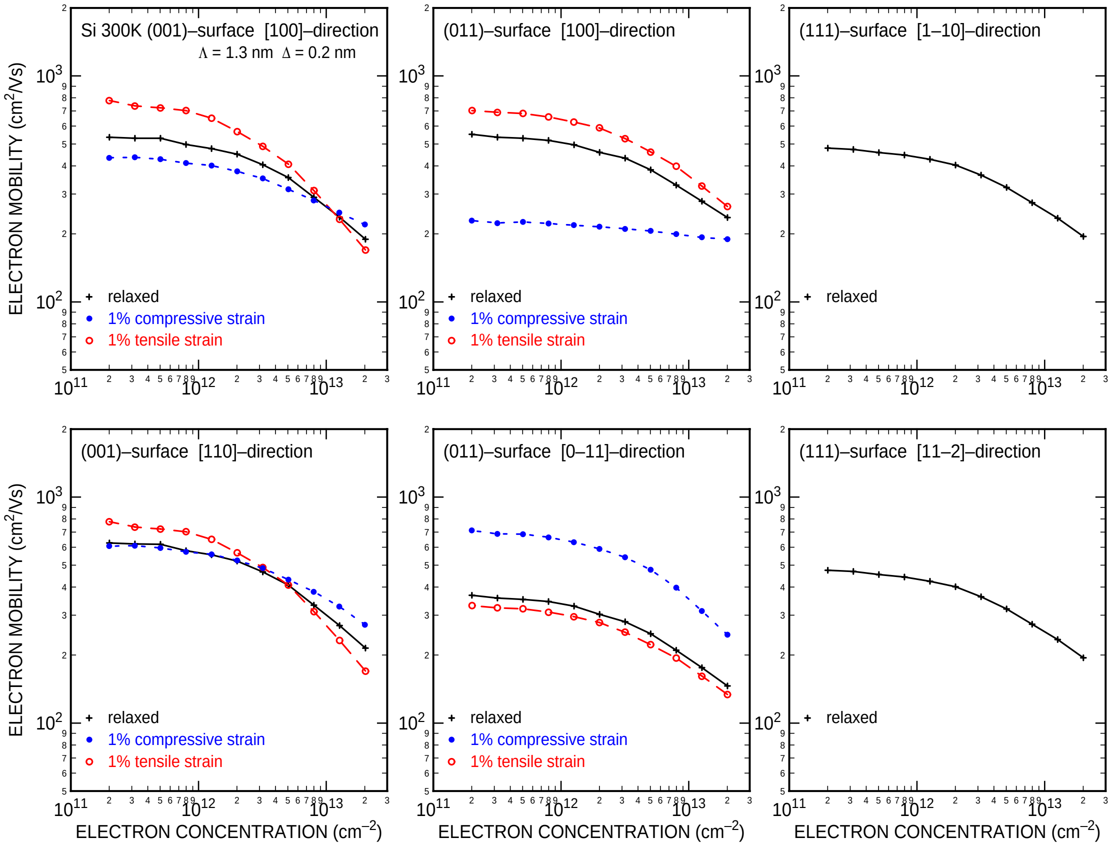
I_{on} is not the whole story, of course...

- Low-field mobility also determines switching speed
- Both I_{on} (or g_m) and μ depend on scattering: Correlated *only* when scattering-dominated



A ‘sad’ conclusion: Should we trust theory?

- A depressing example: We cannot explain the mobility-boost in biaxially stressed (tensile) Si nFETs
- Even more depressing: Nobody cares!
  We (*i.e.*, the system?) reward activities on ‘record breaking’ devices
  We discourage ‘thinking’ and basic experimentation... No time left to ‘think’



Outline/Conclusions

- The 'early days (*i.e.*, when we used to 'think'):
  Science, not much Technology
  - The basics of 'warm electron' transport: The Modena 'standard model'
  - The (oversold?) challenge of 'hot carriers': The 'new standard model'
  - Coulomb interactions
  - Technology? Just calibration of moments methods...
- The future days of the 'end of scaling' (*i.e.*, compute-and-do-not-think):
  Technology, not much Science
  - A little bit of Science: More Coulomb interactions
  - New devices (PD, FD and UTB SOI; Double-gate FETs, ...)
  - New materials (strained Si, Ge, III-Vs,...)
  - Old materials from a new angle ('new' crystal orientation)
- Basic (scientific?) questions at the end of the road:
  - Is ballistic transport a 'pipedream'?
  - Is the low-field mobility meaningful?
- Quantum transport: Science or fashion? Not for me to address...