Physical limits of high-speed SiGe HBT performance

M. Schroter[1,2], G. Wedel[1], J. Krause[1], B. Heinemann[3], C. Jungemann[4], P. Chevalier[5], A. Chantre[5], N. Rinaldi[6]

[1]Technical University Dresden, Chair for Electron Devces & Integrated Circuits, 01062 Dresden, Germany
[2]University of California San Diego, ECE Dept., La Jolla, USA
[3]IHP, Frankfurt (Oder), Germany
[4]RWTH Aachen University, Aachen, Germany
[5]ST Microelectronics, Crolles, France
[6]University of Naples, Naples, Italy

Bordeaux

June 28/29, 2011

OUTLINE

- Introduction
- Methodology overview
- 1D scaling analysis
- Compact model
- 3D scaling analysis
- Electrothermal considerations
- Conclusions

Introduction

why SiGe BiCMOS & HBTs

- future mm-wave applications (e.g. sensors and imaging, communcations) require transistors operating frequencies beyond 500GHz
  => out of reach for CMOS regarding output power, power gain, impedances, analog characteristics

however ...

- Conventional technologies appear to approach their physical limits
  => general interest as to the "mileage left" for this technology
  => explore physical limits of SiGe HBTs

Goals

- provide information on margins left for SiGe HBT technology
- knowing the physical is important for creating a roadmap
  => important for product planning

Introduction

- early predictions of limits
  - Johnson limit: 200 GHz·V for Si BJTs, far exceeded by existing SiGe HBTs
  - eighties: (f_T, f_{max}) = (17,10)GHz for BJTs (0.4μm E width)

Technology predictions are ***very*** hard!

- requires understanding of physical effects occurring in the future
  => need good physical models and simulation tools
- make judgement calls regarding ***practical*** limitations
- alternatives: basically none (or do nothing)

=> use as conservative assumptions as possible

Methodology overview

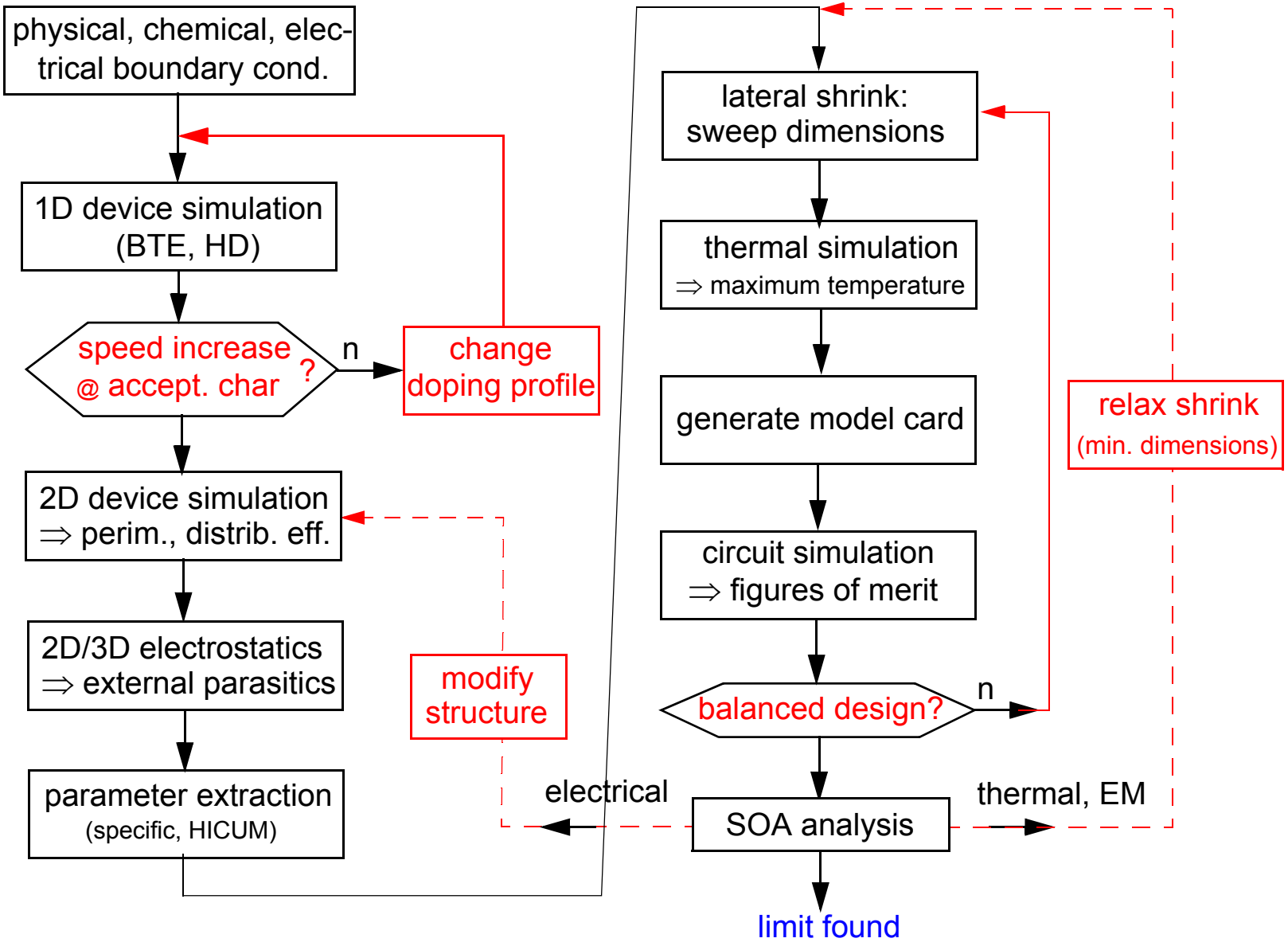
existing approaches

- f_T, f_{max} estimate based on analytical (textbook) equations for most relevant time constants => do not capture physical effects in advanced structures (transport, breakdown, tunneling, 2D/3D, ...)
- mostly ignore parasitics (series resistances, BE spacer capacitance, metallization)
- device simulation using non-calibrated HD models => predictions too optimistic

... vs. this approach

- most advanced and reliable transport models: BTE, calibrated HD
- Schroedinger-Poisson (BU) for tunneling currents (& evaluation of class. models)
- 2D effects (perimeter injection and charge, current spreading) from device simulation
- parasitic effects of BE spacer, metallization from electrostatic simulation
- series resistances from sheet resistances, estimated specific contact resistances
- HICUM parameter extraction and generation for realistic device structure
- circuit simulation for obtaining figures of merit

Flowchart


physical, chemical, elec-
trical boundary cond.
1D device simulation
(BTE, HD)
speed increase
@ accept. char
?
n
change
doping profile
2D device simulation
⇒ perim., distrib. eff.
2D/3D electrostatics
⇒ external parasitics
modify
structure
parameter extraction
(specific, HICUM)
lateral shrink:
sweep dimensions
thermal simulation
⇒ maximum temperature
generate model card
circuit simulation
⇒ figures of merit
balanced design?
n
relax shrink
(min. dimensions)
electrical
SOA analysis
thermal, EM
limit found

Device simulation tools

- Boltzmann transport equation (BTE) solution
  - MC, SHE
  - full band analysis (includes advanced non-parabolic full band fit, II)
  - issue: too slow for profile optimization (& generating characteristics for parameter extraction)

- Hydrodynamic (HD)/energy transport (ET) simulation
  - moments of BTE: energy balance and flux in addition to DD
  - careful calibration of additional parameters (relaxation time, fudge factors!)
  - issues:
    - predictive capability limited to 1 process generation => need to re-calibrate
    - cannot handle "too" steep profiles (e.g. Ge, heterojunctions)

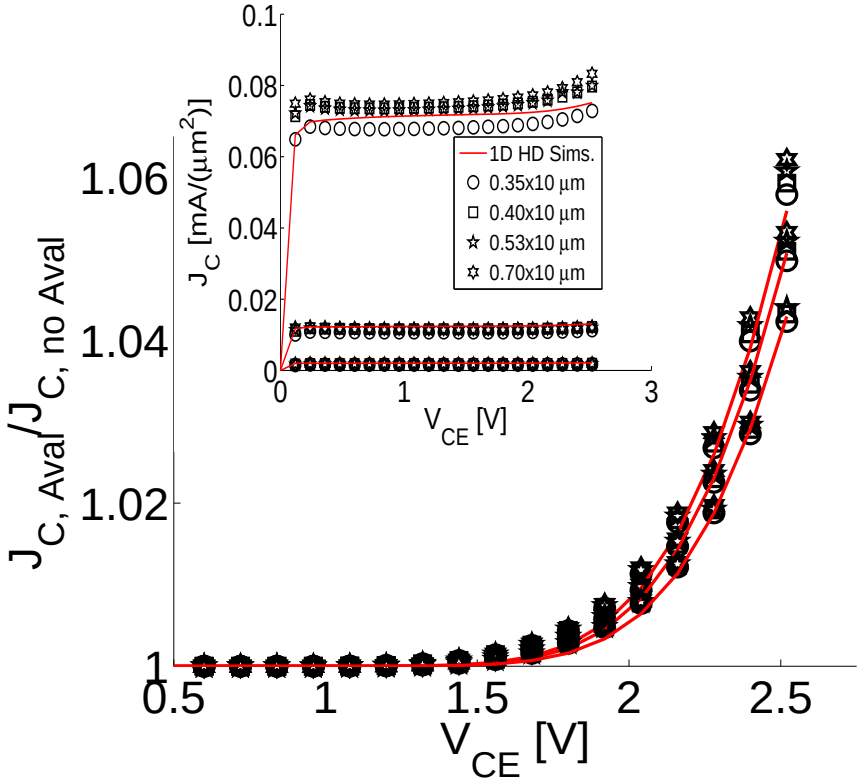
- Calibration
  - examples see next slide

Calibration

selected examples (for details see BCTM 2010 paper)

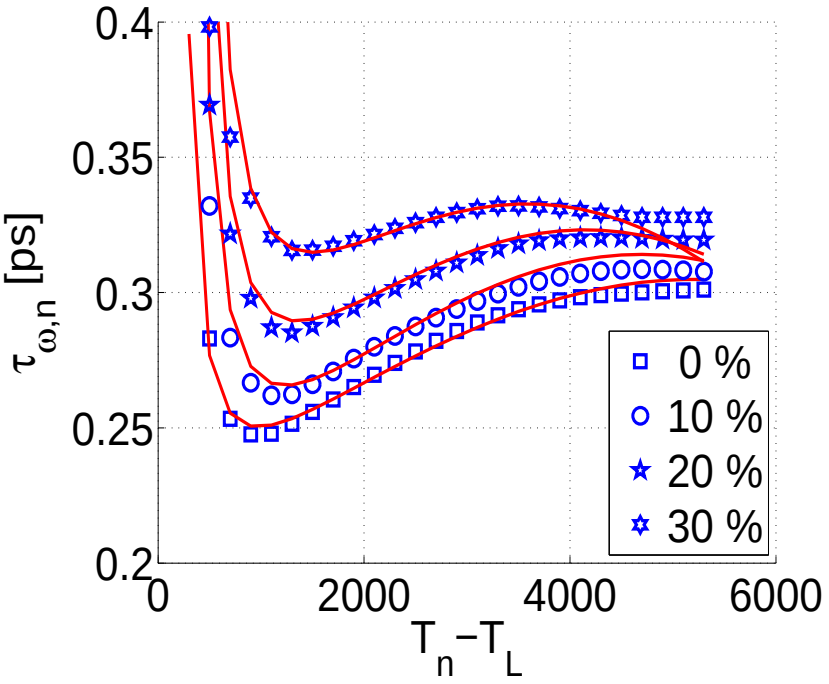
impact ionization

HD vs experimental results



relaxation time

HD vs BTE



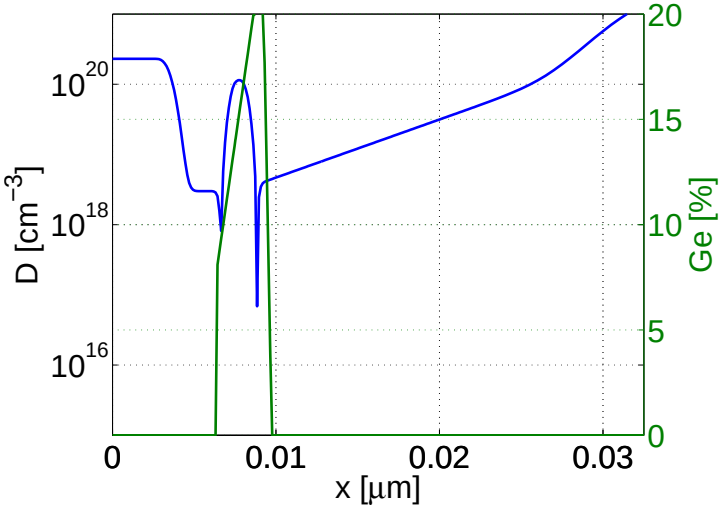
=> valid for process node and subsequent generation only

1D scaling analysis

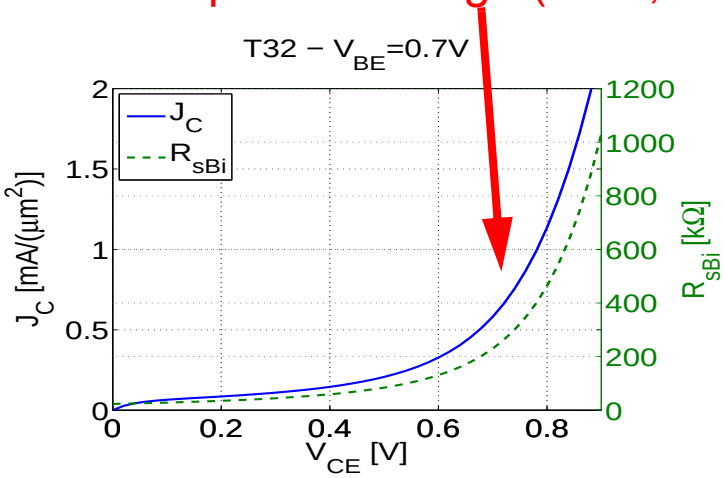
- why 1D loop separately?
  - ultimate limits are expected to be determined by "1D" effects
- known constraints/boundary conditions
  - max doping (B,E, bl)
  - base width => punch-through
  - BE and BC doping => tunneling
  - sufficiently short E width => q.s. charge reduction
- figures of merit
  - f_T (directly), f_{max} with assumed E width of 30nm)
  - $I_C(V_{CE})$ curve shape (avoid negative output conductance)
- scaling steps
  - additionally investigated physical effects: BTB and TA tunneling
    Notes: - SP solution yields lower tunneling current
    - tunneling models to be verified by experimental data
  - investigated structural variations: E/B/C width and doping (incl. vertical spacers)
  - evaluated roughly ***100*** different 1D profiles
  - examples see next slides

Intermediate results and issues

straight scaling of 500GHz profile

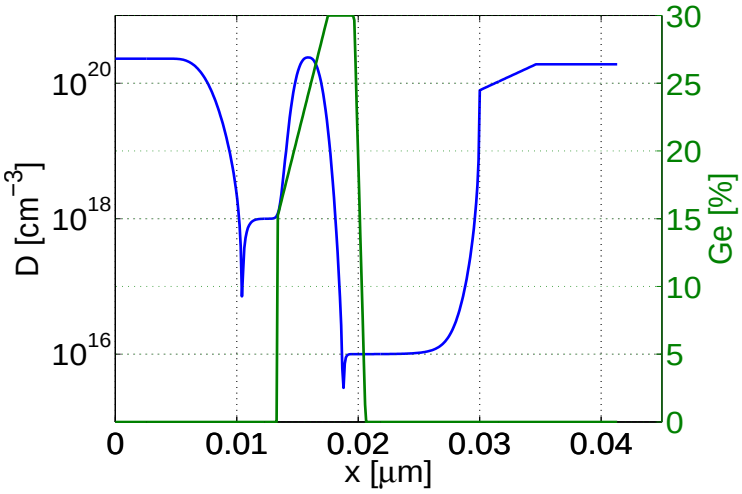


=> base punch-through (vs. II, tun.)

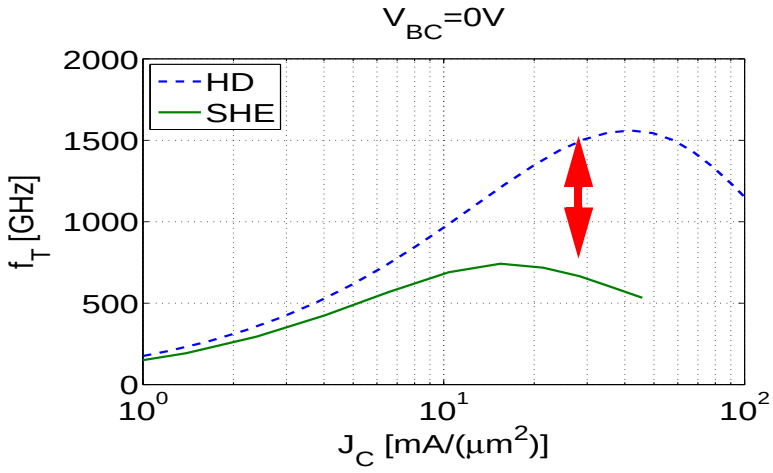


=> optimize profile shape(s)

optimized profile & performance. verific.

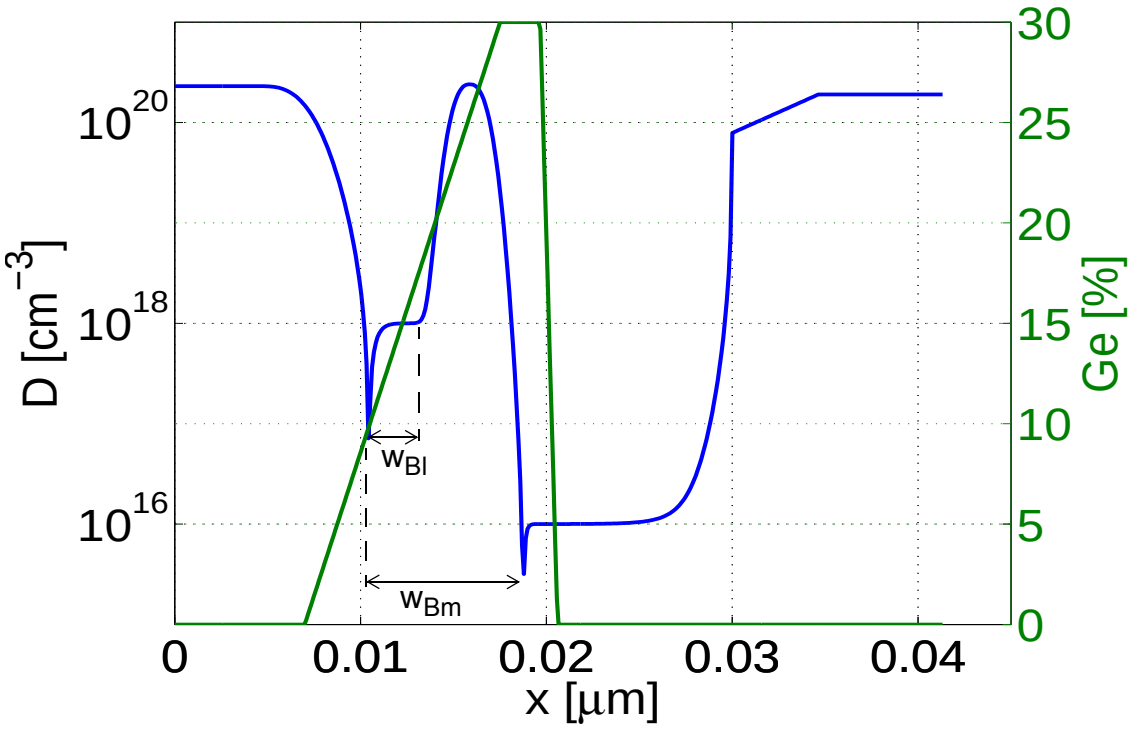


=> HD simulation issues



=> BTE solution used as reference

Final profile



important data:
x_{jE} = 10.5 nm
x_{jC} = 18.8 nm
=> w_{Bm} = 8.3 nm
w_{EI} = 3.3 nm => w_{BE}
w_{Eh} = 10. 3 nm
w_{Ci} = 13.3 nm

peak f_T = 1385 GHz
$J_C(f_{T,peak})$ = 63 mA/μm^2

- lower doped E => reduce tunneling impact on forward characteristics
- base width not at minimum (slight reduction still possible)
- C width shorter than II length => avalanche current does not increase anymore
- lower doped C => reduce TAT; little f_T change if increased to 10^{18} cm^{-3}
- graded Ge through base and "lightly" doped E

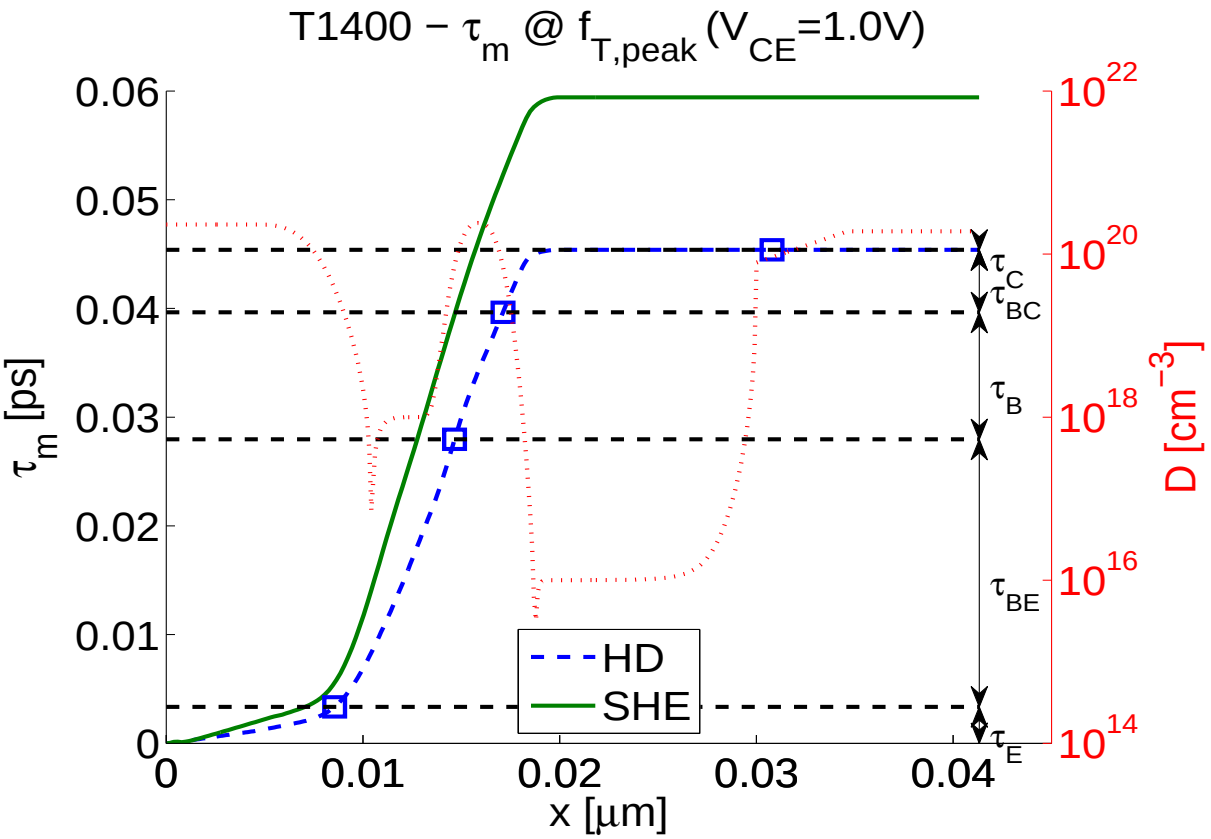
Decomposition of regional storage times

1D profile optimization

accumulated storage time:

$$\tau(x) = q\int_0^x \frac{dm}{dJ_C}\bigg|_{V_{CE}} dx$$

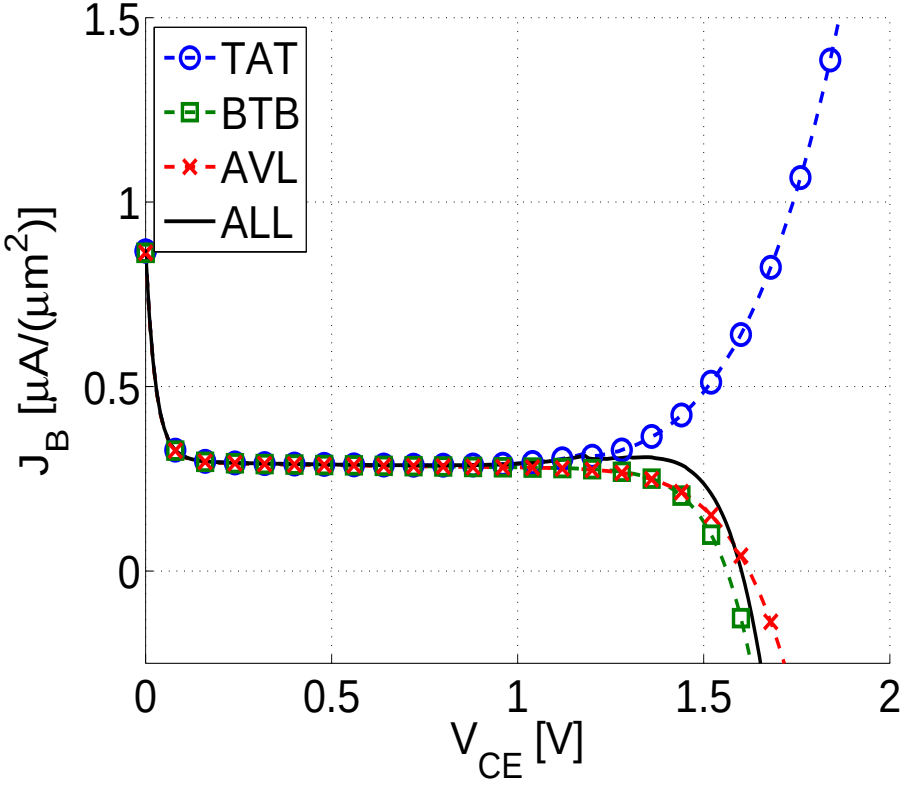
m: minority carrier density



BE region causes largest contribution
(regardless of lightly doped p or n spacer)

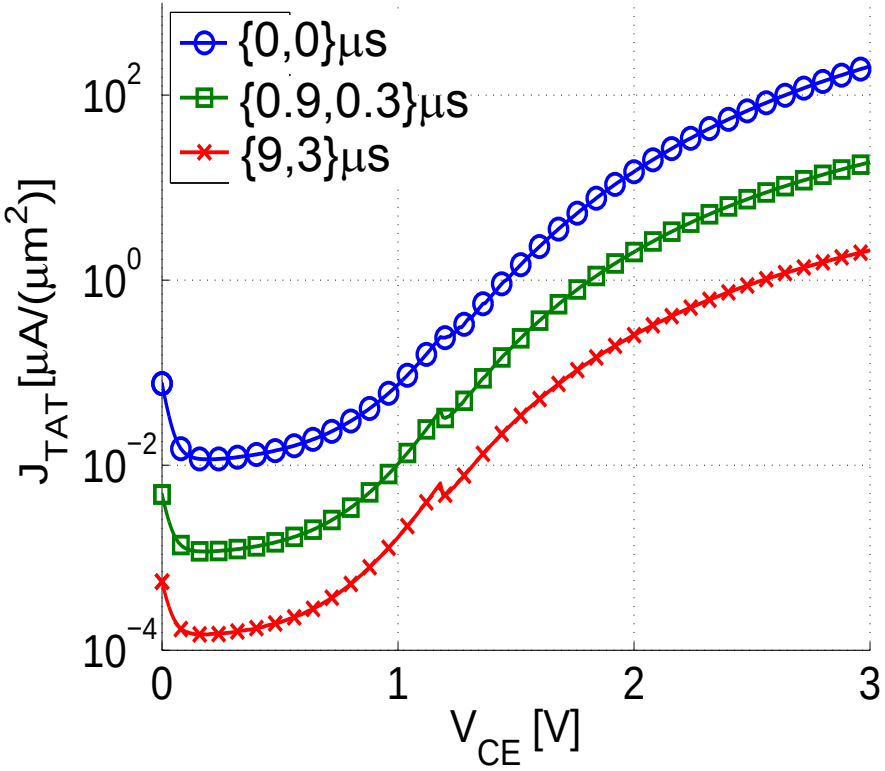
Collector breakdown mechanisms

base current components



compensation of BTB, AVL by TAT

TAT component

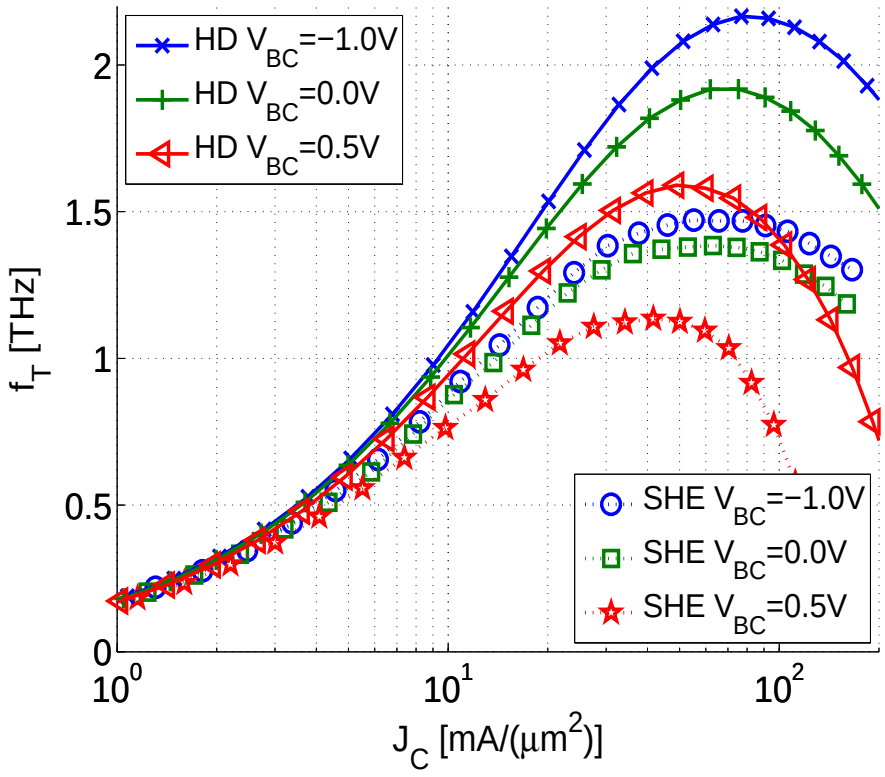


variation of carrier lifetime
uncertainty due to lack of exp. data

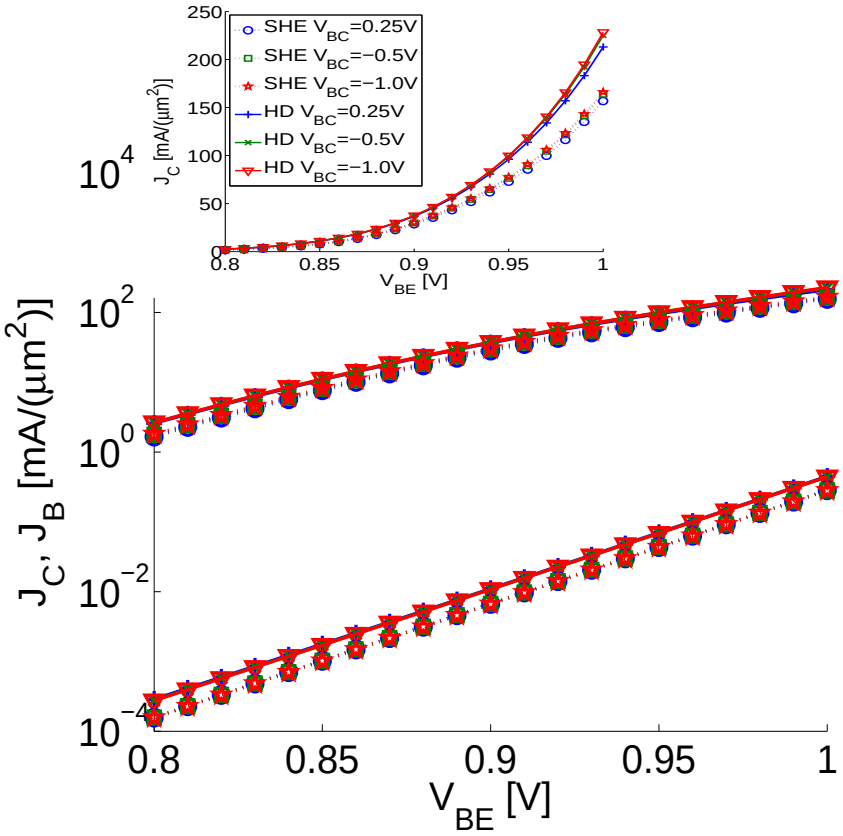
=> BTB tunneling likely to become dominant mechanism

1D electrical characteristics (final profile)

transit frequency



Gummel characteristics



- reducing w_{El} to zero (=> conventional emitter doping profile)
  - 10% higher peak f_T, but lower f_T at low J_C and higher tunneling current impact at low V_{BE}

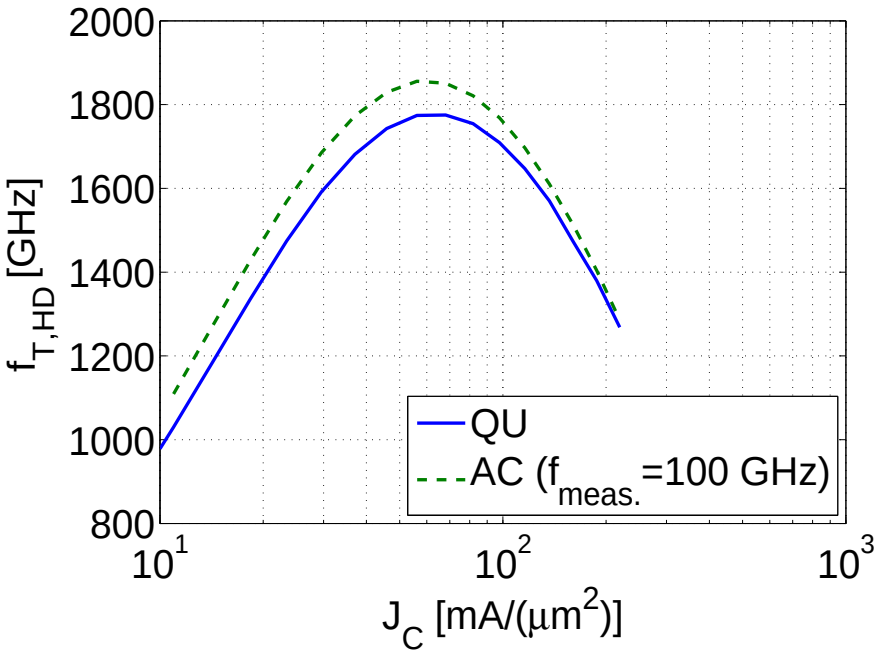
=> f_T = 1.5 THz appears to be (roughly) the ***isothermal*** limit

Summary of 1D process parameters

parameters	initial profile	ultimate limit
N_{Bmax} (cm^{-3})	$6\ 10^{19}$	$2.4\ 10^{20}$
w_{Bm} (nm)	9	8.3
w_{Bl} (nm)	0	3.3
w_{Ci} (nm)	58	13.3
f_T (THz) @ V_{BC} = -1V	0.46	1.47 (BTE)
J_C (mA/μm^2) @ peak f_T	13	65 (BTE)
BV_{CEO} (V) @ V_{BE} = 0.7V	1.37	1.4 (HD)
R_{SBi0} (Ω/sq)	6100	2770
C_{jEi0} (fF/μm^2)	7.8	14.1
C_{jCi0} (fF/μm^2)	4.3	7.9

Verification of f_T determination

- usually obtained from quasi-static (QU) method => allows regional analysis
- for long neutral regions (cf. IMEC emitter profile) => non-QS effects
  => QU method yields much lower transit frequency
- correct values obtained from applying measurement (AC) method (extrap. from |ß|)

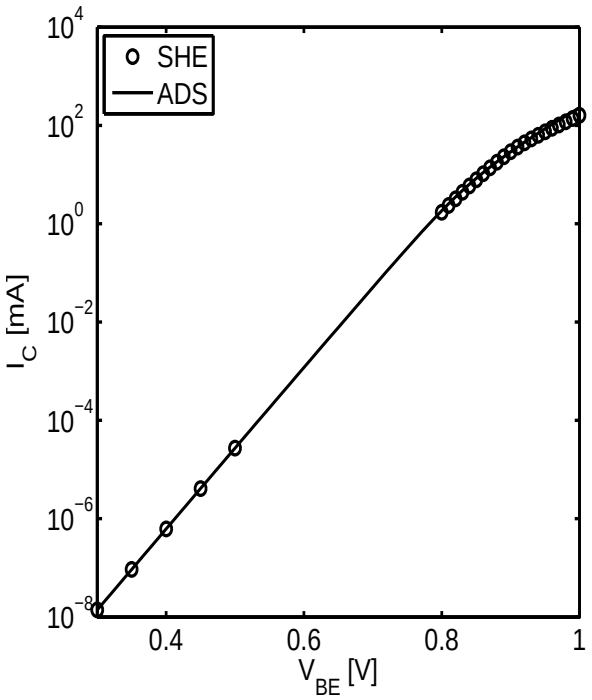


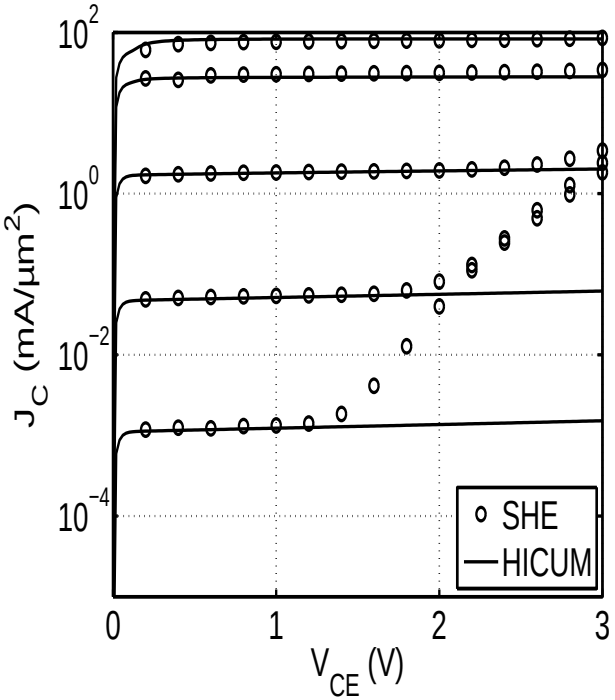
- AC method not available for BTE solution
- used HD simulation (as "proof of concept")
- QU method slightly underestimates f_T at low and medium current densities for the structure(s) found
- longer E region still yields similar f_T (from AC method) as proposed structure

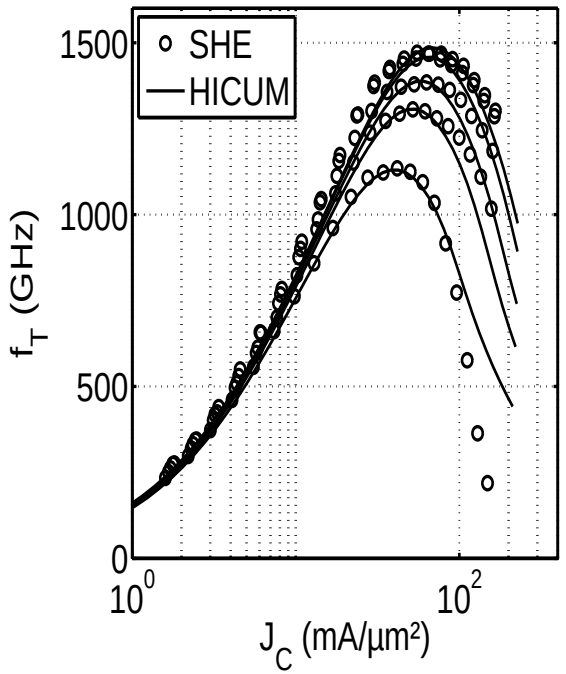
=> optimization result using QU method yields correct f_T

Compact model

- HICUM v2.3 (!!)
- parameter extraction (as physics-based as possible) => 1D results





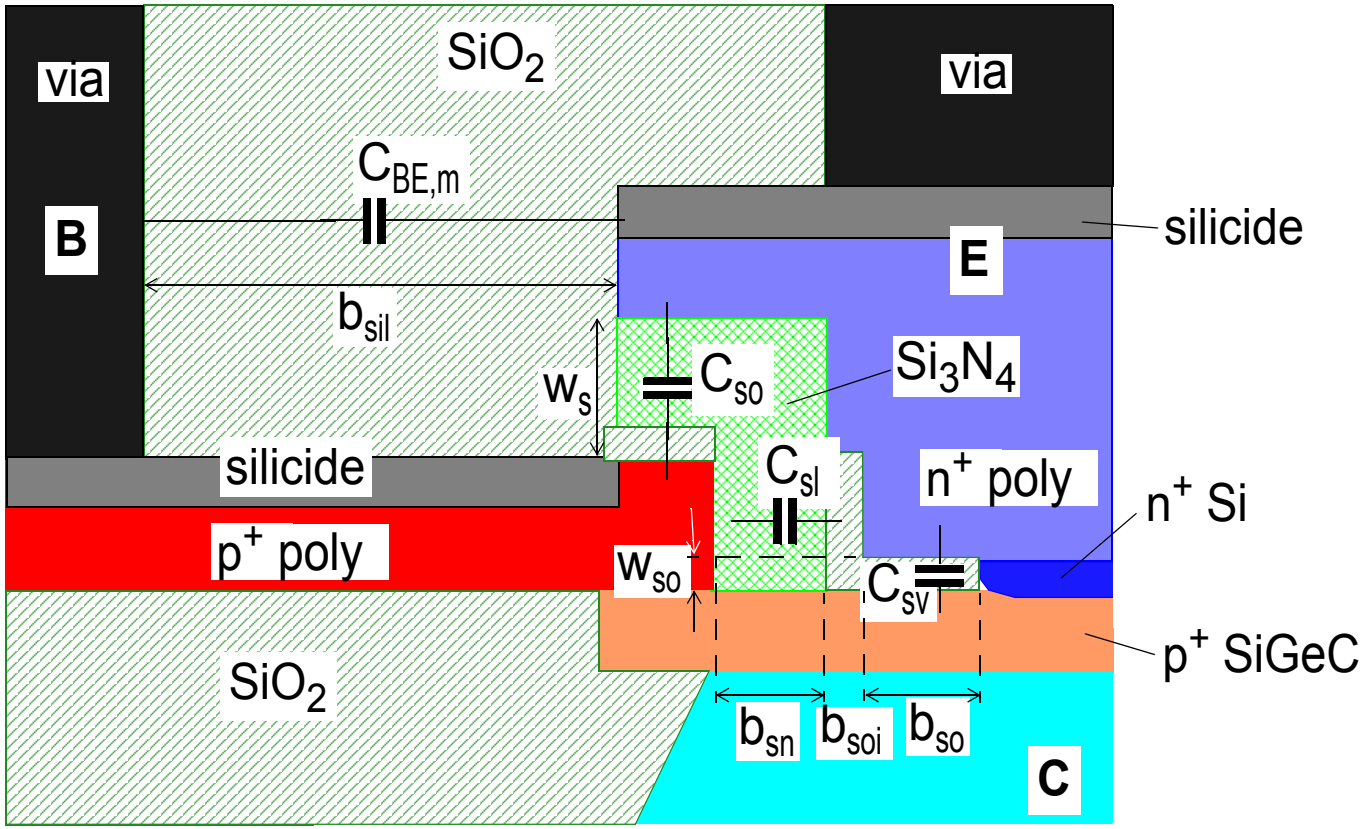


=> excellent accuracy over relevant bias range

=> suitable as basis for 2D/3D simulations of figures of merit

2D/3D effects and parasitics

- junction perimeter to area currents and capacitances via ratio parameter
- BE spacer and contact metallization capacitance



b_{sil} = [50...250]nm

w_s = [5...25]nm

b_{sn} = [10...20]nm

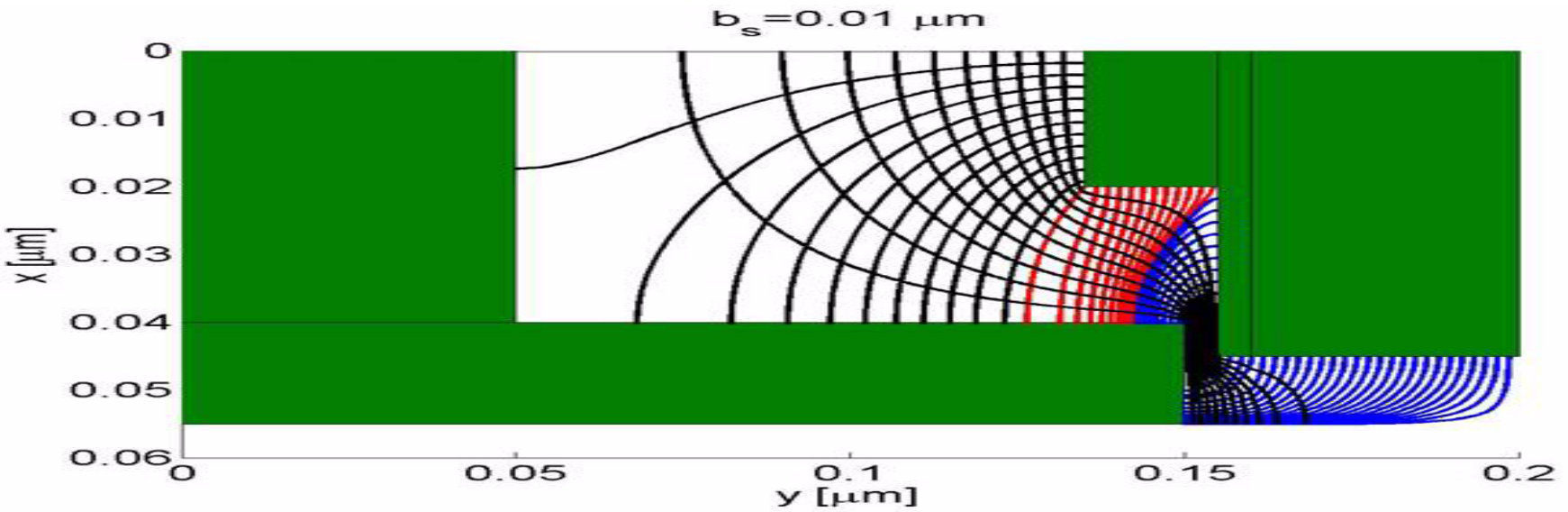
$w_{so} = b_{soi}$ = [2.5...5]nm

b_{so} = [10...20]nm

electrostatic analysis of spacer and contact (BE, BC) structure
=> scalable analytical model

Example: BE spacer electrostatic analysis

field lines from Poisson solver



(plot laterally stretched to reflect true dimensions)

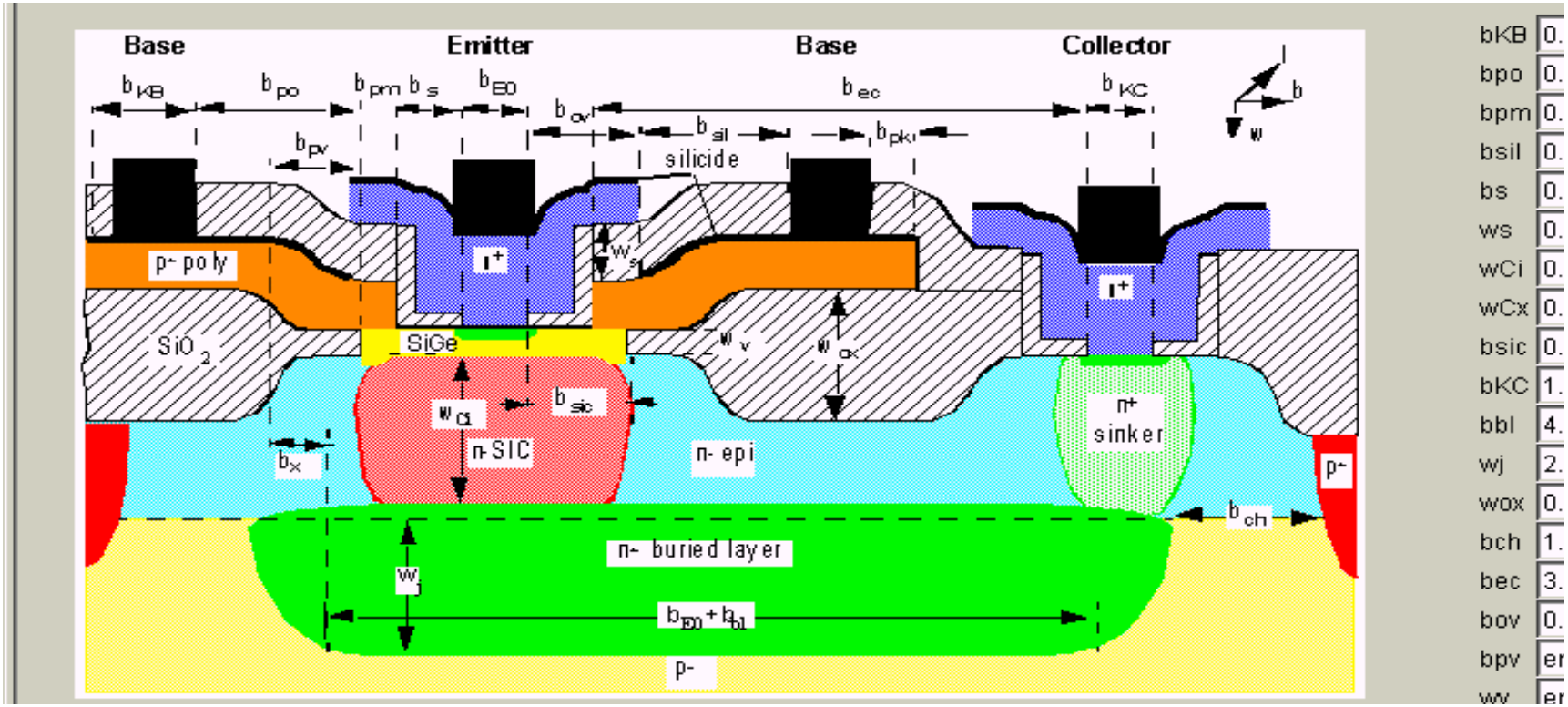
- similar analysis for parasitic BC and contact metallization capacitances

=> all relevant parasitic capacitances included

3D scaling analysis

Goal: find lateral dimensions yielding balanced device design ($f_{max} \geq f_T$)

- starting point: B4T design rules, B30x extrapolations
  => simultaneous lateral shrink of *all* dimensions using TRADICA scaling factor
- assumed device structure is still fairly conventional



=> conservative estimate, leaves margin for innovative changes

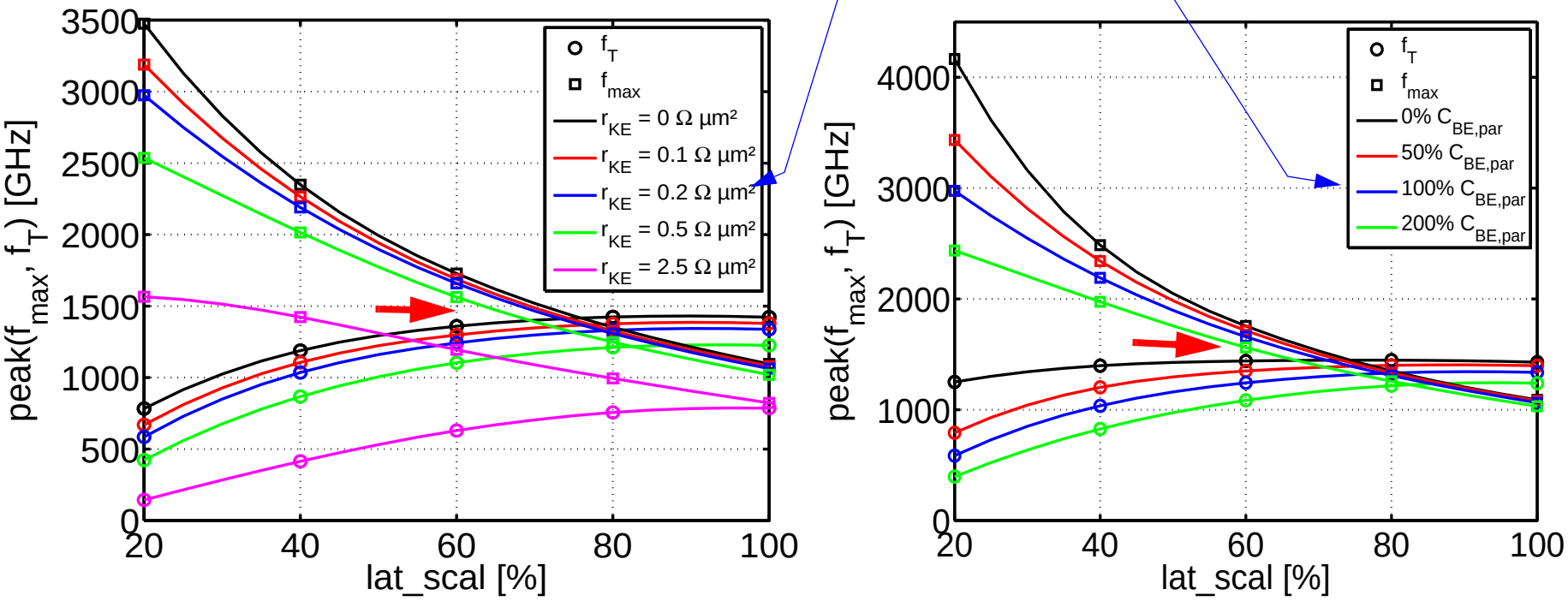
Sensitivity w.r.t. selected critical parameters

emitter contact resistance

parasitic BE capacitance

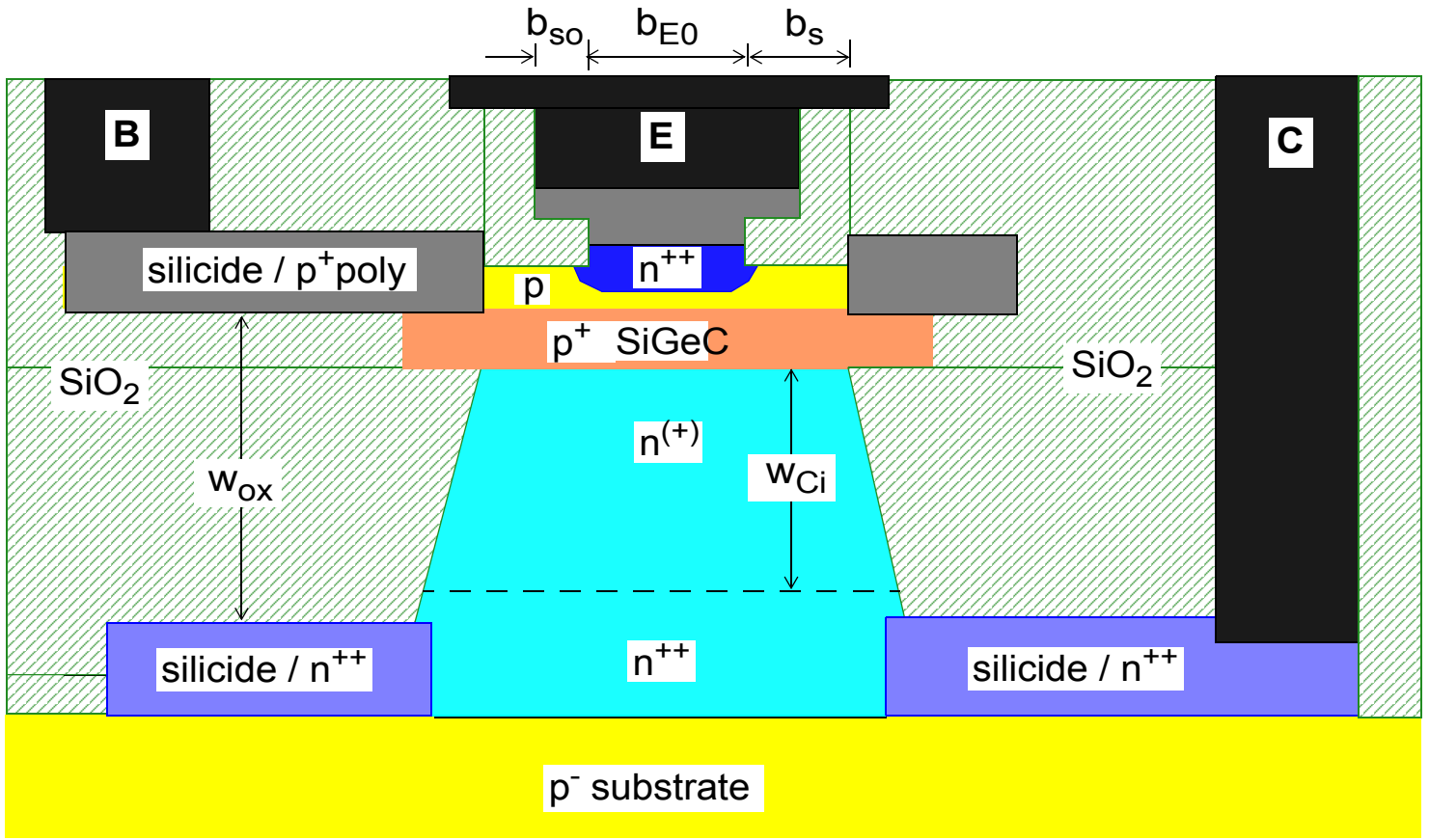
reference transistor size = 0.05*1 μm^2

(at intercept: b_{E0} = 37 nm, R_E = 5 $\Omega\mu$m, $C_{BE,par}$ = 0.36 fF/μm)



=> increasing parasitics at given process "node" slightly shifts lateral scaling for ***balanced*** design to the right

HBT structure towards ultimate limits

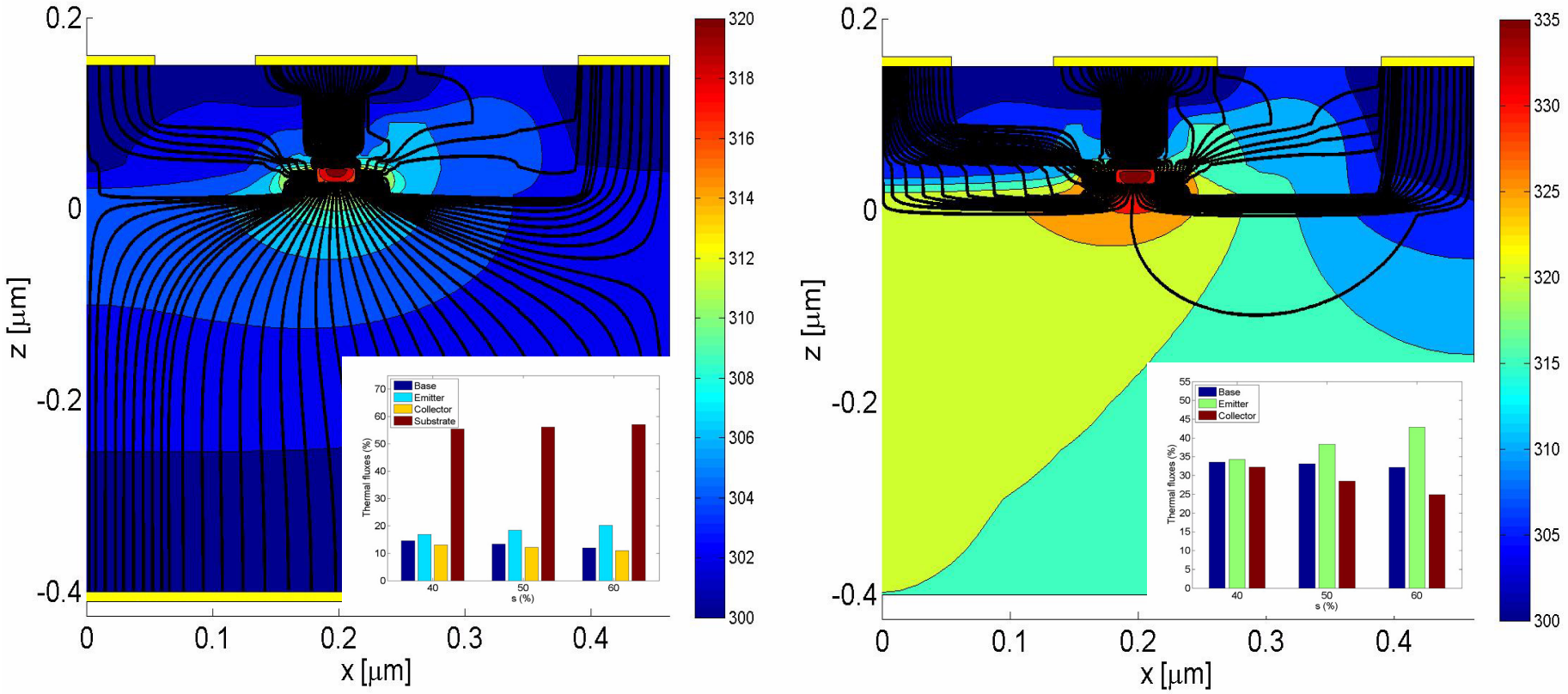


- almost 1D current flow
- no deep trench necessary
- low-ohmic buried layer (possibly silicided)

Electrothermal considerations

heat flux and temperature distribution (b_{E0} = 0.03 μm, l_{E0} = $3b_{E0}$)

bulk structure P_{dis} = 0.4mW SOI structure



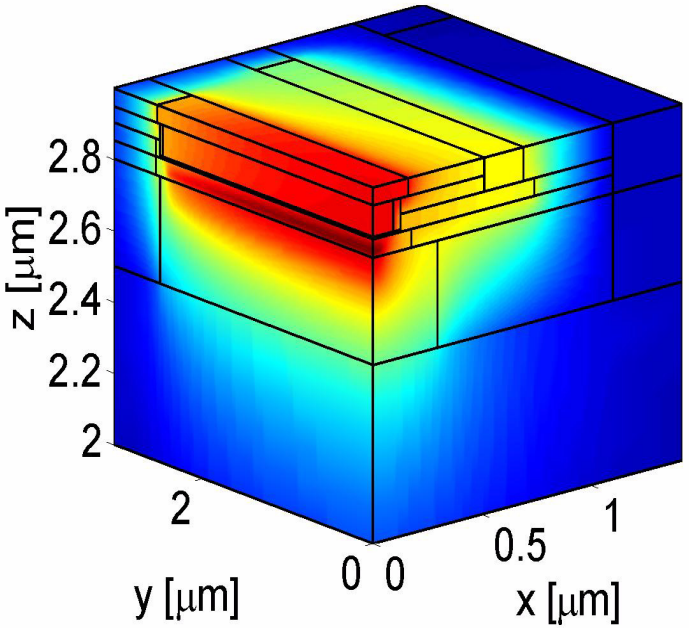
heat flux mostly through bulk

heat flux equally through contacts

=> bulk structure fabrication appears feasible (acc. to process eng.)

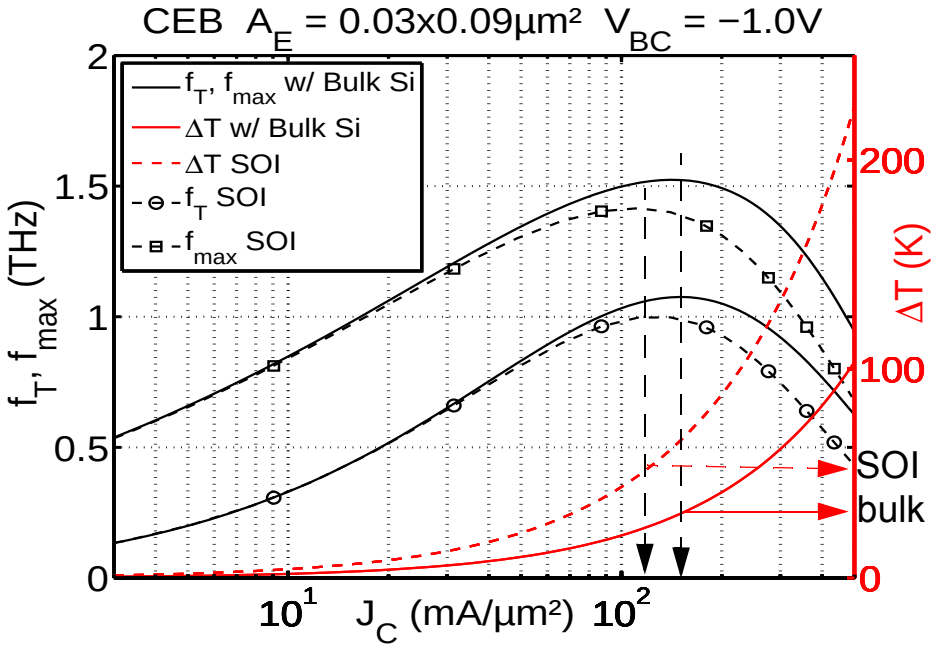
Electrothermal considerations

- R_{th} from 3D thermal simulator, C_{th} *estimated* from R_{th} and time constant
- temperature coefficients from combination of device simulations (mostly 1D elements) and of experimental data on existing processes (external elements)
- temperature increase for 60% scaling



example for 3D T distribution

=>

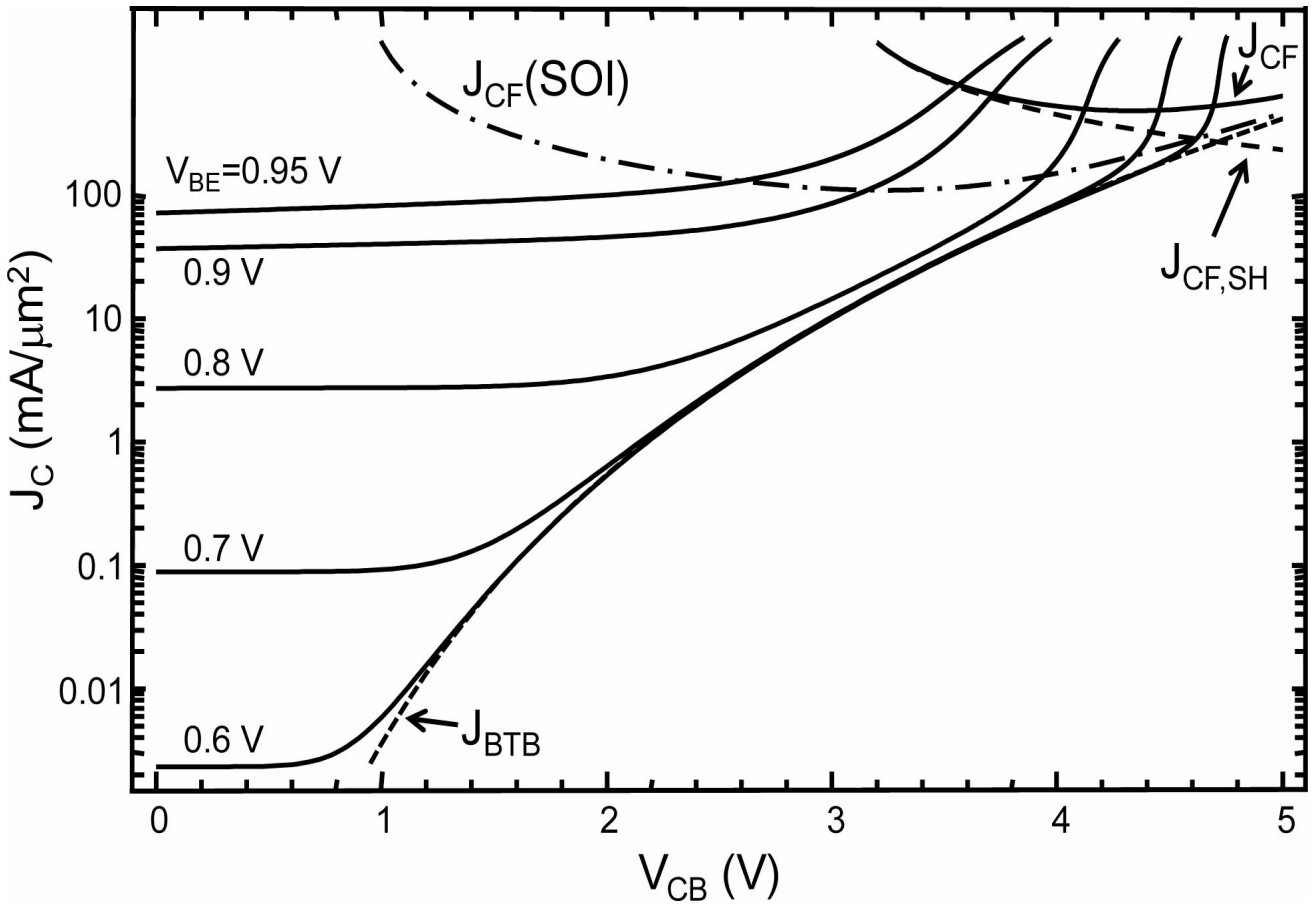


=> moderate drop in peak values

=> tolerable T increase up to peak f_T, f_{max}

Safe Operating Area

- calculated anaytically from TCs, including BTB Tunneling and avalanche current



snapback current density

snapback s.h. only

=> surprisingly high BC breakdown voltage in useful bias range

Summary of 3D scaling

scaling factor electrical parameters	60%	50%	40%
b_{E0} (nm)	30	25	20
l_{E0} (nm)	90	75	60
R_{Th} (K/mW)	84	105	142
f_{max} (THz) @ J_C (mA/μm^2)	1.37 @ 106	1.60 @ 109	1.91 @ 119
peak f_T (THz) @ J_C (mA/μm^2)	1.01 @ 131	0.99 @ 144	0.95 @ 155
τ_{CML} (ps) @ J_C (mA/μm^2)	0.57 @ 280	0.52 @ 301	0.52 @ 419

Issues

- electromigration (J_C > 150mA/μm^2 at peak f_T, f_{max})
- steep doping profiles (especially base) and integration into CMOS
- parasitics: metallization, contacts (especially emitter), access regions (base link)

Conclusions

- A set of calibrated simulation tools was used for predicting performance limit of SiGeC HBTs

- 1D f_T limit is around 1.5 THz
- 2D/3D limit is around (f_T, f_{max}) = (1.2, 1.5) ... (1.1, 2.2) THz at BVCEO > 1V and emitter contact width of 30 ... 20 nm
- Further shrink improves f_{max}, t_{CML} somewhat but at expense of significant f_T drop

- Biggest challenges
  - high current density at peak (f_T, f_{max}) => exceeding existing electromigration limits
  - reduction of emitter resistance to at least $0.5 \Omega \mu m^2$
  - steep doping profiles

=> significant innovation required to achieve physical limits

=> prediction of SiGeC HBT performance limit facilitates roadmap generation