Negative Bias Temperature Instability: Modeling Challenges and Perspectives

Tibor Grasser

Christian Doppler Laboratory for TCAD in Microelectronics
Institute for Microelectronics, TU Vienna
Gußhausstraße 27–29, A-1040 Wien, Austria

Acknowledgments

This work would not have been possible without the support of ...

B. Kaczer and G. Groeseneken (IMEC)

Longstanding collaboration, most of the measurement data, discussion/theory

M. Nelhiebel, Th. Aichinger, J. Fugger, and O. Häberlen (Infineon Villach)

Financial support, measurement data, and discussion

R. Minixhofer and H. Enichlmair (austriamicrosystems)

Financial support, measurement data, and discussion

H. Reisinger, C. Schlünder, and W. Gustin (Infineon Munich)

Ultra fast measurement data

My reliability Ph.D. students

R. Entner, W. Gös, Ph. Hehenberger, F. Schanovsky, O. Triebl, and P.-J. Wagner

Reliability community

P. Lenahan, J. Campbell, G. Bersuker, V. Huard, N. Mielke, A.T. Krishnan, S. Mahapatra, S. Zafar, E. Islam, ... (I bet I forgot someone, sorry ...)

Overview

Introduction

Selected experimental observations

What is NBTI?

What would we expect from a good/physics based/predictive NBTI model?

Published NBTI models

How do existing models live up to these expectations?

Where are the biggest discrepancies?

Summary

Where are we? Where do we need to go?

What is Negative Bias Temperature Instability?

Negative bias temperature stress of pMOSFETs[1] [2] [3]

  Large negative gate voltage ($\approx 5 - 8\,\mathrm{MV/cm}$), all other terminals grounded

  Elevated temperatures (typically $100\,^{\circ}\mathrm{C} - 200\,^{\circ}\mathrm{C}$, but also at room temperature)

Degradation of critical device parameters

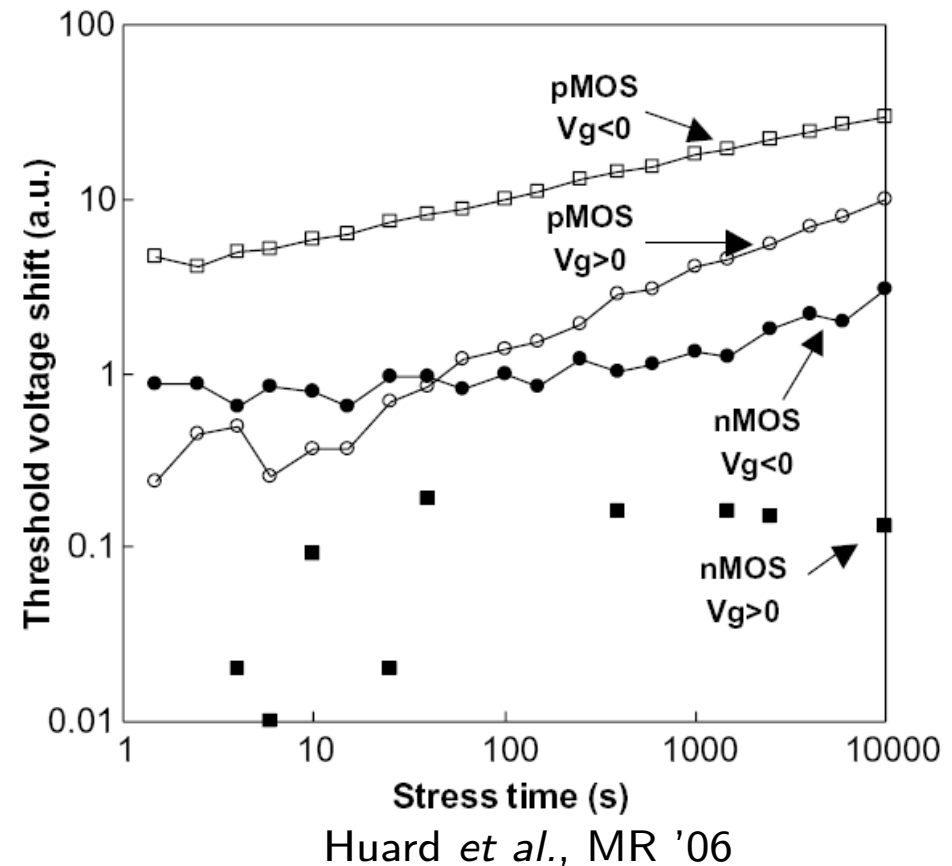 Threshold voltage

  Subthreshold slope

  Transconductance

  Mobility

  Drain current

  ...

Occurs in all four configurations

  Strongest in pMOS with negative bias

Huard *et al.*, MR '06

One of the most important reliability concerns in pMOSFETs

[1] Schroder and Babcock, JAP '03   [2] Alam and Mahapatra, MR '05   [3] Huard *et al.*, MR '06

What is Negative Bias Temperature Instability?

What happens during negative bias temperature stress?

Creation of SiO_2/Si interface defects (dangling Si bonds, P_b centers)
- Pre-existing, but passivated by hydrogen anneal
- Si–H bonds can be broken
- Results in trapping sites inside the Si bandgap
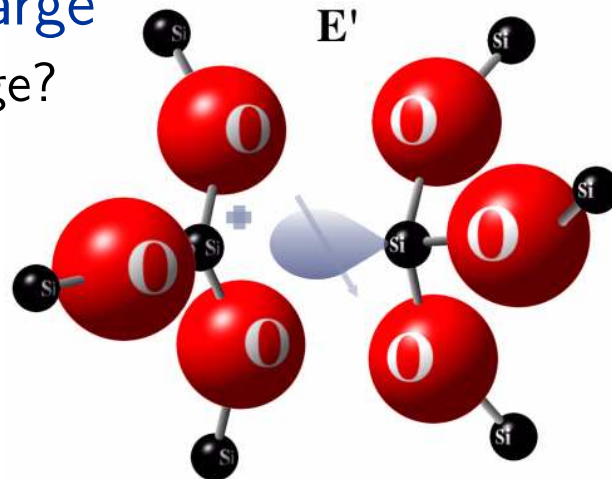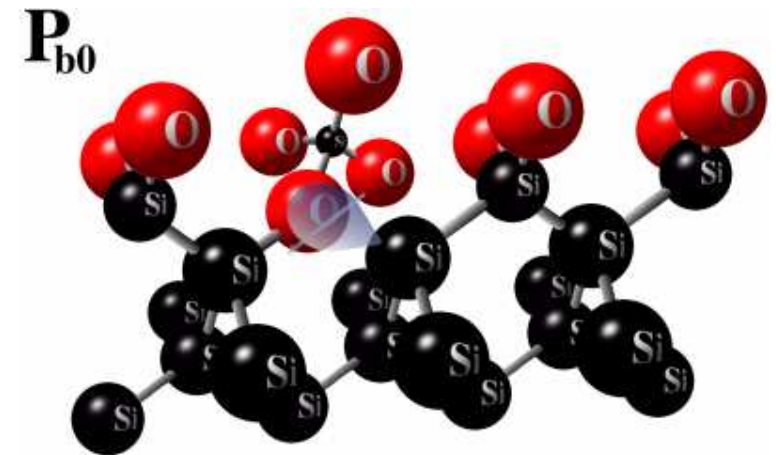- Universally acknowledged [1] [2]
- Different defect in SiON? [3]

P_{b0}

[Courtesy: PennState Univ.]

Creation of oxide charge
- Fixed positive charge?
- Trapped protons?
- Controversial! [4]

E'

[Courtesy: PennState Univ.]

[1] Mahapatra *et al.*, IRPS '07   [2] Huard *et al.*, MR '06   [3] Campbell *et al.*, TDMR '07   [4] Mahapatra *et al.*, IRPS '07

NBTI Measurement Techniques

Main problem: it is impossible to perfectly measure NBTI

As soon as stress is removed, extremely fast recovery is observed[1] [2]

Strong bias dependence, in particular to positive bias[3] [4] [5]

Different techniques proposed which measure certain aspects of NBTI

OTF techniques gives us ΔI_D at $V_\mathrm{G} = V_\mathrm{stress}$[6]

ΔI_D has to be converted to ΔV_th using empirical relationships/compact model

Suffer from initial delay, mobility variations, compact modeling errors[7] [8] [9]

Measure/stress/measure (MSM) techniques give us ΔV_th[10]

Suffer from recovery due to measurement delay

However, much smaller sensitivity to mobility variations

Charge-pumping and DCIV techniques: interface vs. oxide charges[11]

Suffer from huge delays and positive bias during measurement

Biggest problem: results do not match!!!

No exact theory available that unanimously links and explains all the data

[1] Ershov *et al.*, IRPS '03   [2] Reisinger *et al.*, IRPS '06   [3] Ang, EDL '06   [4] Huard *et al.*, MR '06

[5] Grasser *et al.*, IEDM '07   [6] Denais *et al.*, IEDM '04   [7] Islam *et al.*, IEDM '07   [8] Grasser *et al.*, IIRW '07

[9] Reisinger *et al.*, TDMR '07   [10] Kaczer *et al.*, IRPS '05   [11] Neugroschel *et al.*, IEDM '06

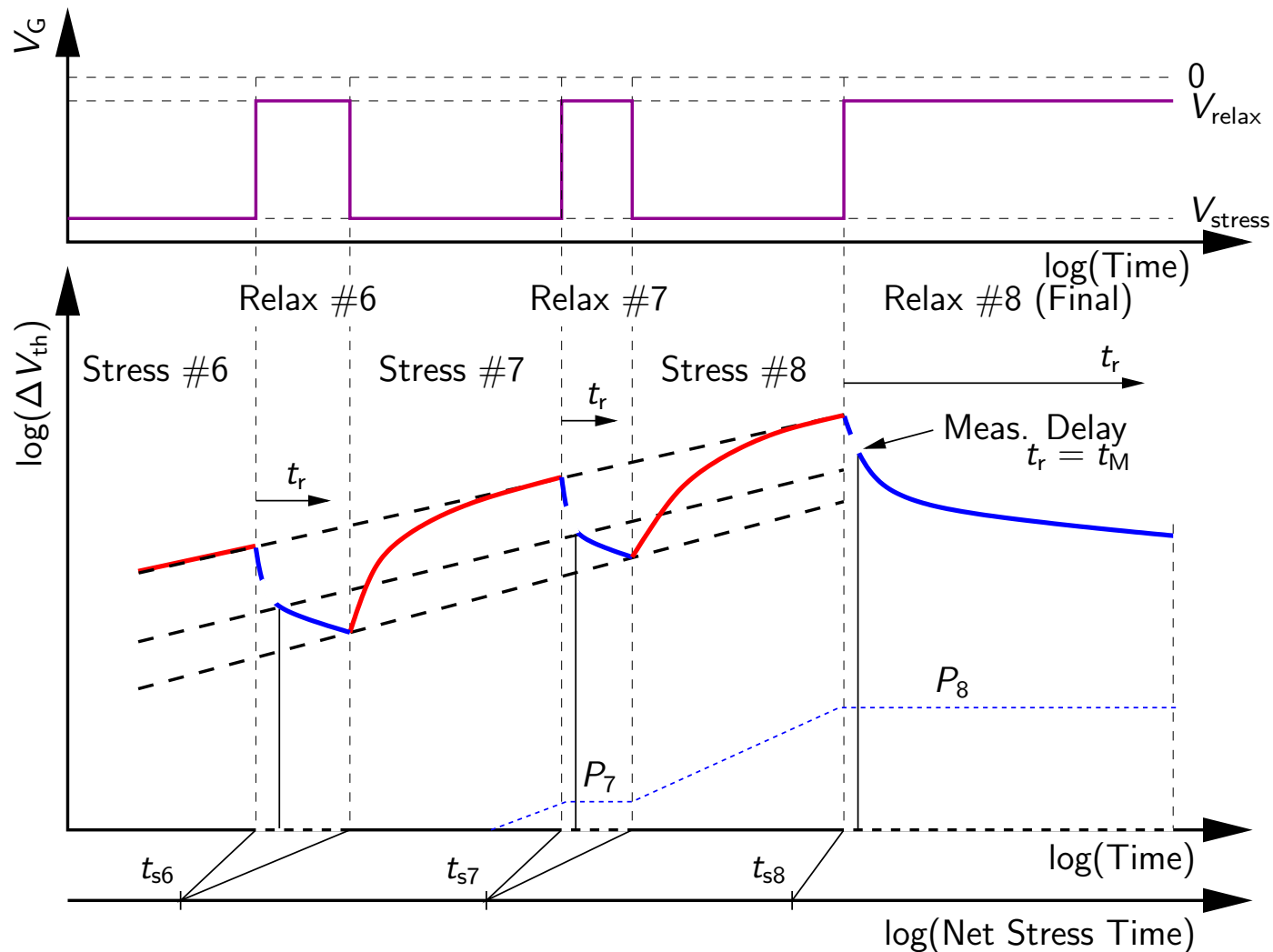# Experimental Observations

(What should we actually model?)

**There are many more,
just a selection of some nasty ones
(from a modeling point of view)**

Extended measure/stress/measure (MSM) technique[1] [2] [3]

Application of several stress/recovery phases in one measurement

Record as much stress and relaxation data during one measurement as possible

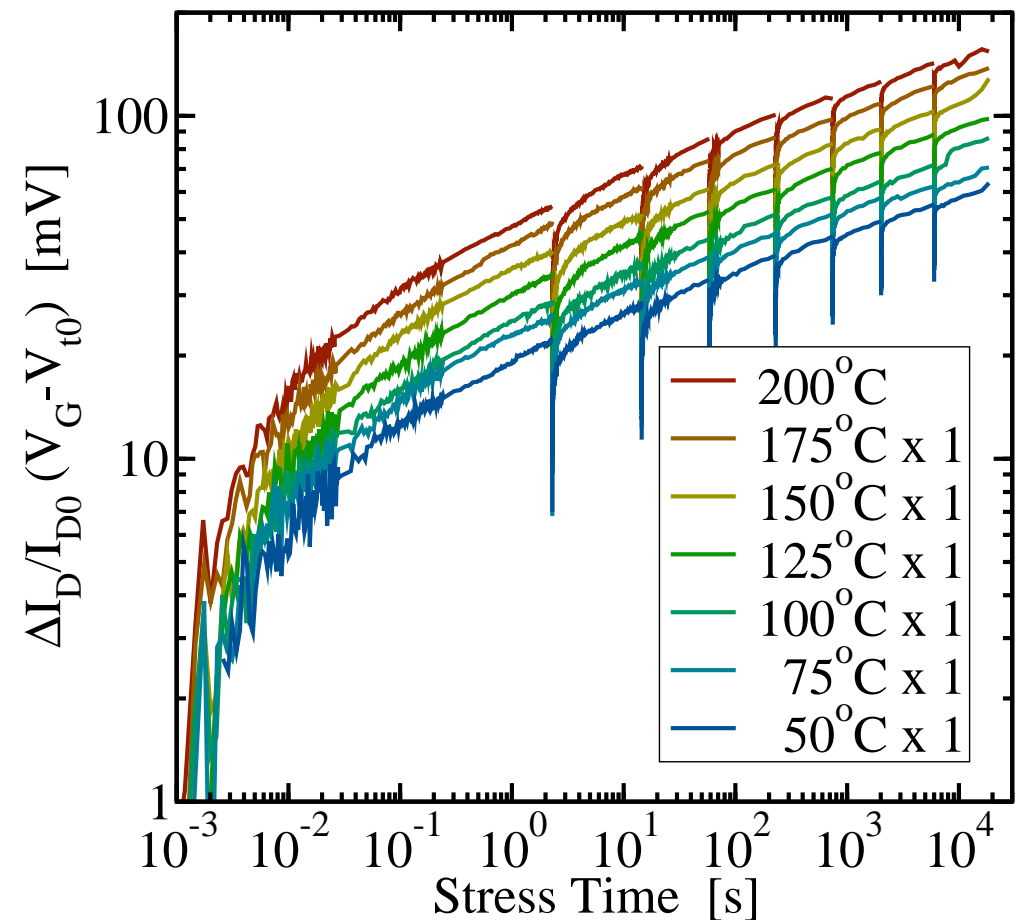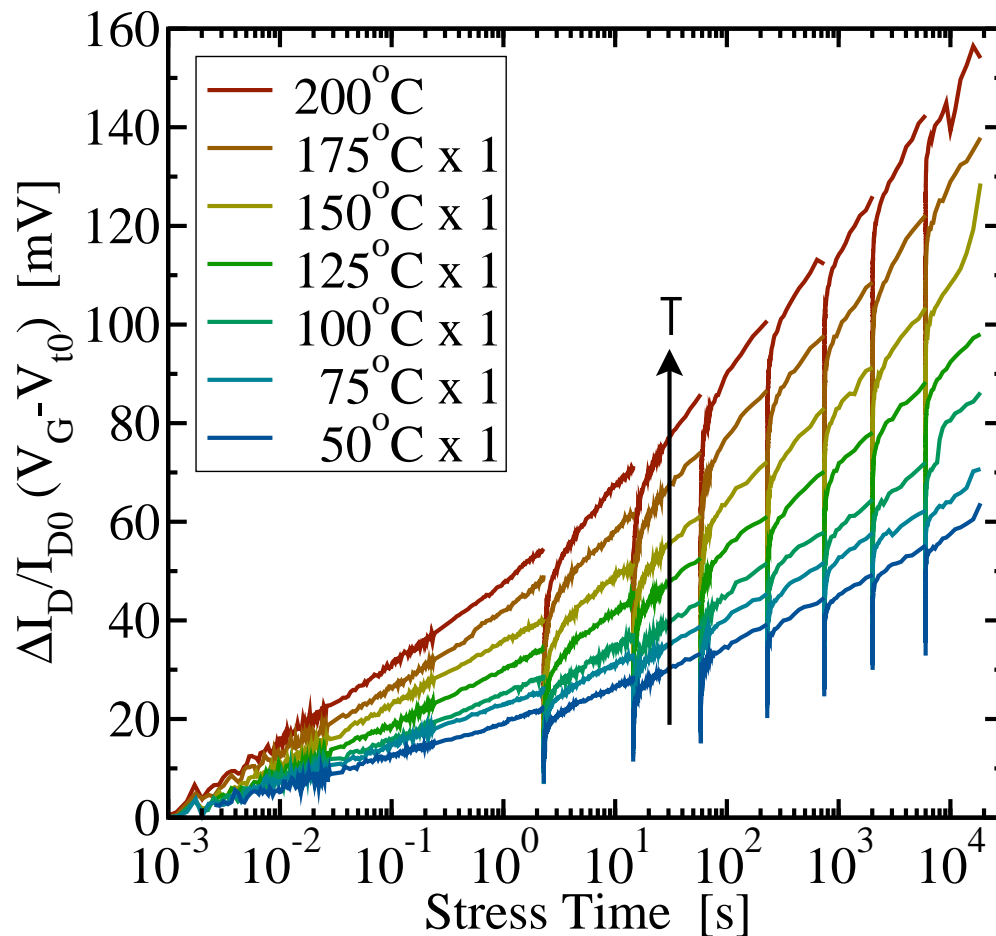[1] Kaczer et al., IRPS '05   [2] Grasser et al., IEDM '07   [3] Kaczer et al., IRPS '08

Temperature Scaling: Stress

Strong temperature activation of stress behavior[1]

On-the-fly recording of $I_D(t_s)$, monitor change relative to $I_D(1\,ms)$

Conversion of ΔI_D using compact-model under the assumption of constant mobility

Initially log(t), then power-law? [2] [3]

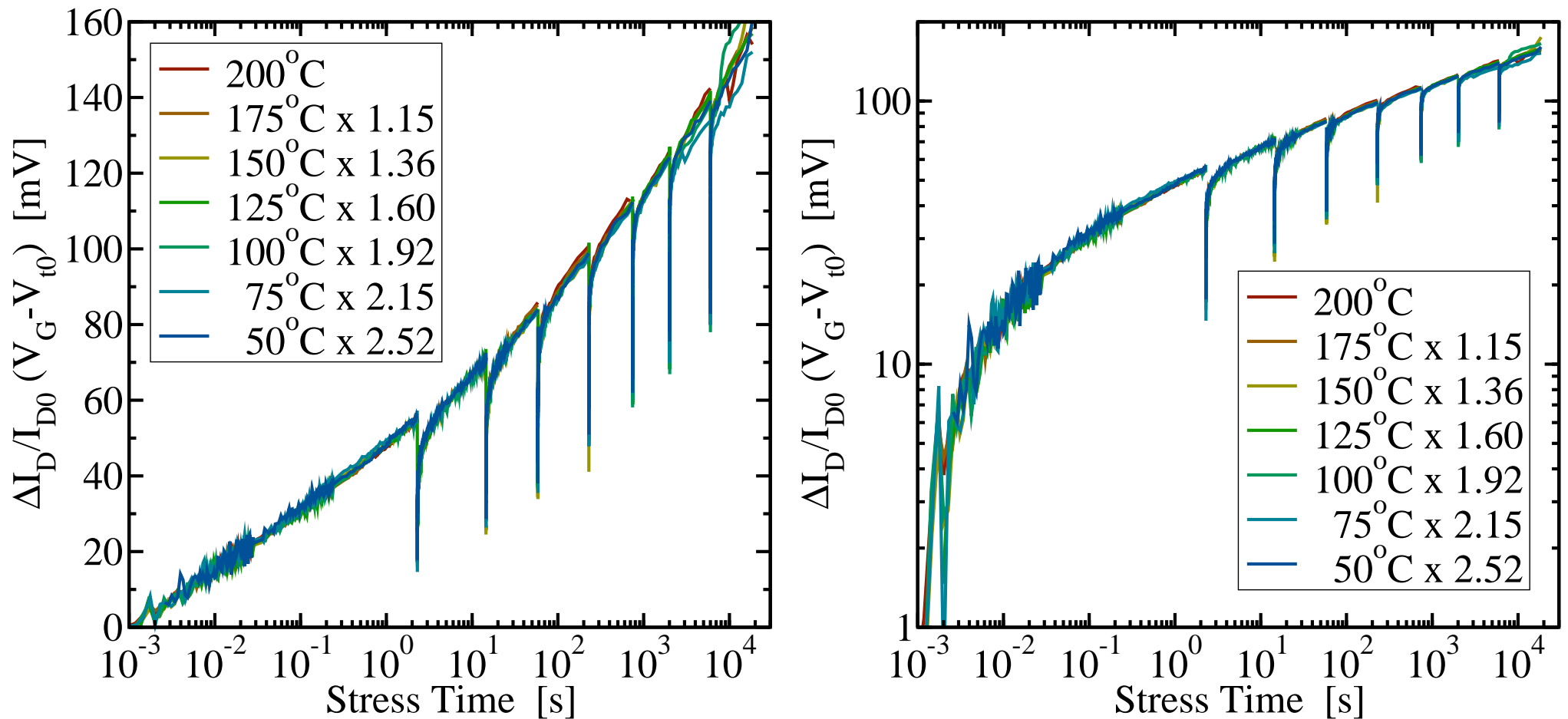[1] Mahapatra *et al.*, IRPS '07   [2] Reisinger *et al.*, IRPS '06   [3] Huard *et al.*, IEDM '07

Temperature Scaling: Stress

Stress: IMEC data can be made to overlap by simple scaling[1]

$$\Delta V_{th}(t_s, T) \approx \theta(T) \times \Delta V_{th}(t_s)$$

Does this indicate the dominance of a single mechanism?

Power-law approximation: slope is (roughly) independent of temperature

[1] For a contrary example see Reisinger *et al.*, IRPS '06, Fig. 4
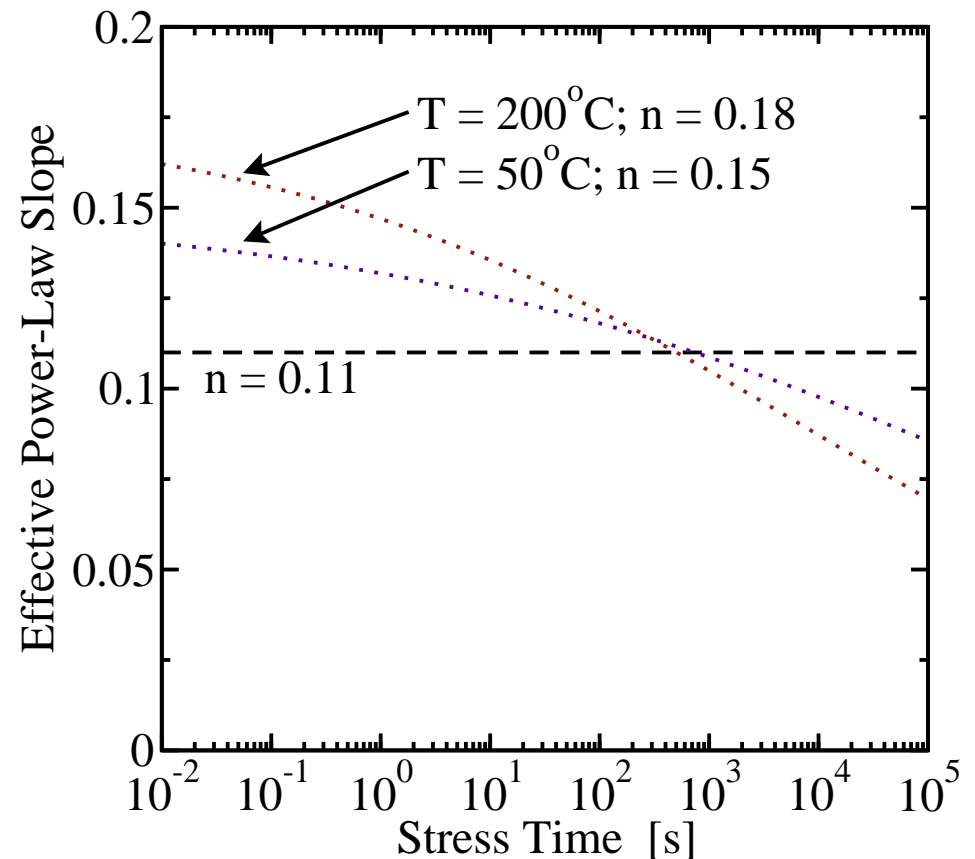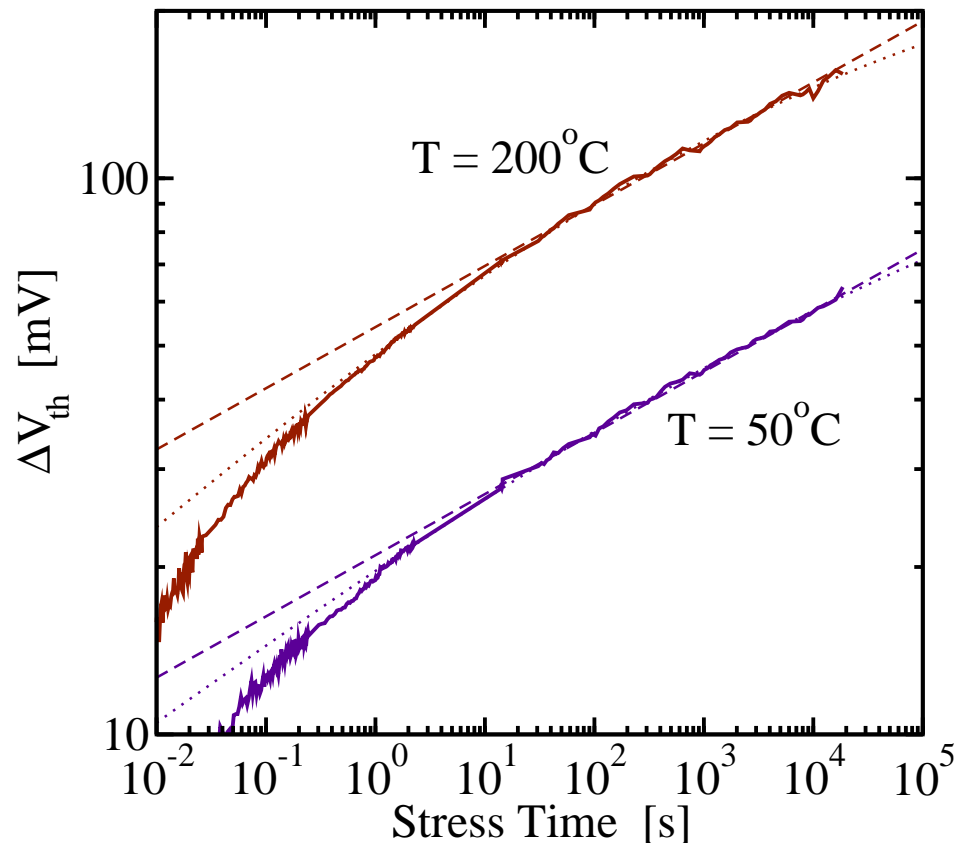
Is it a Power-Law?

Although data may look like a power-law, it is not really a power-law!

$$\Delta V_{th}(t_s) \approx A\, t_s^n$$
Fit to long-time data: n temperature-independent?

Data better fitted by the following expression[1] (implies saturation)

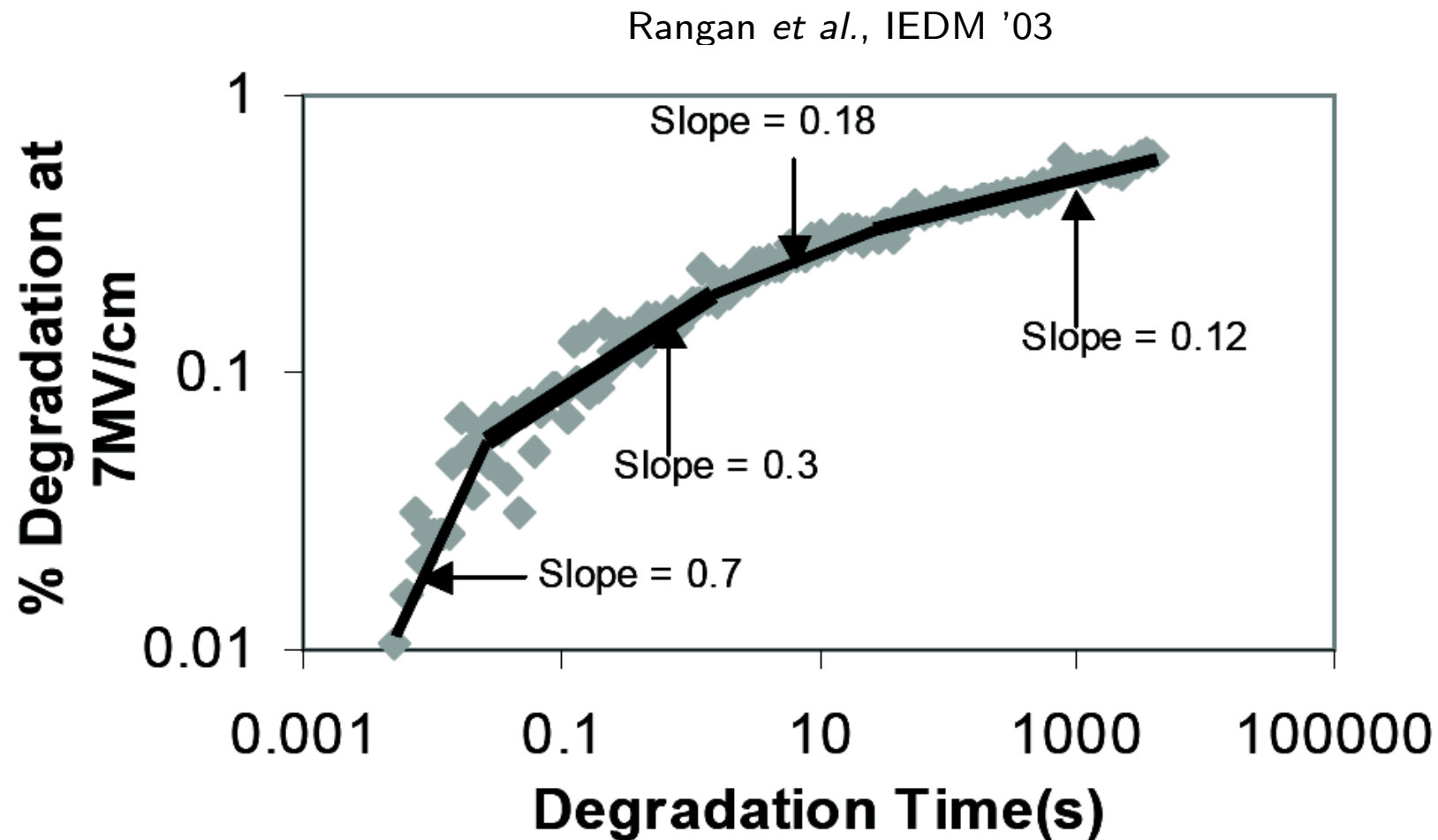$$\frac{\Delta V_{th}(t_s)}{\Delta V_{th,max}} \approx \frac{1}{1 + (t_s/\tau)^{-n}}$$

Fit interval? n may be temperature-dependent

[1] Huard et al., MR '06

Is it a Power-Law?

Degradation does not follow a perfect power-law[1] [2] [3]
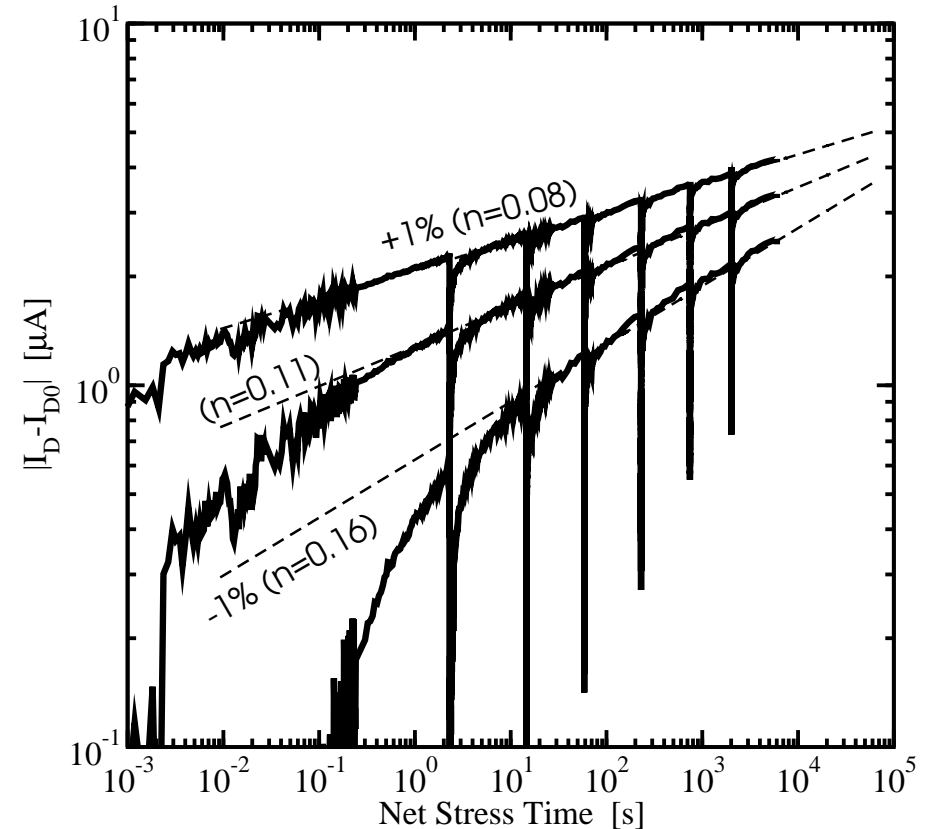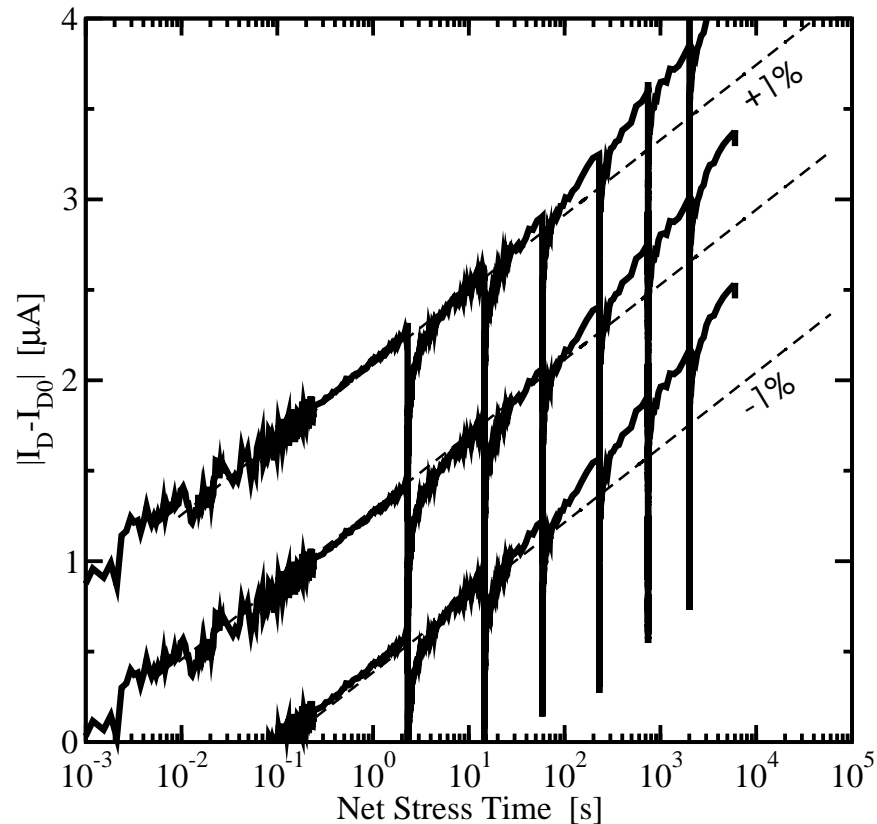

Rangan *et al.*, IEDM '03

[1] Rangan *et al.*, IEDM '03   [2] Huard *et al.*, MR '06   [3] Reisinger *et al.*, MR '07

The Power-Law Approximation

One of the dangers of using a power-law approximation[1]

Strong sensitivity to first measurement point, e.g. 1% accuracy in $1\,\mu s$ [2] [3]

Nevertheless, in the following the term '(power-law) slope' will also be used

Main reason: many people are familiar with typical numbers, e.g. 1/4 vs. 1/6 [4]

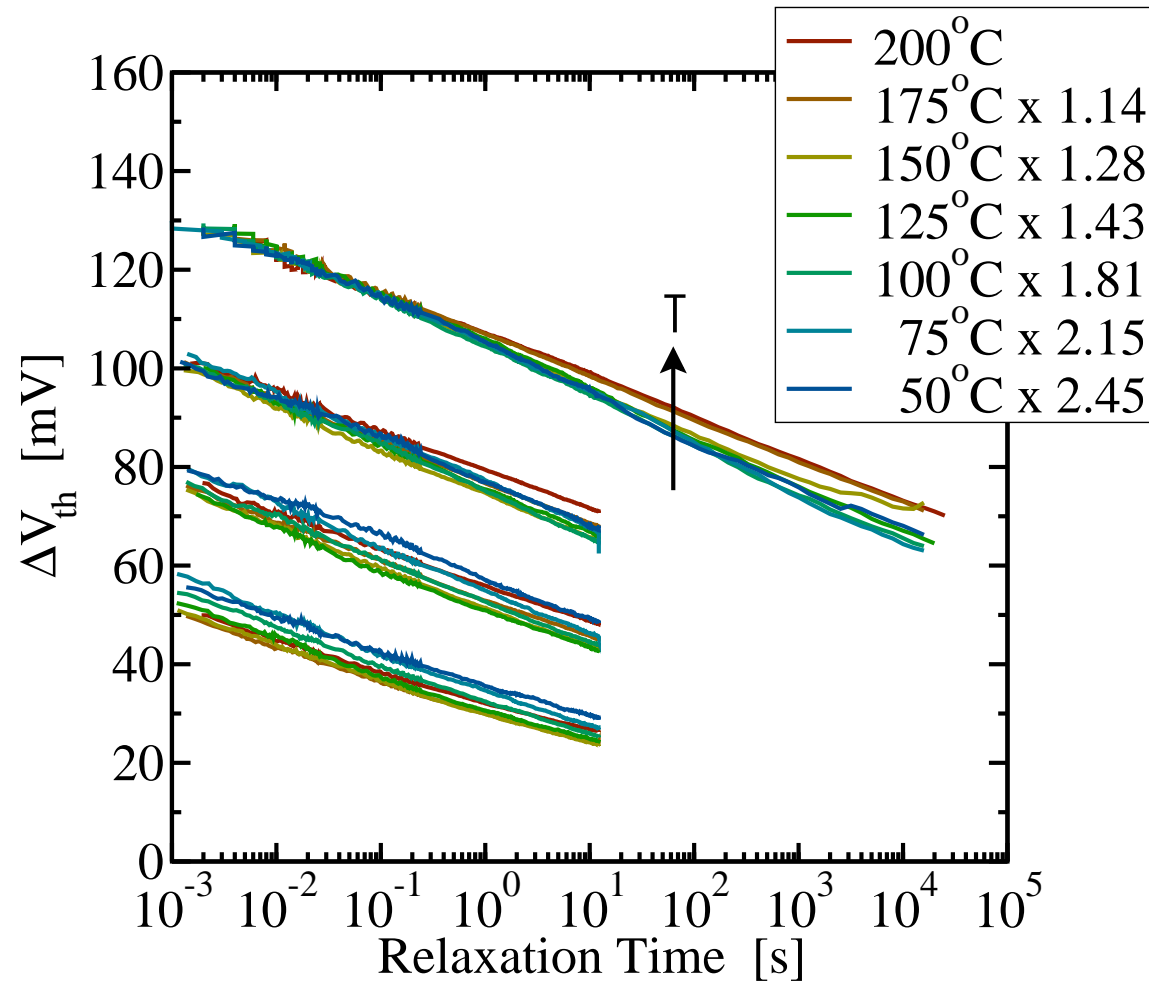Power-law slope: derivative of data on a log-log plot (or, locally fitted to At_s^n)

We just need to remember to be careful with this term!

[1] Shen *et al.*, IEDM '06   [2] Reisinger *et al.*, TDMR '07   [3] Kumar *et al.*, IEDM '07   [4] Mahapatra *et al.*, IRPS '07

Temperature Scaling: Recovery

Recovery behavior

Higher temperature seems to result in slower recovery (larger permanent damage?)

Stress behavior
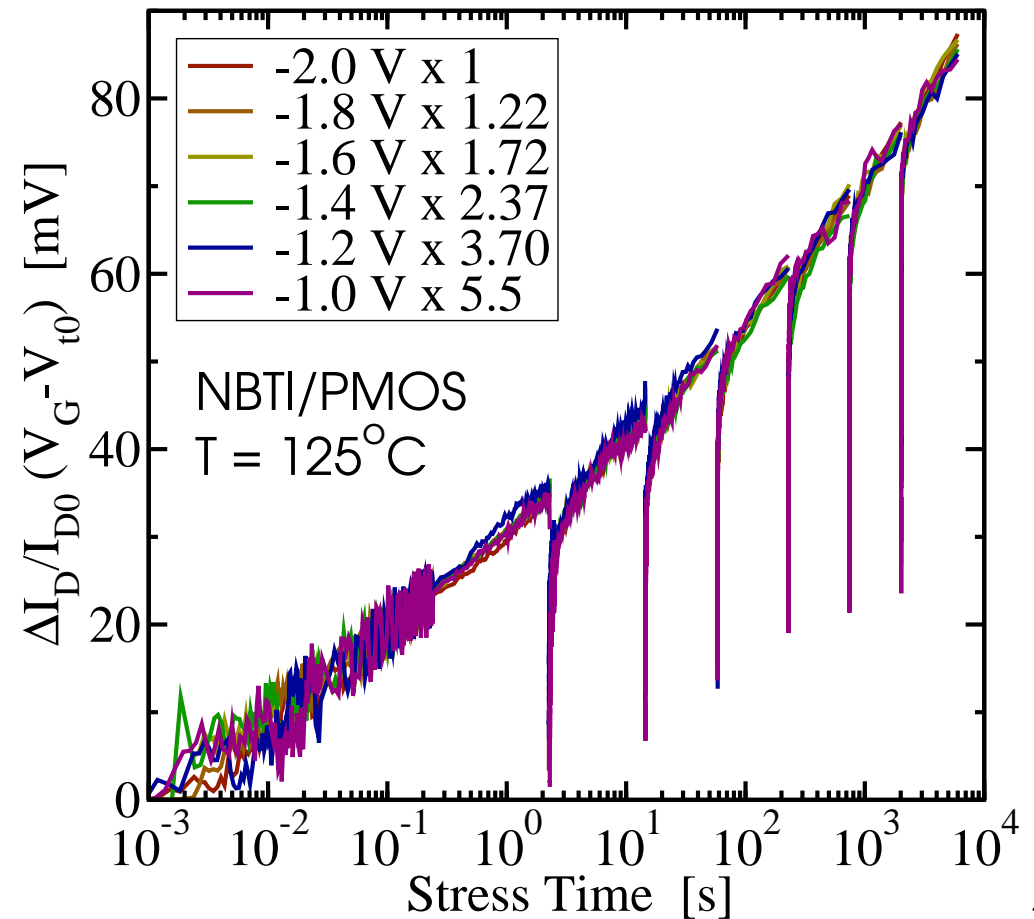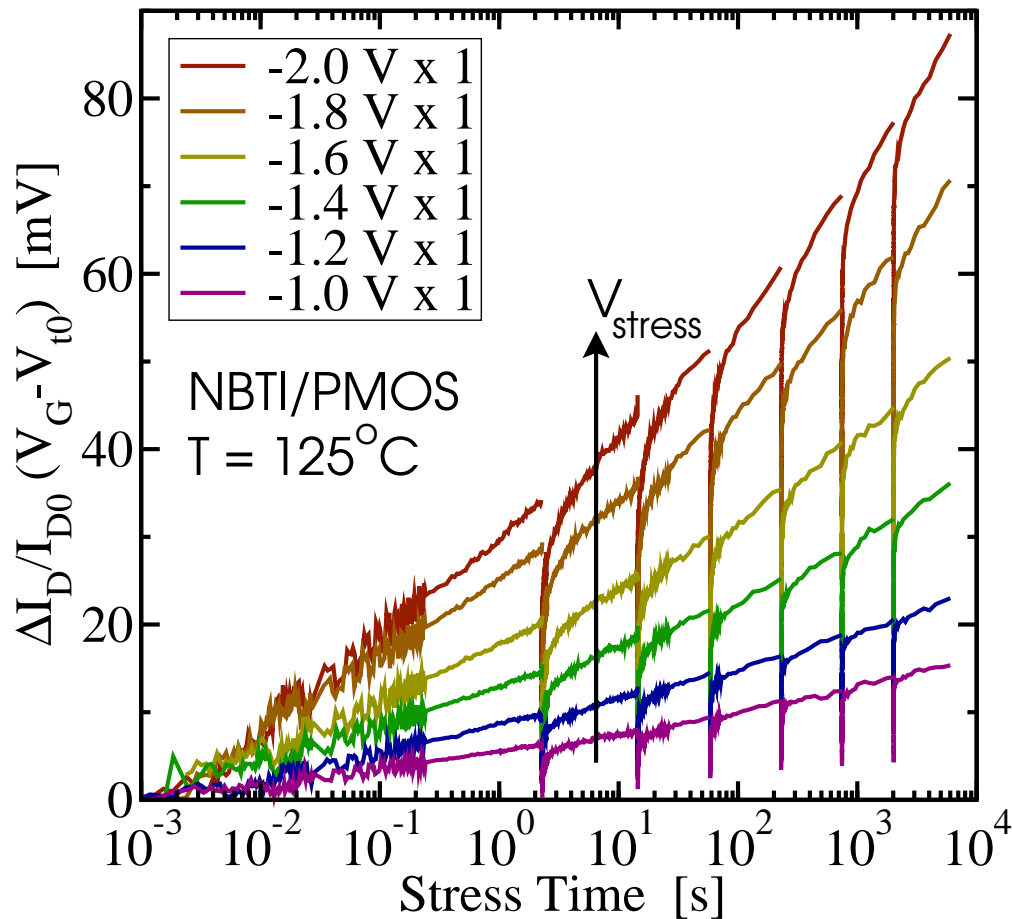
As with temperature activation, data can be made to overlap

$$\Delta V_{\text{th}}(t_s, V_{\text{stress}}) \approx \varsigma(V_{\text{stress}}) \times \Delta V_{\text{th}}(t_s)$$

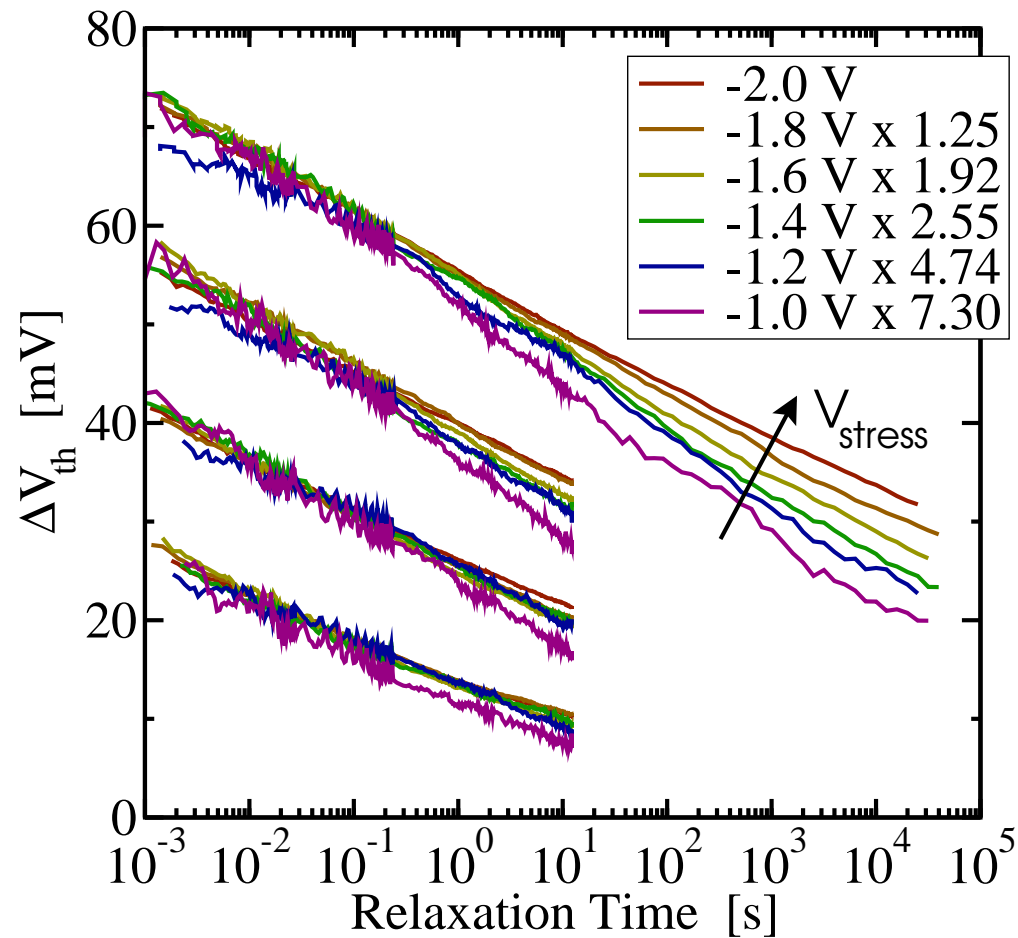Again, does this indicate the dominance of a single mechanism?

Is this possible with 2 independent mechanisms with different T/V activation?

15

Voltage Scaling: Recovery

Recovery behavior

Higher voltages seems to result in slower recovery (larger permanent damage?)

One or Two Components?

Some groups claim that NBTI is dominated by a single mechanism[1][2]

    If there is any hole trapping at all, then only in initial data

    Dominance of a single mechanism supported by voltage/temperature scaling

Other groups claim that two components contribute to NBTI

    Hole trapping gives the fast component, N_{it} follows reaction-diffusion theory[3][4][5]

    Hole trapping can be permanent/slowly relaxing[6][7]

    Hole trapping is fast and recoverable, N_{it} is permanent[8][9]

$\Rightarrow$ Literature is anything but consistent regarding hole-trapping

    Strong technology/processing dependence?

[1] Mahapatra *et al.*, IRPS '07 [2] Kumar *et al.*, IEDM '07 [3] Shen *et al.*, IEDM '06 [4] Reisinger *et al.*, IRPS '06

[5] Ang, APL '06 [6] Zhang *et al.*, TED '04 [7] Ang and Wang, EDL '06 [8] Huard *et al.*, MR '06
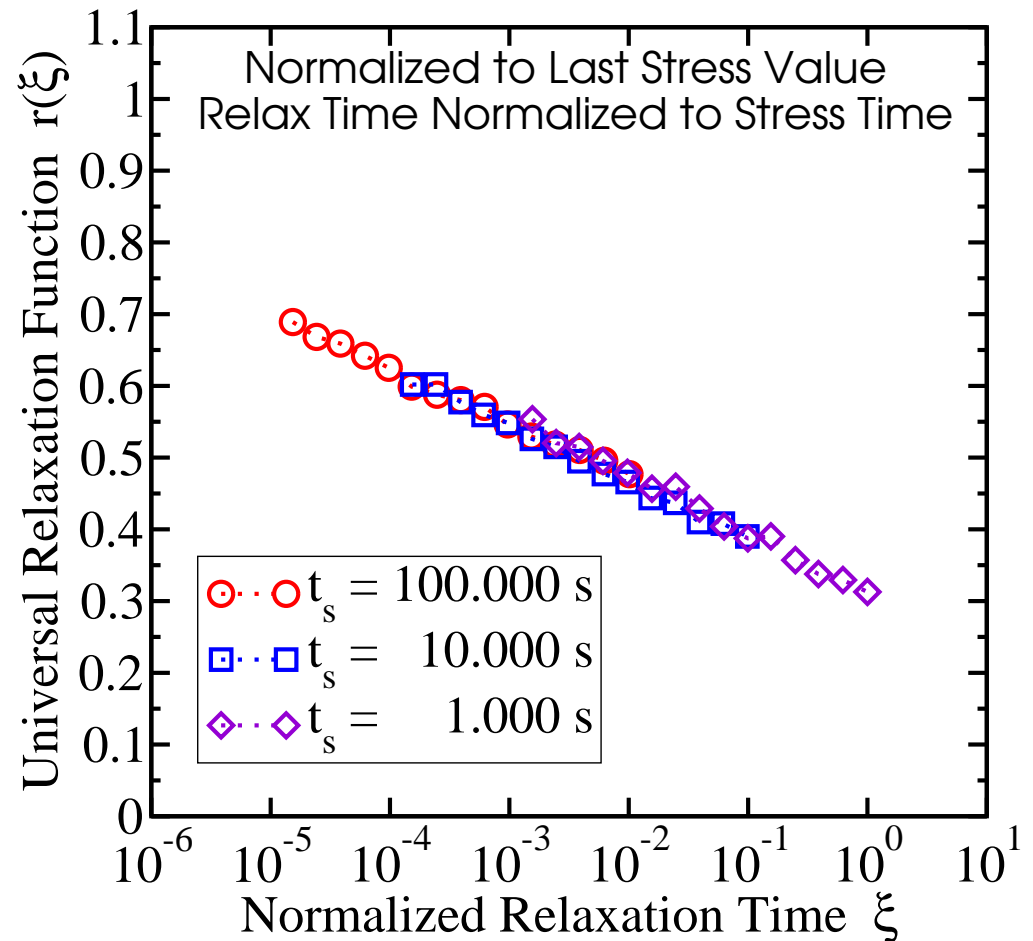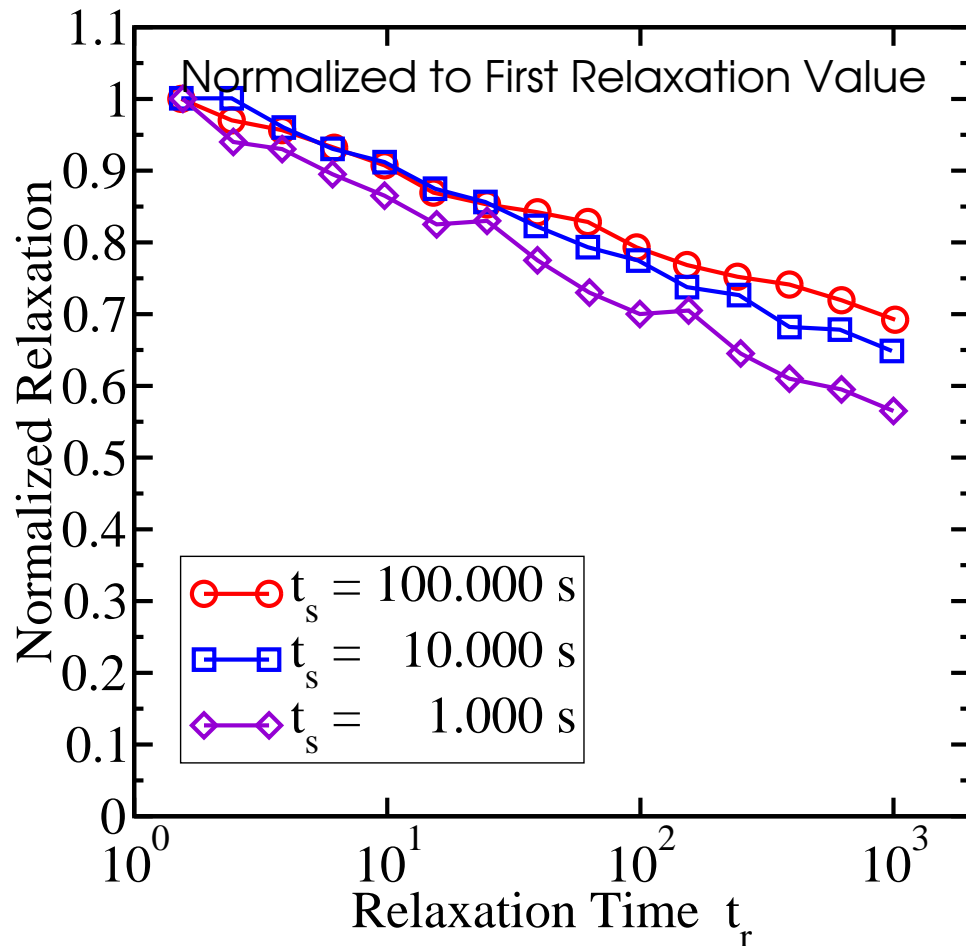
[9] Huard *et al.*, IEDM '07

Universal Relaxation of On-The-Fly-Data

Interesting aspect of relaxation: universality

On-the-fly data[1]

Normalized relaxation depends only on normalized relaxation time $\xi = t_r/t_s$

Universal Relaxation: Analytic Expression

Possible empirical expressions[1]

Normalized relaxation time

$$\xi = t_r / t_s$$

Power-law

$$r(\xi) = 1/(1 + B\xi^\beta)$$

Log-like, variant A

$$r(\xi) = 1 - \beta \log(1 + B\xi)$$

Log-like, variant B

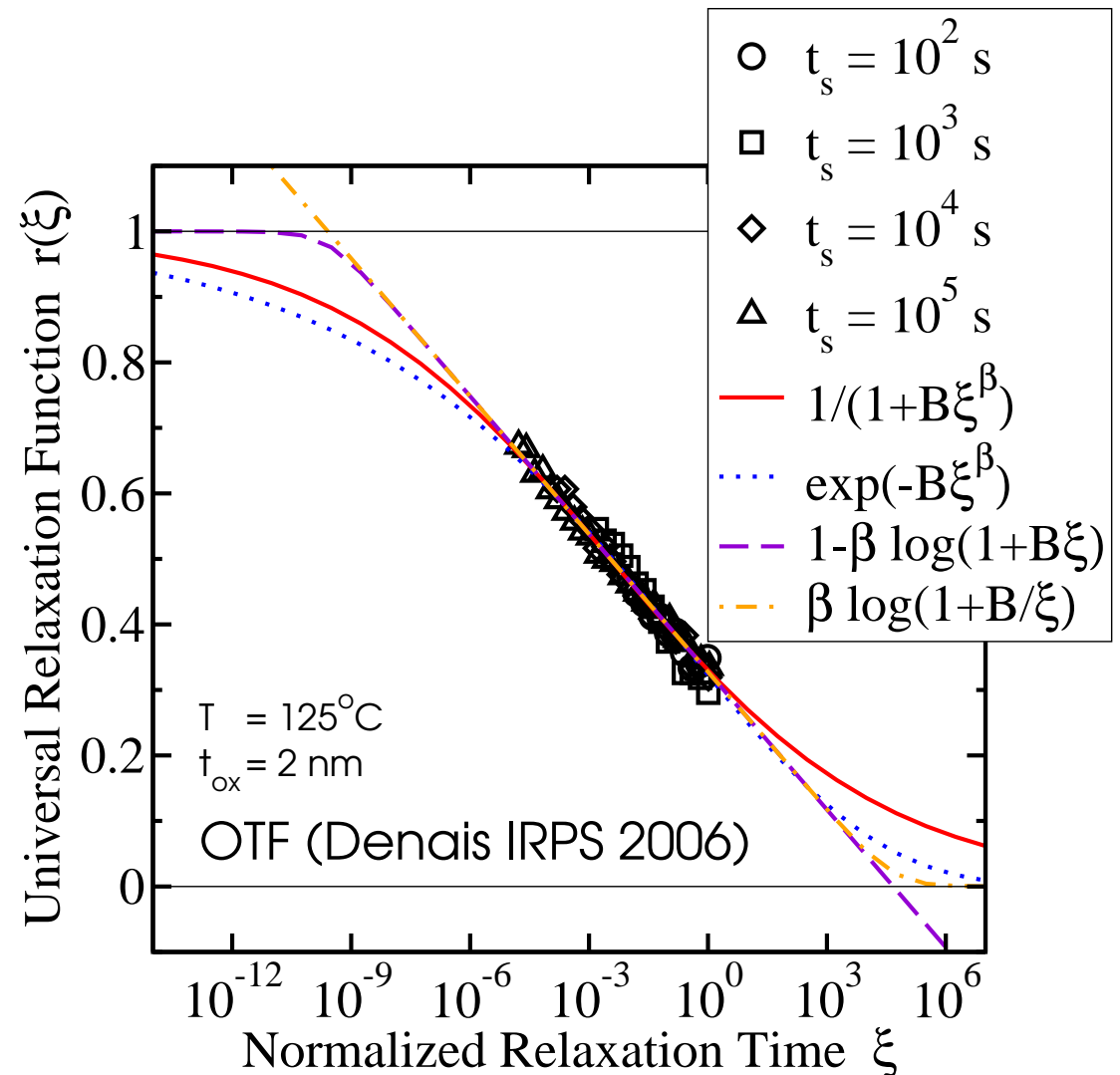$$r(\xi) = \beta \log(1 + B/\xi)$$

Stretched exponential

$$r(\xi) = \exp(-B\xi^\beta)$$

We choose $1/(1 + B\xi^\beta)$

Gave the best fits to more detailed relaxation data

Can be made to describe theoretical models (e.g. reaction-diffusion model)

$\circ$ $t_s = 10^2$ s
$\square$ $t_s = 10^3$ s
$\diamond$ $t_s = 10^4$ s
$\triangle$ $t_s = 10^5$ s
— $1/(1 + B\xi^\beta)$
···· $\exp(-B\xi^\beta)$
– – $1 - \beta \log(1 + B\xi)$
–·– $\beta \log(1 + B/\xi)$

$T = 125^\circ$C
$t_{ox} = 2$ nm

OTF (Denais IRPS 2006)

Universal Relaxation Function $r(\xi)$

Normalized Relaxation Time ξ

[1] Grasser *et al.*, IRPS '07

Universal Relaxation: Application to MSM Data

Universal relaxation and MSM data[1]
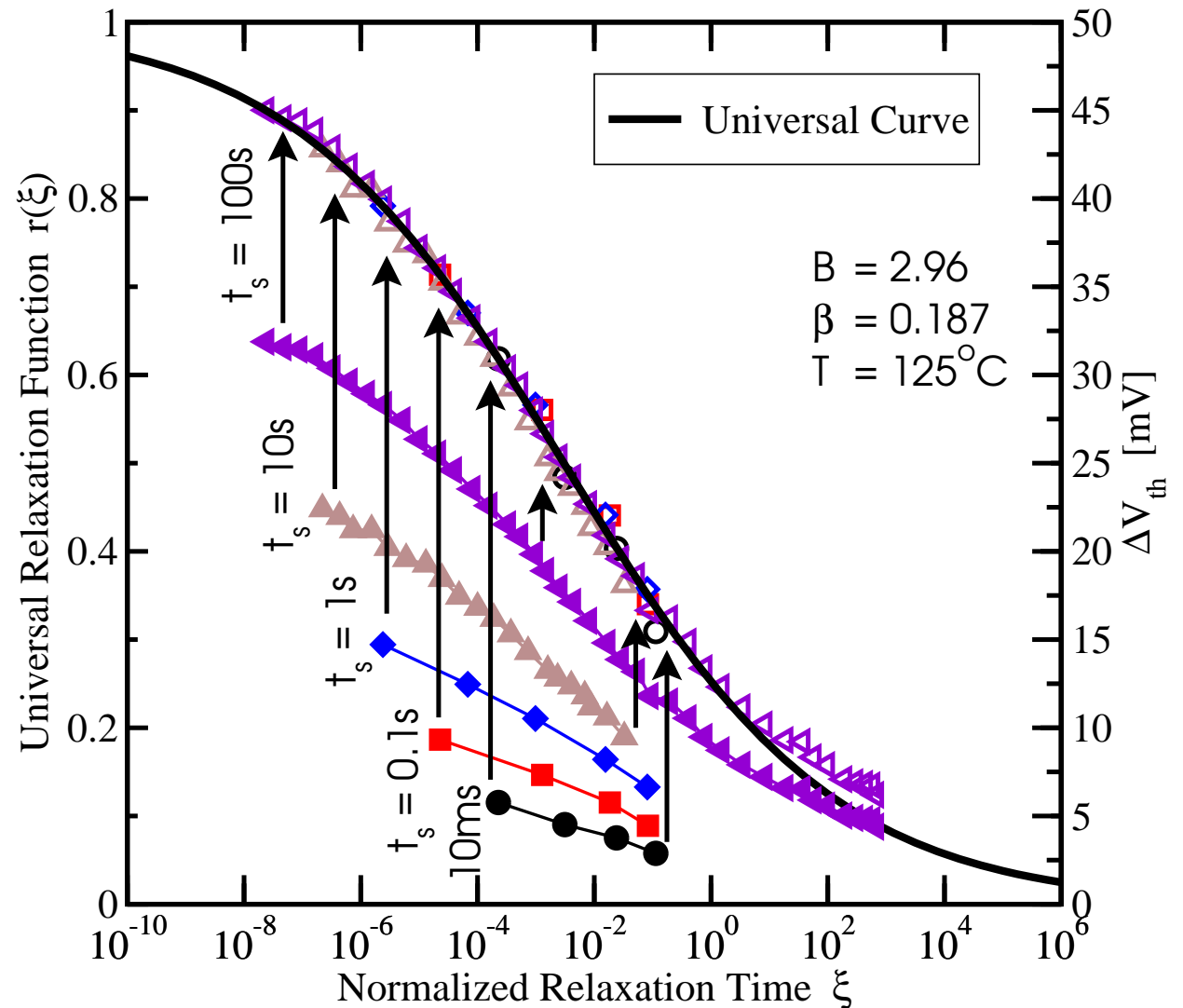
$$r(\xi) = \frac{R(t_r)}{R(0)}$$

$R(0)$ is unknown

We only have $R(t_M)$

Universal relaxation

$$R(t_M) = R(0)\ r(t_M/t_s)$$

Consequently

$$r(\xi) = \frac{R(t_r)}{R(t_M)/r(t_M/t_s)}$$

B = 2.96
β = 0.187
T = 125°C

[1] Grasser et al., IRPS '07

Universal Recovery: Permanent Component!

More detailed data reveals recoverable R and permanent component P[1]

$$S(t_s) = R(t_s, 0) + P(t_s)$$

Universal relaxation covers only recoverable component R

$$r(\xi) = \frac{R(t_s, t_r)}{S(t_s) - P(t_s)} = \frac{R(t_s, t_r)}{R(t_s, 0)}$$

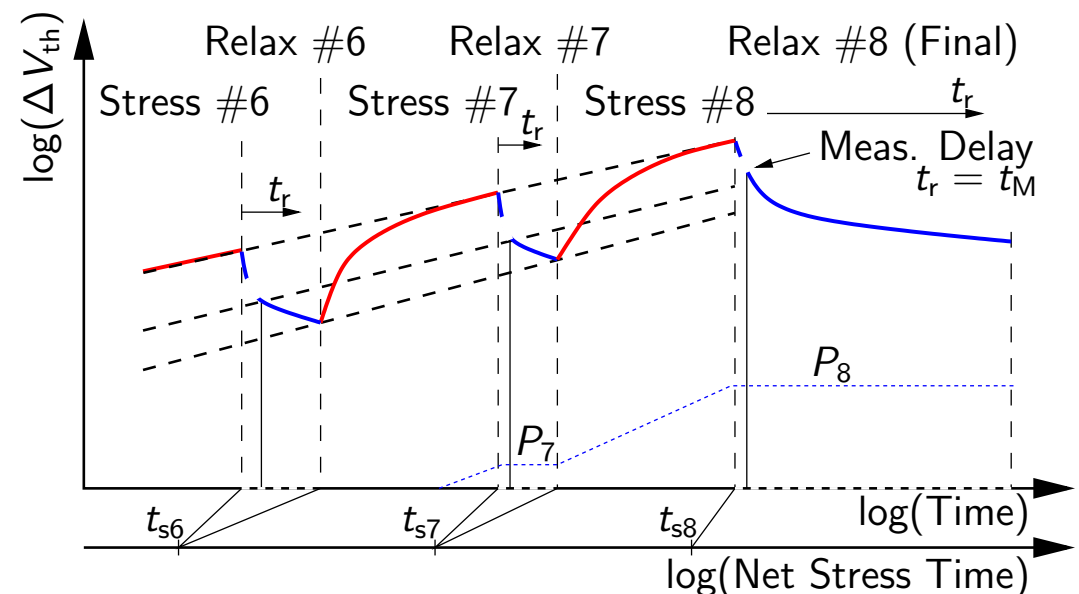Individual (constant) samples of P_i for each relaxation phase at $t_{s,i}$

$$S(t_{s,i}, t_r) = \frac{S_M(t_{s,i}, t_M) - P(t_{s,i})}{r(t_M/t_{s,i})} r(t_r/t_{s,i}) + P(t_{s,i})$$

$$= R(t_{s,i}) r(t_r/t_{s,i}) + P_i$$

Important:

No assumptions on R and P
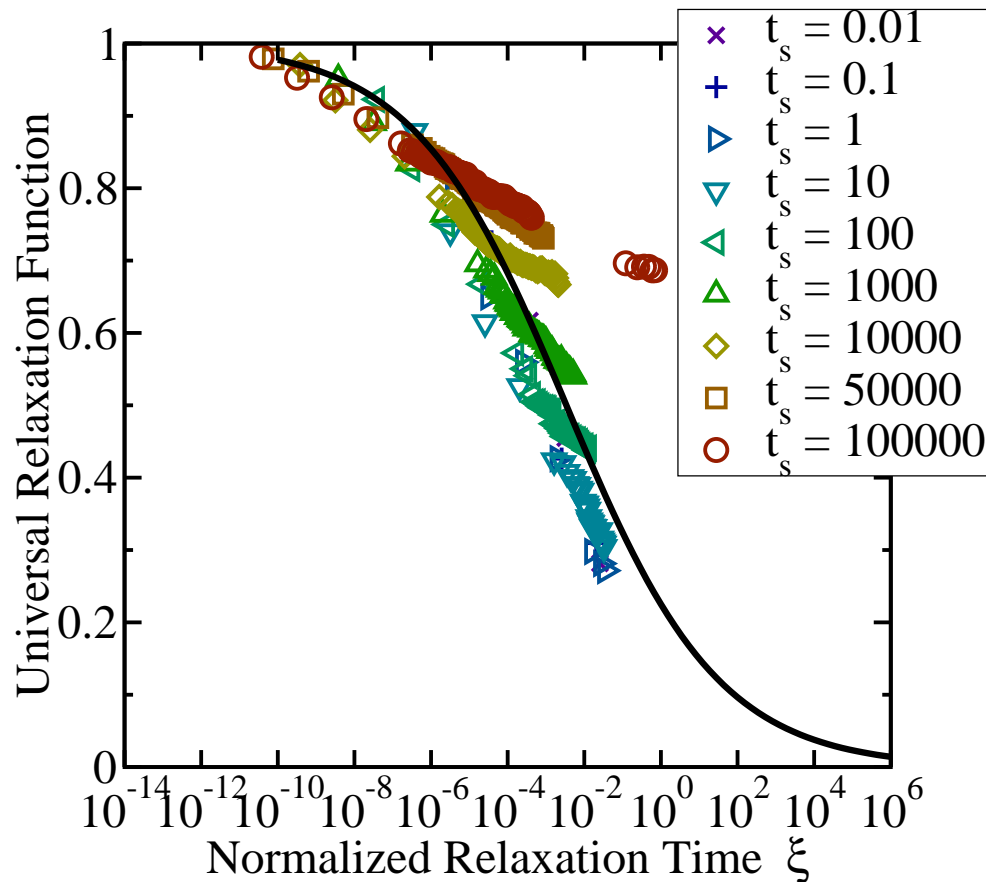
Empirical assumption on r only
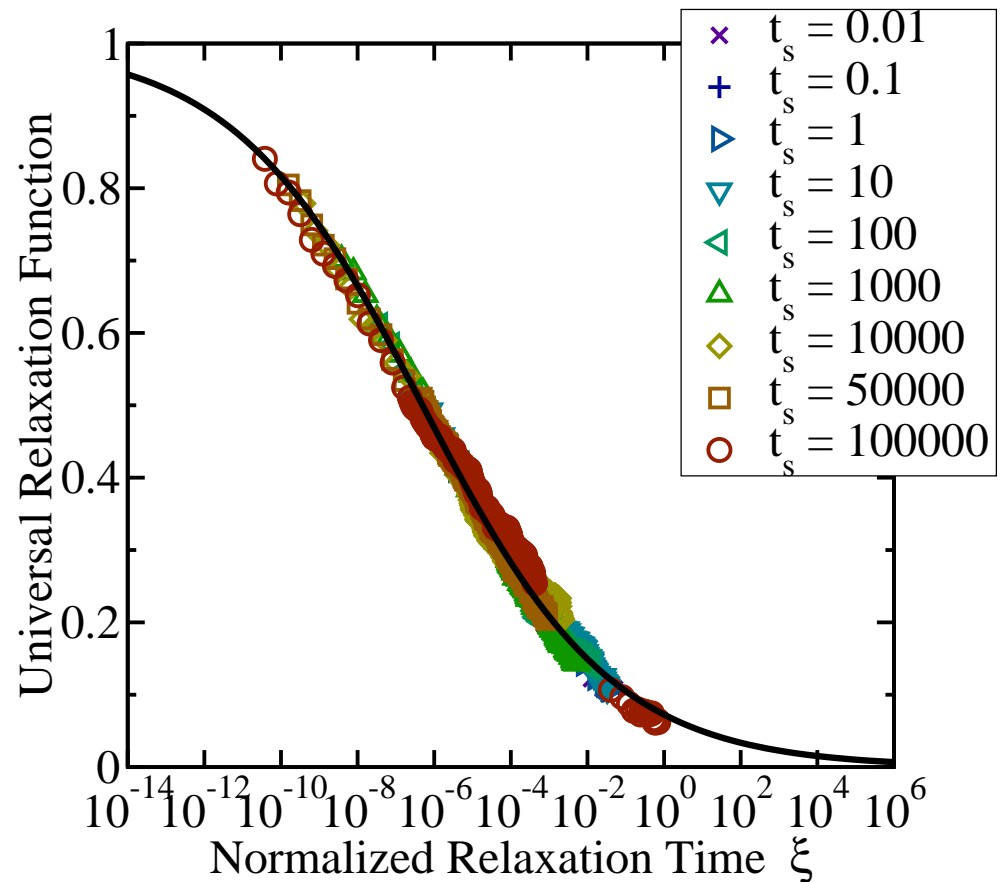

[1] Grasser *et al.*, IEDM '07

Extraction of the universal component

Now with extremely detailed relaxation data

Reconstructed time evolution during stress and recovery
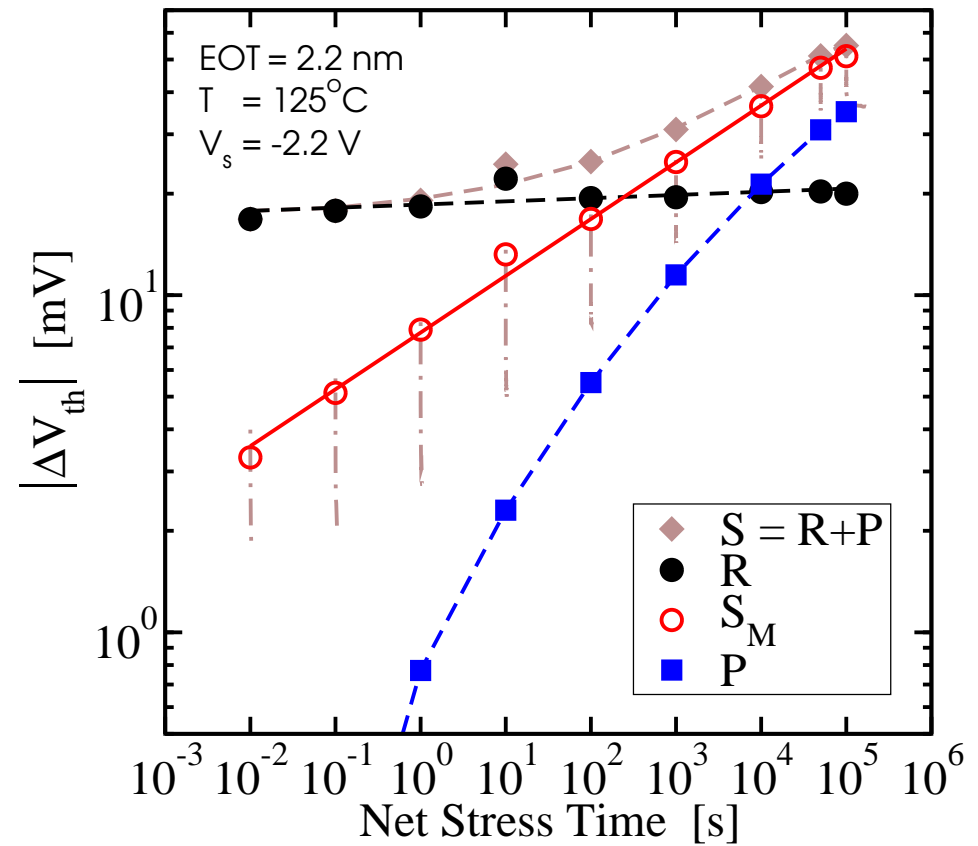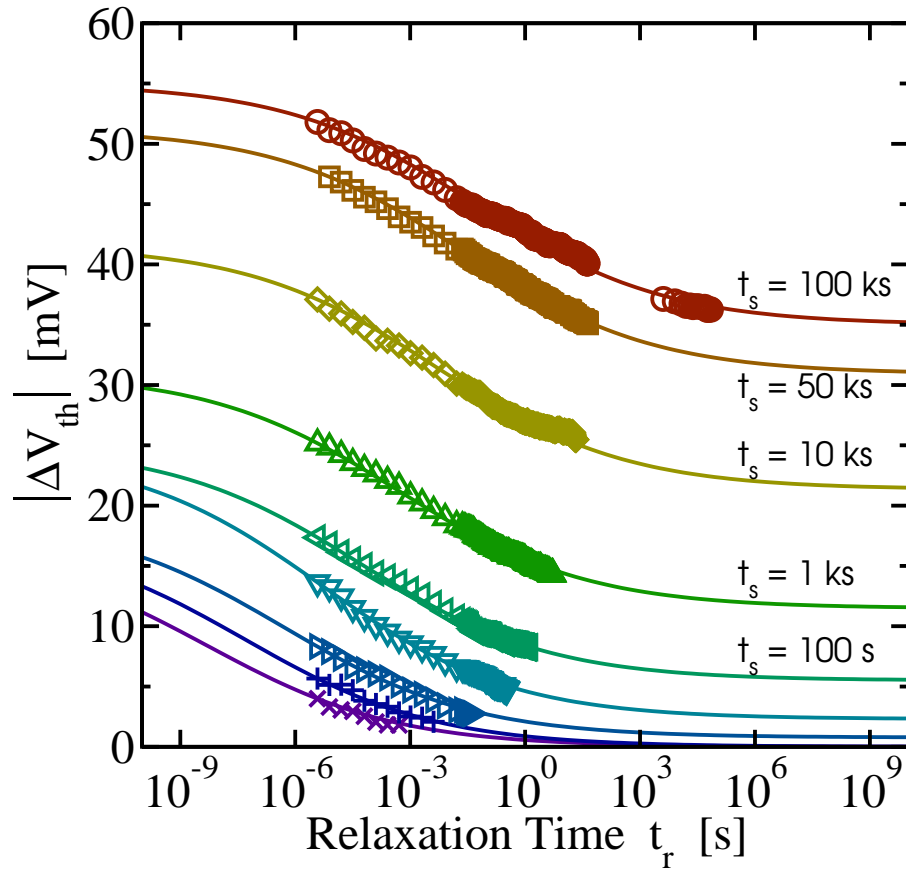
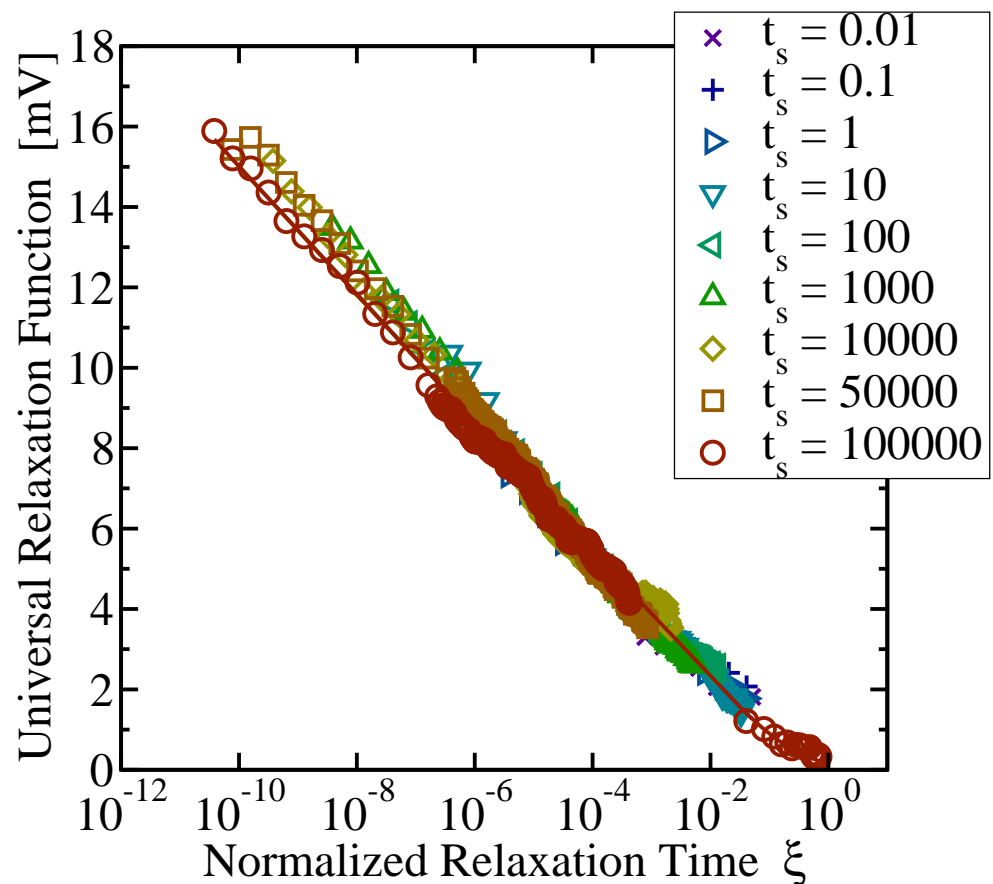# Universal Relaxation: Absolute vs. Relative

Striking contradiction: what is universal?

The relative recovery? [1] [2]

$$r(\xi) = \frac{S(t_s, t_r) - P(t_s)}{S(t_s, 0) - P(t_s)}$$

... or the absolute recovery? [3]

$$R(\xi) = S(t_s, t_r) - P(t_s)$$
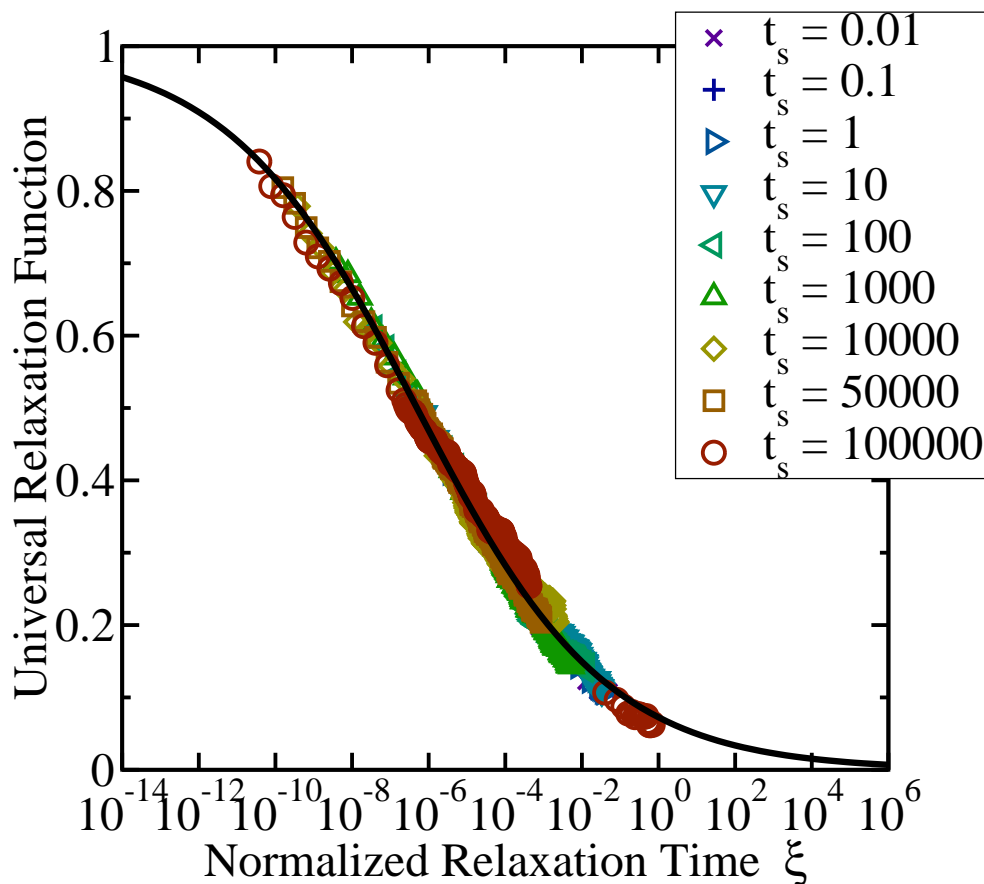
[1] Denais et al., IRPS '06   [2] Grasser et al., IEDM '07   [3] Huard et al., IEDM '07

Striking contradiction: both can be fitted to experimental data

Universality of the relative recovery?

Allows extrapolation to $t_r \to 0$, is extrapolation correct?

Universality of the absolute recovery?

Requires additional parameter for extrapolation to $t_r \to 0$, how to extract?

Both methods give roughly the same P for this case

Depends on the device and stress conditions, not true in general

How can both the relative and absolute recovery be universal?

Universality of the relative recovery

$$r(\xi) = \frac{S(t_s, t_r) - P(t_s)}{S(t_s, 0) - P(t_s)} = \frac{R(t_s, t_r)}{R(t_s, 0)}$$

Universality of the absolute recovery

$$R(\xi) = S(t_s, t_r) - P(t_s)$$

Recoverable component has a very small power-law exponent

$$R(t_s, 0) = S(t_s, 0) - P(t_s) \approx A\, t_s^{0.02}$$

EOT = 2.2 nm
T = 125°C
V_s = -2.2 V

S = R+P
R
S_M
P(relative)
P(absolute)

$|\Delta V_{th}|$ [mV]

Net Stress Time [s]

Caveats Universal Relaxation

Universal relaxation function is empirical

We need a solid theory to support it

We can only monitor a small window of the recovery behavior

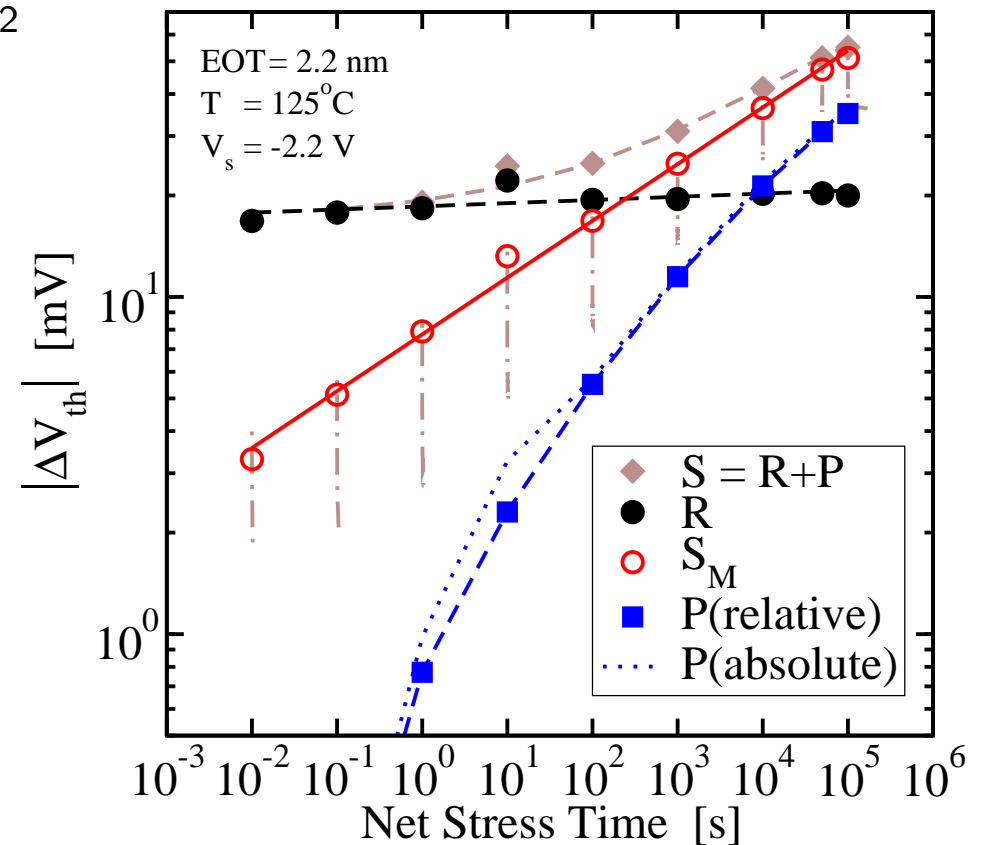Recovery may last over 12 decades in time

Based on this data alone justification of empirical models will be very difficult

For a too small window, non-universal data may appear universal

Even if future theory predicts 'perfect' universal recovery

Short time data will not be universal

Long time data will not be universal either (saturation)

Nevertheless: universality is likely a good approximation

Saturated mechanisms can definitely be ruled out

Universality rules out 'bumps' in the relaxation data

Gives good approximations for R and P from a single measurement

NBTI versus PBTI

In oxynitrides, many similarities between negative and positive stress

PBTI weaker than NBTI (at least when only the sign in V_{stress} is changed[1])

PBTI also creates positive charge (gives negative ΔV_{th} shift, just as NBTI)

Relaxation after PBT stress is universal, but somewhat different to NBT recovery
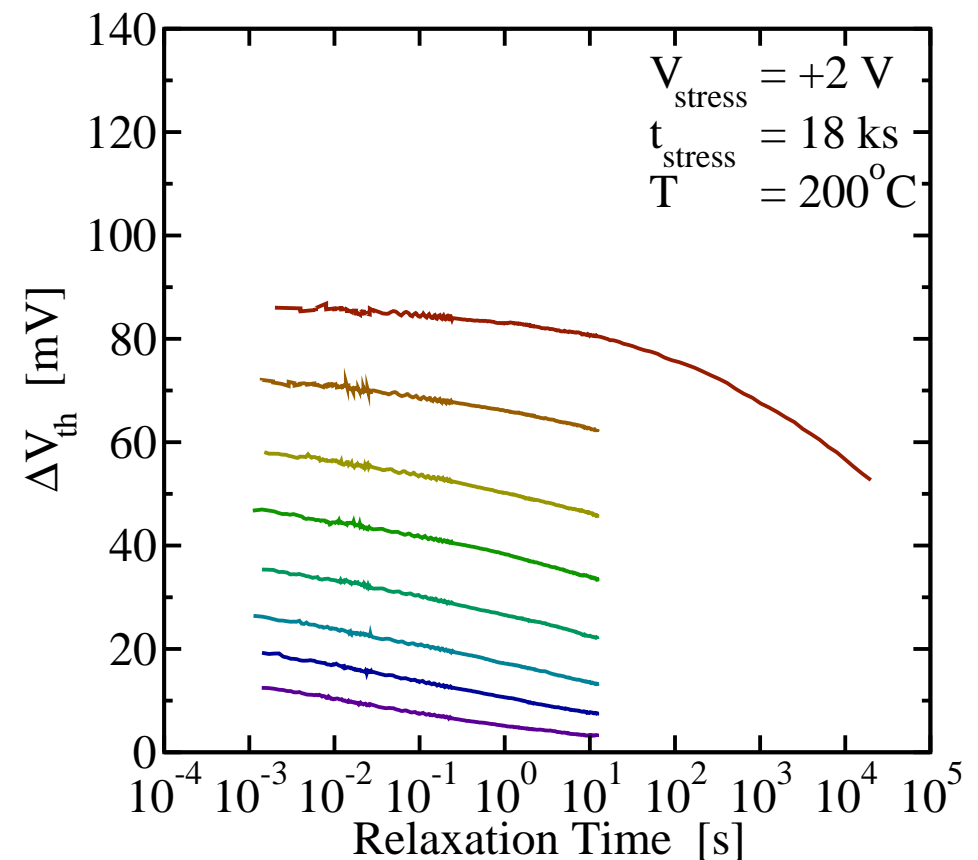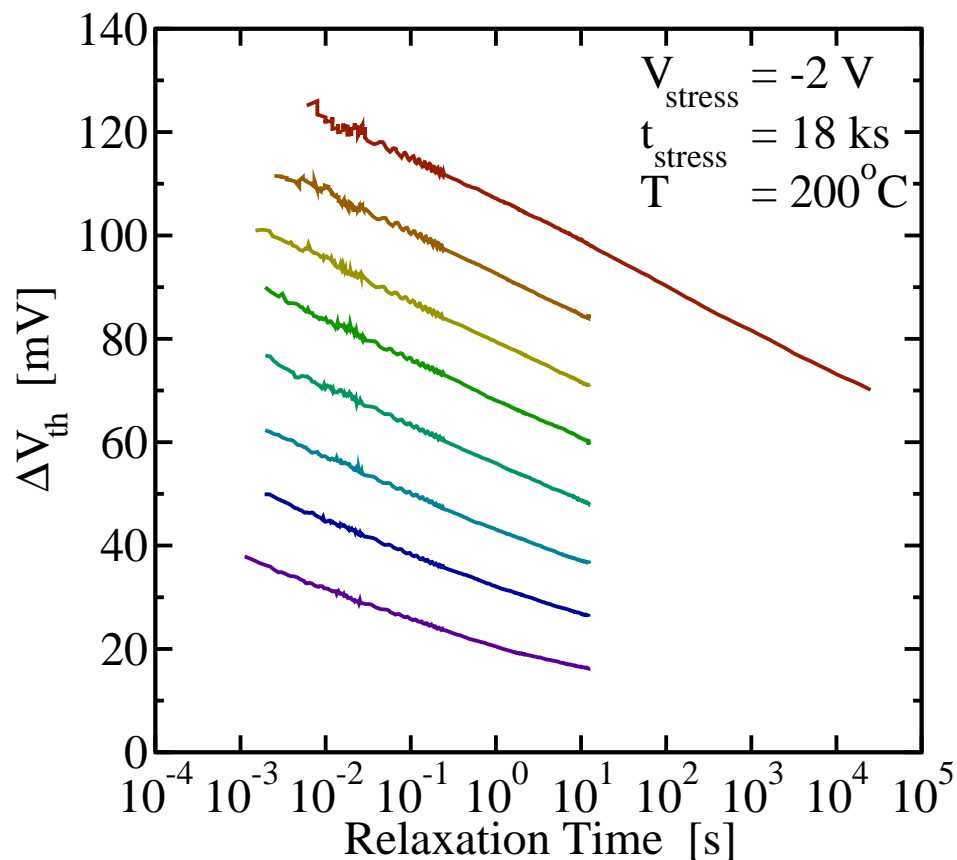
[1] Correction for flatband voltage required, see Alam and Mahapatra, MR '05

Bias Dependent Recovery

Frequently used assumption[1][2]: $R \doteq$ trapped holes

$\Rightarrow$ accelerated recovery at positive bias (electron injection, hole annihilation)
Suggested as a method to efficiently extract P, looks consistent at first
However, devices degrade after application of positive bias!!! [3][4]
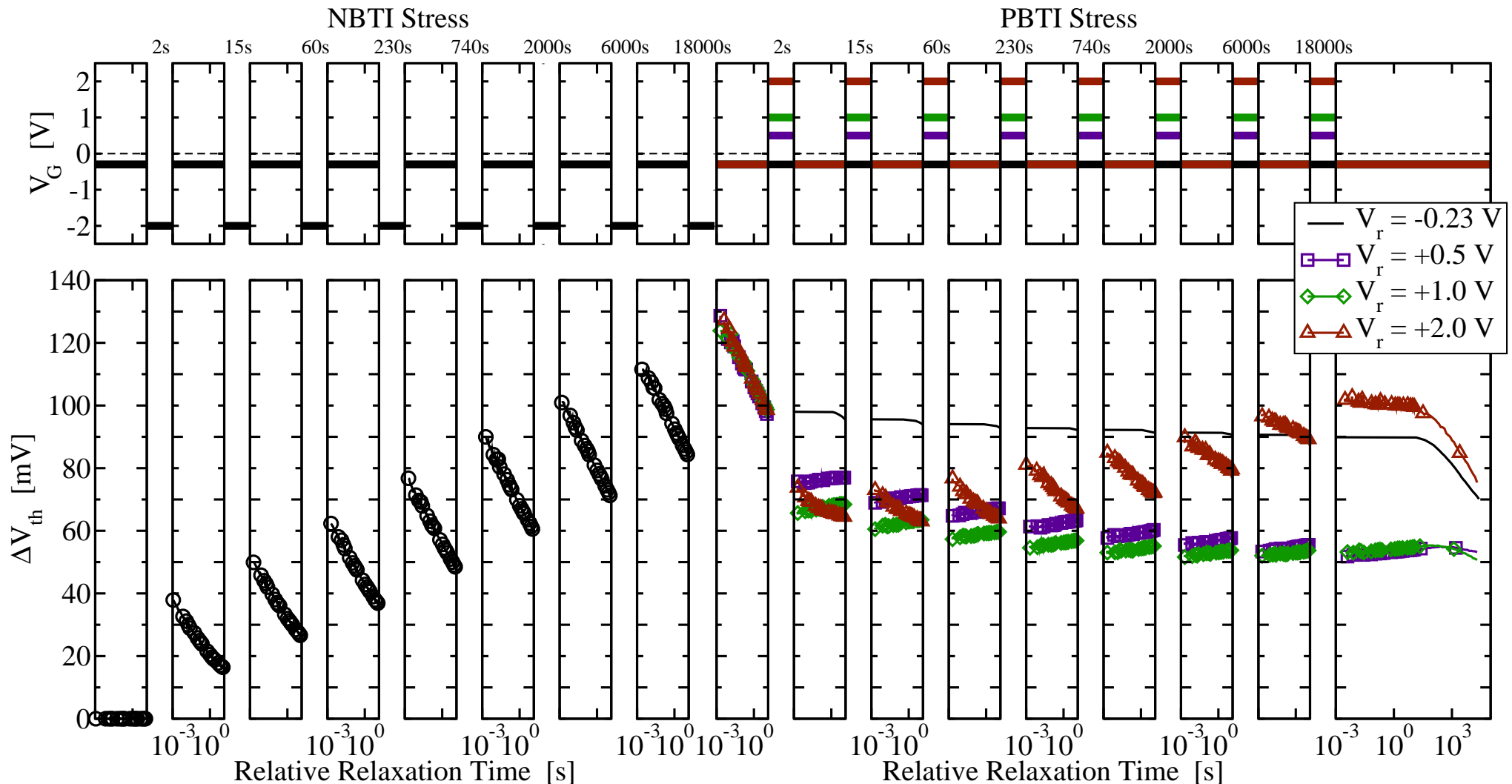
[1] Huard *et al.*, IEDM '07   [2] Shen *et al.*, IEDM '06   [3] Grasser *et al.*, IEDM '07   [4] Grasser *et al.*, IRPS '08

Duty Factor Dependence

Strong duty factor dependence[1][2]

Strong impact for DF → 100% and DF → 0%

Weak impact for DF around 50% (plateau)

The following are MSM measurements with a delay of 1 ms
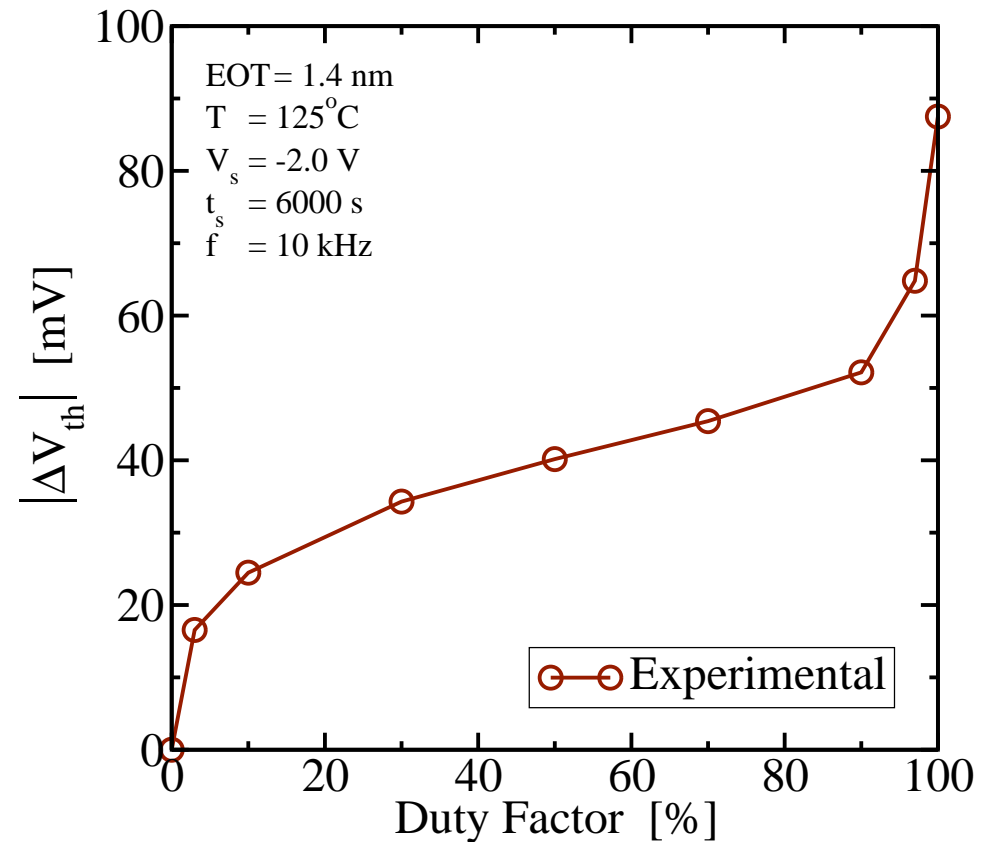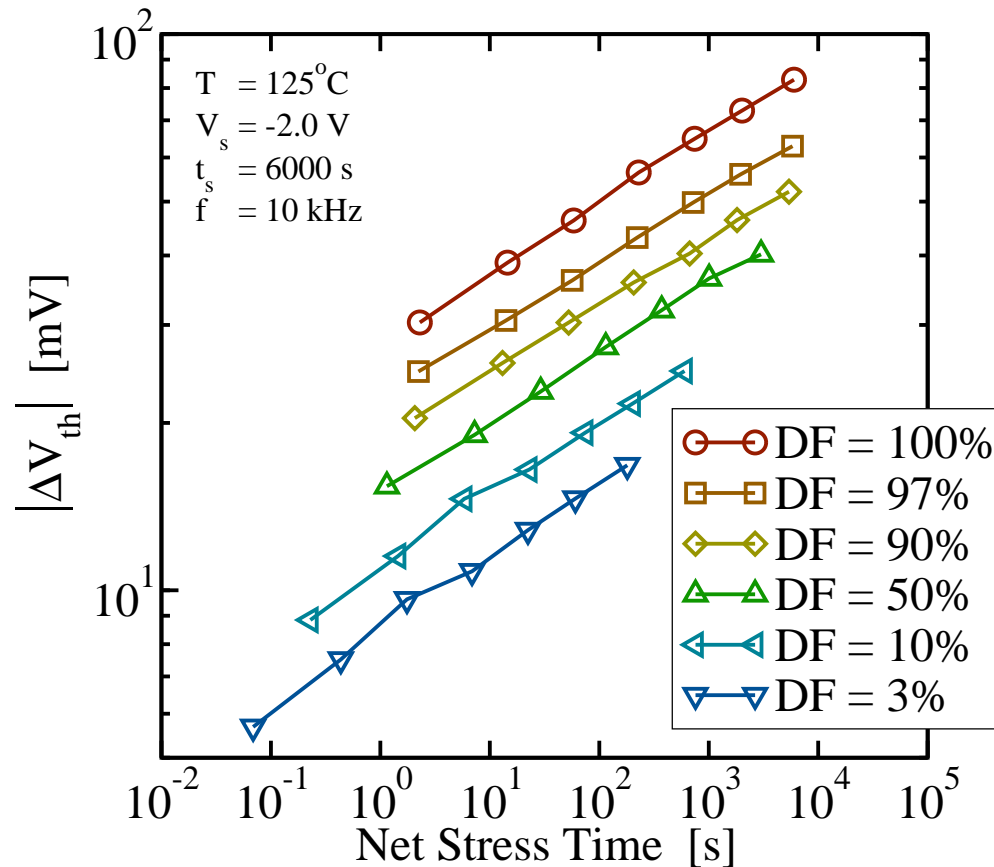
[1] Fernandez et al., IEDM '06   [2] Grasser et al., IEDM '07

Measurement delay has a significant impact on measurement[1] [2] [3]

Curvature in data becomes more obvious, larger (time-dependent) 'slope'

Impact of delay does not disappear at longer stress times

Impact of delay is temperature dependent
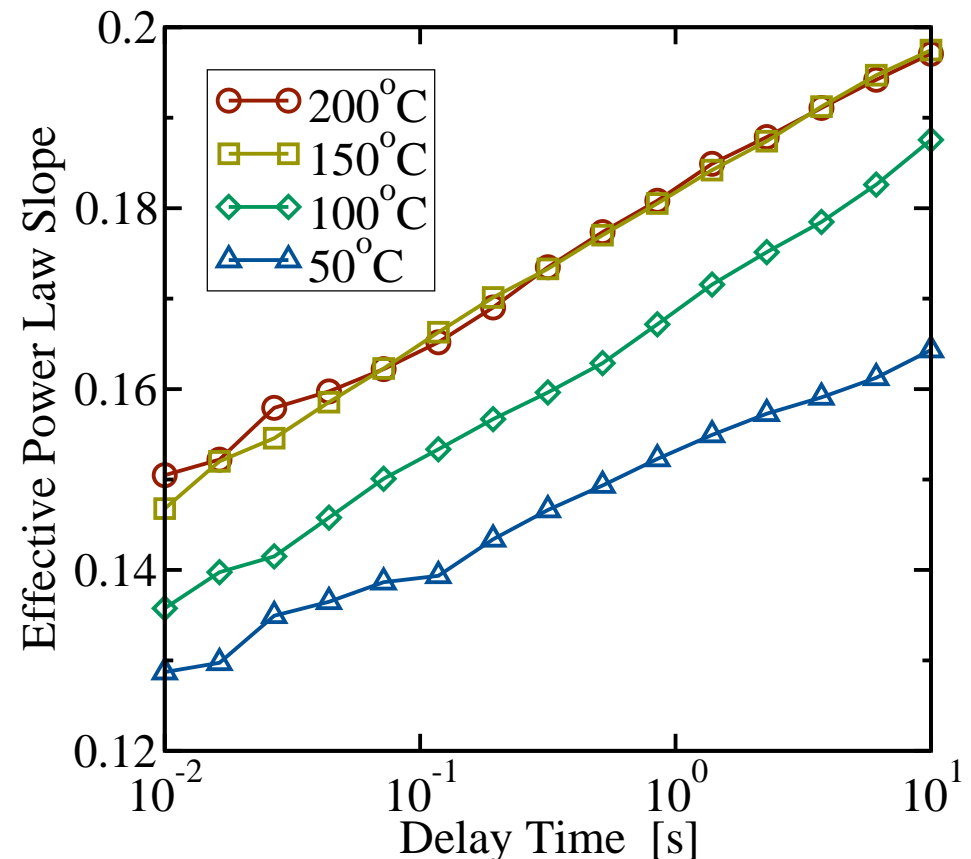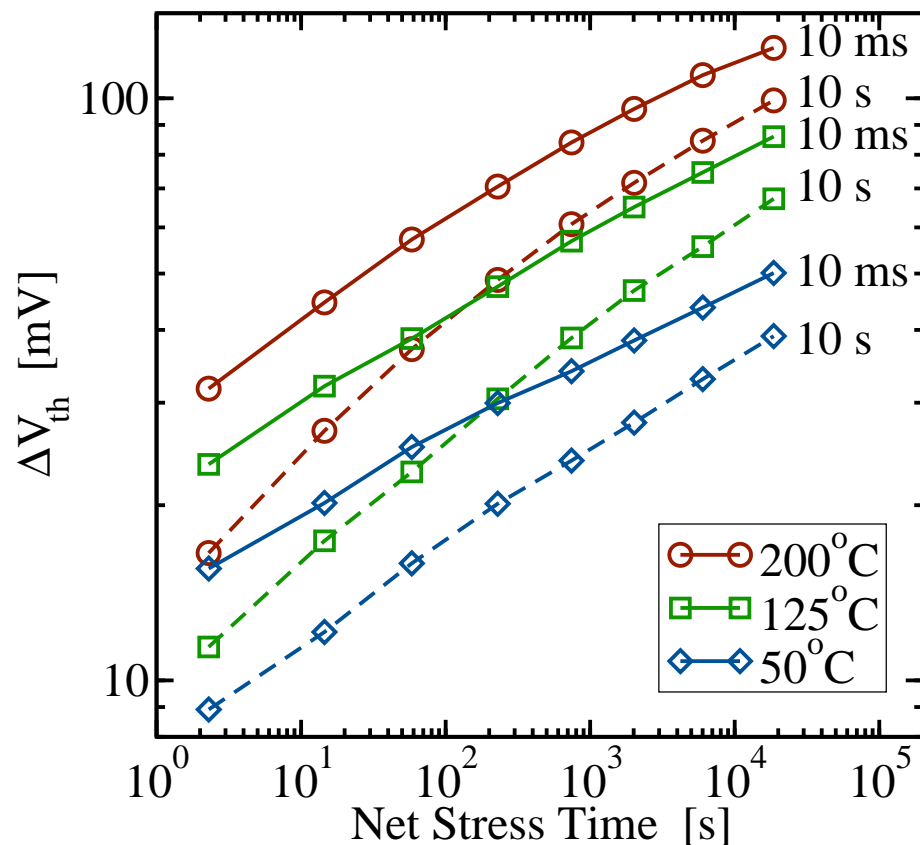
[1] Ershov *et al.*, IRPS '03   [2] Denais *et al.*, IEDM '04   [3] Kaczer *et al.*, IRPS '05

Analytic expression for delayed data (without P)[1]

$$S(t, t_M) = \frac{A \, t^n}{1 + B(t_M/t)^\beta}$$

Analyze properties of delayed data

MSM slope is time-dependent

$$n_M(t, t_M) = \frac{d \log S(t, t_M)}{d \log t}$$

$$= n + \frac{\beta B}{B + (t/t_M)^\beta}$$

How define an 'average slope'?

Slope is temperature-dependent

Slope is technology-dependent

Measurement issues?

Slope is not a good parameter to characterize NBTI

Only useful for a rough characterization, say $n = 0.25$ vs. $n = 0.1$

Should not be used for life-time extrapolation (worst-case estimate?)
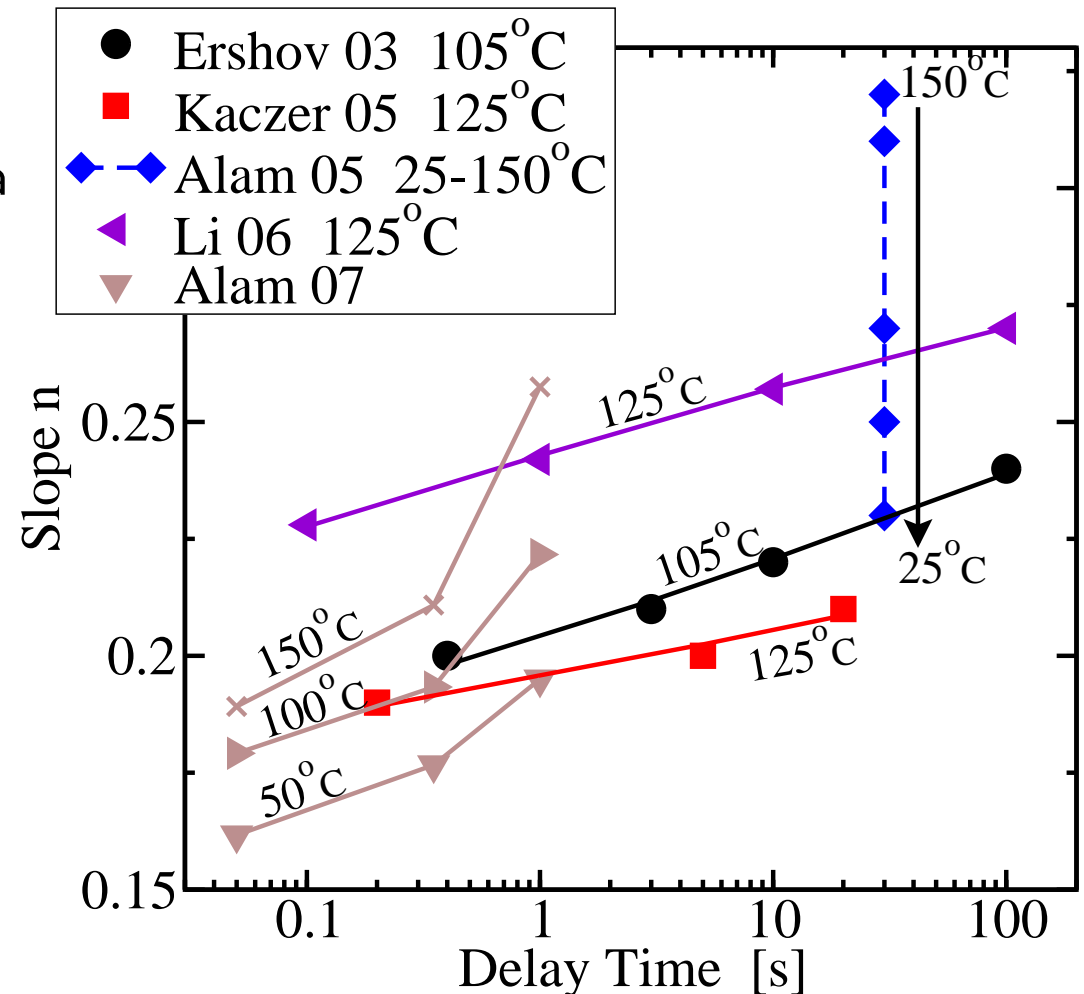
[1] Grasser *et al.*, IRPS '07

Additional Issues

Most reports seem to indicate that NBTI is frequency independent

  At least up to $1\,MHz$ [1]

  Or is it even up to $1\,GHz$? [2]

  Or is there a frequency dependence after all? [3]

  What is the impact of the max/min oscillations on the measurement [4]

Gate oxide thickness dependence

Gate oxide material dependence (nitrogen, fluoride) [5] [6] [7]

NBTI may be independent of the gate material [8]

What do charge-pumping and DCIV measurements see

  Compared to OTF and MSM

  Rigorous link to models missing

And many more fun things to worry about ...

[1] Alam, IEDM '03   [2] Fernandez *et al.*, IEDM '06   [3] Krishnan *et al.*, IEDM '05   [4] Shen *et al.*, IEDM '06

[5] Mahapatra *et al.*, IRPS '07   [6] Mitani *et al.*, IRPS '07   [7] Kumar *et al.*, IEDM '07   [8] Zafar *et al.*, VLSI Symp '06

Model Checklist

Stress: undelayed measurements

Log-like initial behavior

Temperature and voltage independent slope, scaling property

Stress: delayed measurements

Temperature-dependent power-law exponent/slope (proportional to temperature?)

Delay distorts data even at larger stress times

Recovery

Long relaxation tails of log-like nature (more than 12 decades in time?)

Strong bias sensitivity, particularly for positive bias

Universal behavior (or at least in good approximation)

Fast recoverable and slowly recovering/permanent component

Duty factor dependence

Plateau around DF $= 50\%$, strong sensitivity for very large and small DF

May be a matter of taste, just a few nasty points included ...

Published NBTI models

Can they do all this?

General Remarks

All models assume breaking of Si–H bonds $\Rightarrow$ dangling bonds (P_b centers)

Two-step process
Reaction rates for bond breaking/annealing
Subsequent hydrogen diffusion/removal

Which step determines the time-dynamics?
Is NBTI reaction-limited?
Is NBTI diffusion-limited? If so, how to model hydrogen diffusion?

Has been described using a generalized reaction-diffusion framework [1] [2]

Many authors have suggested hole-trapping to be important[3] [4] [5] [6]

Might explain initial log(t) behavior

Might explain fast initial recovery, or even very slow component

Which aspect of NBTI can the models describe? Which not?

All models have been implemented in a numerical device simulator solver[7]

Avoid uncertainties in approximate analytic expressions

Study the predictions during relaxation where analytic expressions are rarely given

[1] Alam *et al.*, MR '06   [2] Grasser *et al.*, TDMR '08   [3] Huard *et al.*, IEDM '07   [4] Shen *et al.*, IEDM '06

[5] Reisinger *et al.*, IRPS 06   [6] Ang and Wang, EDL '06   [7] Minimos-NT, IµE, TU-Vienna

Standard Model: Reaction-Diffusion Model

Classic reaction-diffusion model for hydrogen-related species H[1][2][3][4][5]

Interface reaction:

$$\frac{\partial N_{it}}{\partial t} = k_f(N_0 - N_{it}) - k_r N_{it} H_{it}^{1/a}$$

Boundary condition:

$$\frac{1}{a}\frac{\partial N_{it}}{\partial t} = \frac{\partial H_{it}}{\partial t} = \mathbf{F}_H \cdot \mathbf{n}$$

Bulk transport:

$$\frac{\partial H}{\partial t} = -\nabla \cdot \mathbf{F}_H$$

Particle flux:

$$\mathbf{F}_H = -D_c\left(\nabla H - \frac{Z_H H}{V_T}\mathbf{E}\right)$$

a is kinetic exponent: $a = 1$ for H^0 and H^+, $a = 2$ for H_2

Z_H is particle charge: $Z_H = 0$ for H^0 and H_2, $Z_H = 1$ for H^+

[1] Jeppson and Svensson, JAP '77   [2] Ogawa and Shiono, PRB '95   [3] Alam, IEDM '03
[4] Chakravarthi et al., IRPS '04   [5] Alam and Mahapatra, MR '05
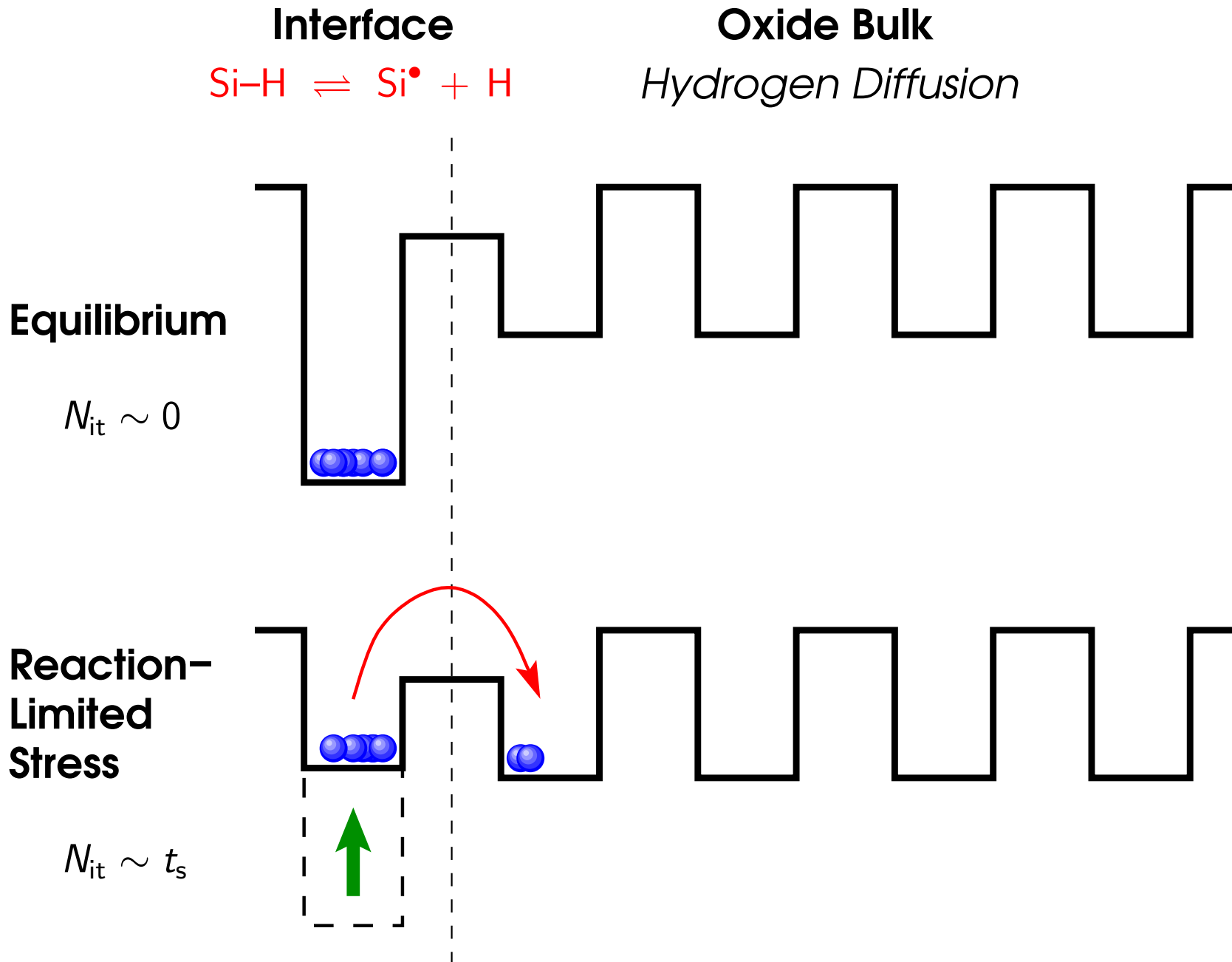
Interface

Si–H $\rightleftharpoons$ Si$^\bullet$ + H

Oxide Bulk

Hydrogen Diffusion

Equilibrium

$N_{it} \sim 0$

**Reaction-
Limited
Stress**

$N_{it} \sim t_s$

H Diffusion

Balance

**Diffusion–
Limited
Stress**

$N_{it} \sim t_s^{1/4}$

H Back Diffusion

**Diffusion–
Limited
Relaxation**

$N_{it} \sim \dfrac{1}{1 + \sqrt{\dfrac{t_r}{t_s}}}$

39

Interface

Oxide Bulk

Equilibrium

$N_{it} \sim 0$

● *Atomic Hydrogen*

⬤⬤ *Molecular Hydrogen*

Reaction-Limited Stress

$N_{it} \sim t_s$

H Diffusion Barrier

H$_2$ Diffusion Barrier

Equilibrium (Mass Action Law)
H + H ⇌ H$_2$

Balance

H$_2$ Diffusion

Diffusion-Limited Stress

$N_{it} \sim t_s^{1/6}$

H$_2$ Diffusion Barrier

H$_2$ Back Diffusion

Diffusion-Limited Recovery

$N_{it} \sim \dfrac{1}{1 + \sqrt{\dfrac{t_r}{t_s}}}$

H$_2$ Diffusion Barrier
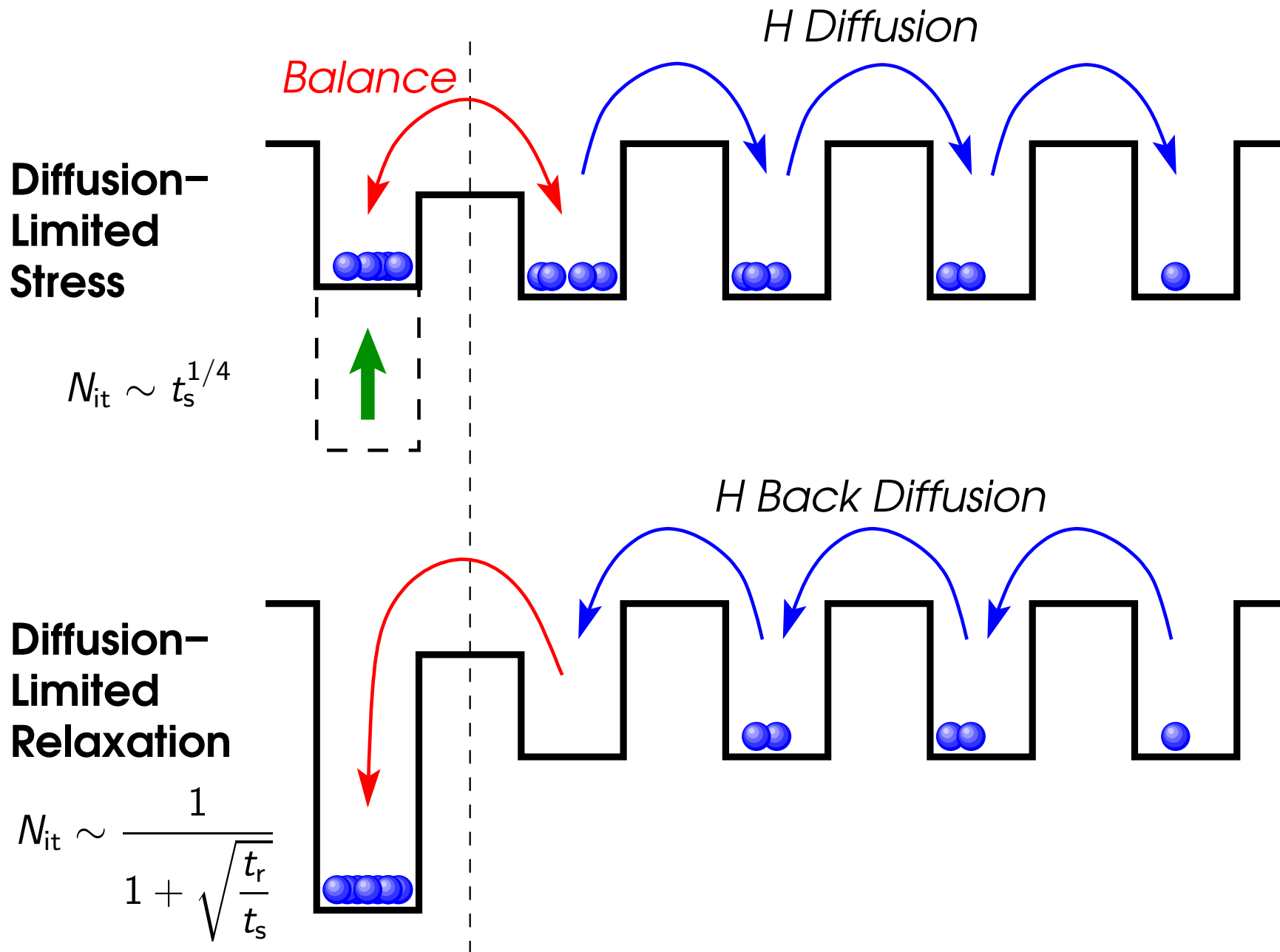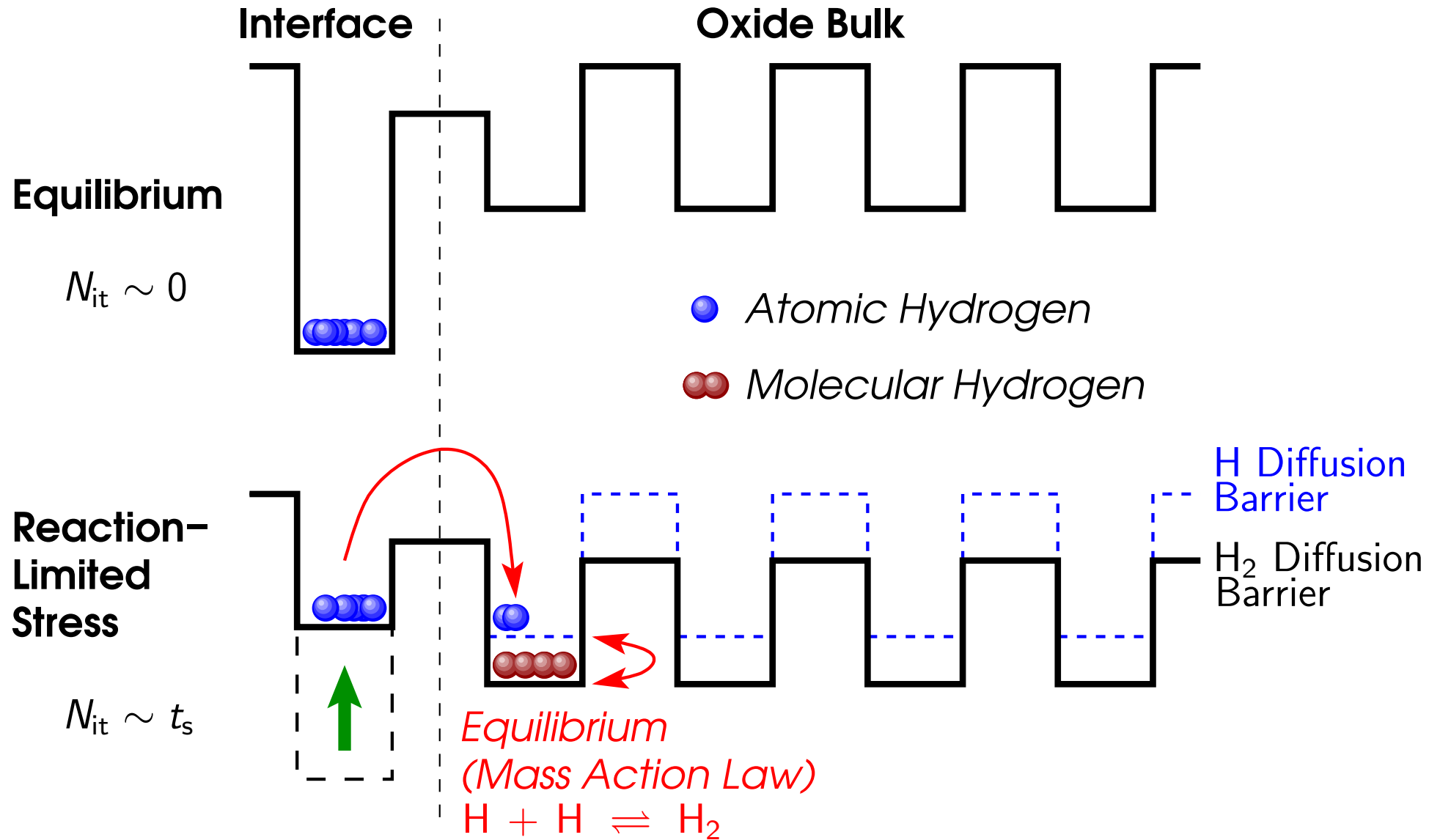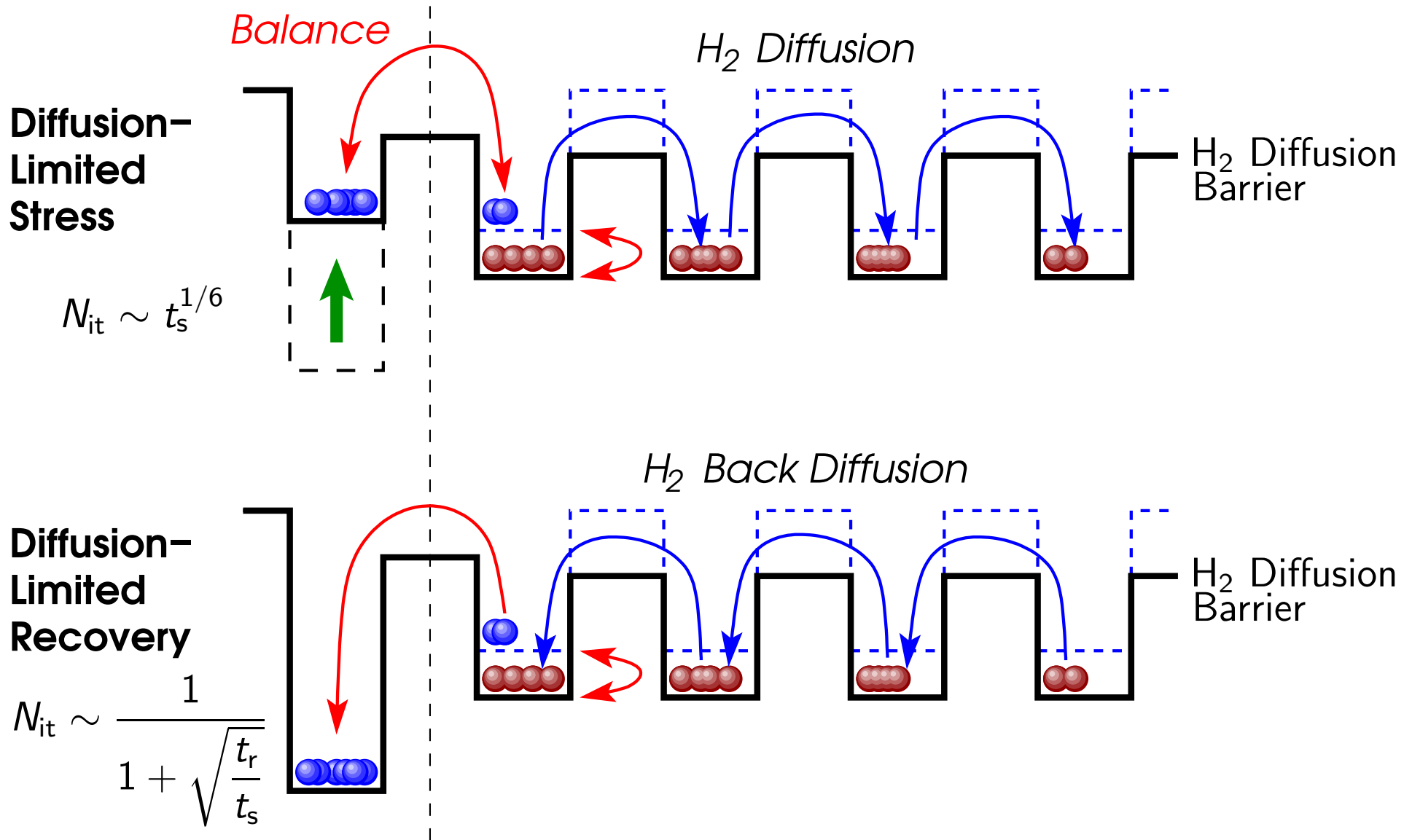
Reaction-Diffusion Model: Stress Behavior

Analytic 1D solution in the diffusion-limited regime $\partial N_{it}/\partial t \approx 0$ gives[1][2]

H^+: $N_{it}(t) \propto A\, t^{1/2}$ (not observed experimentally)

H^0: $N_{it}(t) \propto A\, t^{1/4}$

H_2: $N_{it}(t) \propto A\, t^{1/6}$
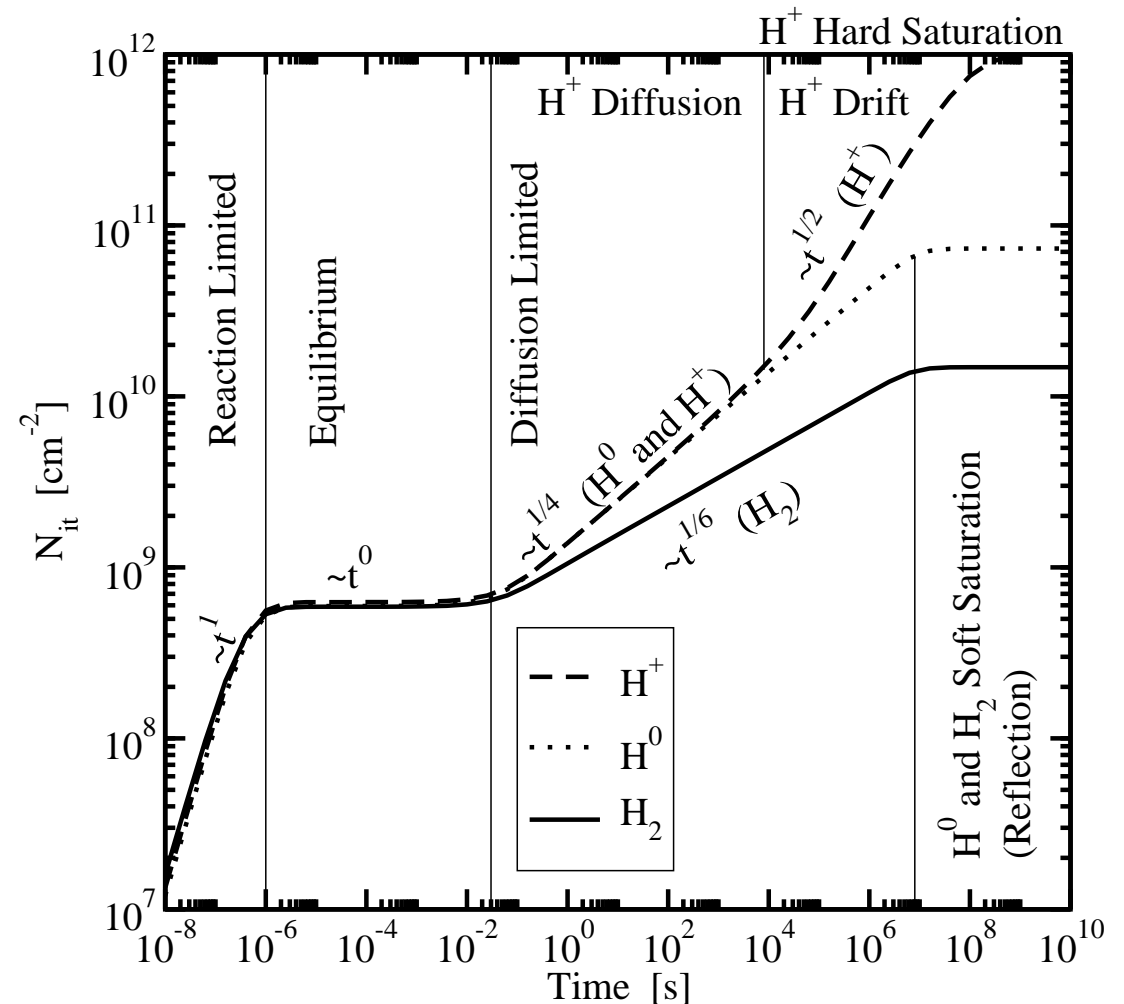
Gives fixed slopes

Fractional power-law

Only abrupt saturation possible

Field-dependent k_f and k_r [3]

Responsible for soft-saturation?

[1] Chakravarthi *et al.*, IRPS '04   [2] Alam and Mahapatra, MR '05   [3] Islam *et al.*, TED '07

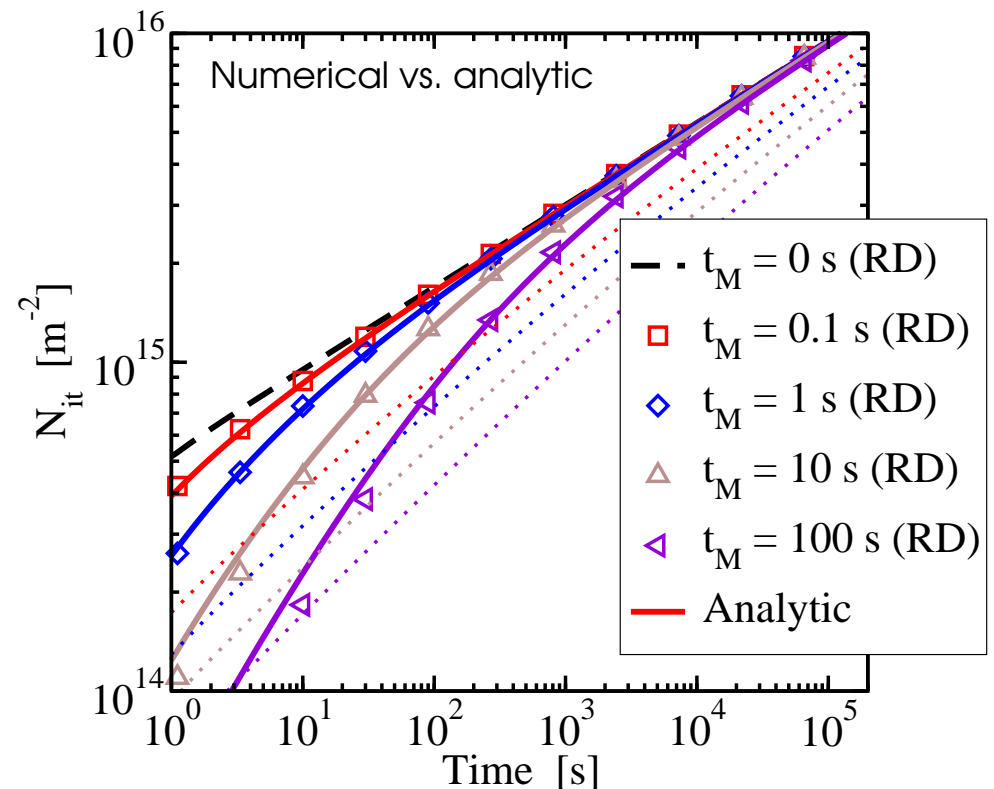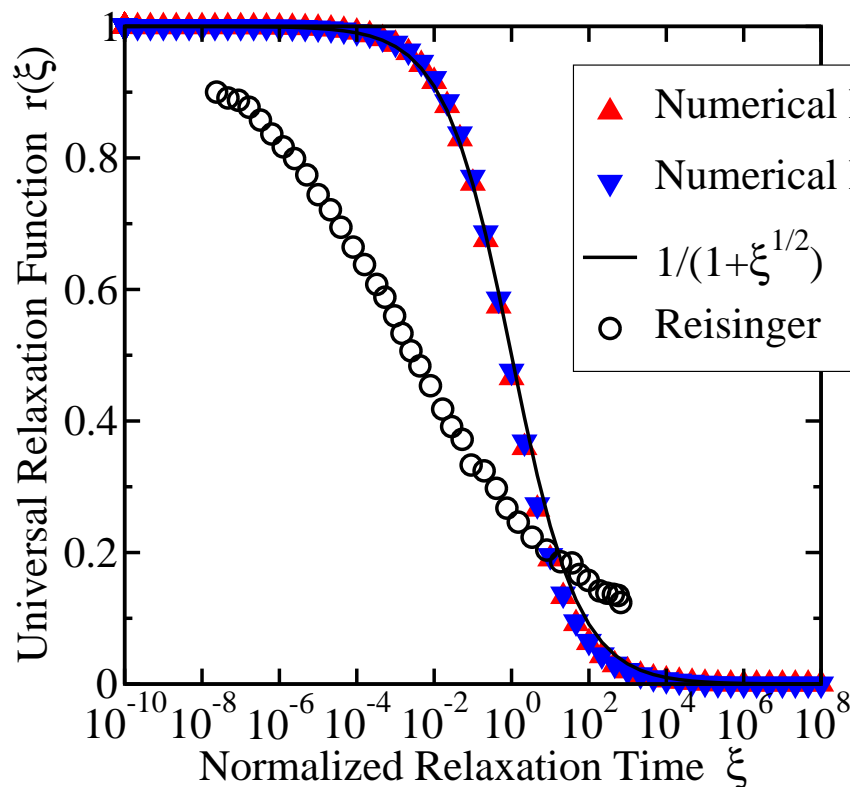Relaxation predicted by the reaction-diffusion model is universal[1]

$$r(\xi) = \frac{\Delta V_{th}(t_s, t_r)}{\Delta V_{th}(t_s, 0)} \approx \frac{1}{1 + \xi^{1/2}}$$ Independent of all model parameters/species!

Poor agreement with measurement data (cannot be fitted!)

Strong time dependence in slope, minimal influence already at $t_s > 10^3$ s

Rule of thumb: 3 decades and 50% recovery at $t_r = t_s$ ($\xi = 1$)
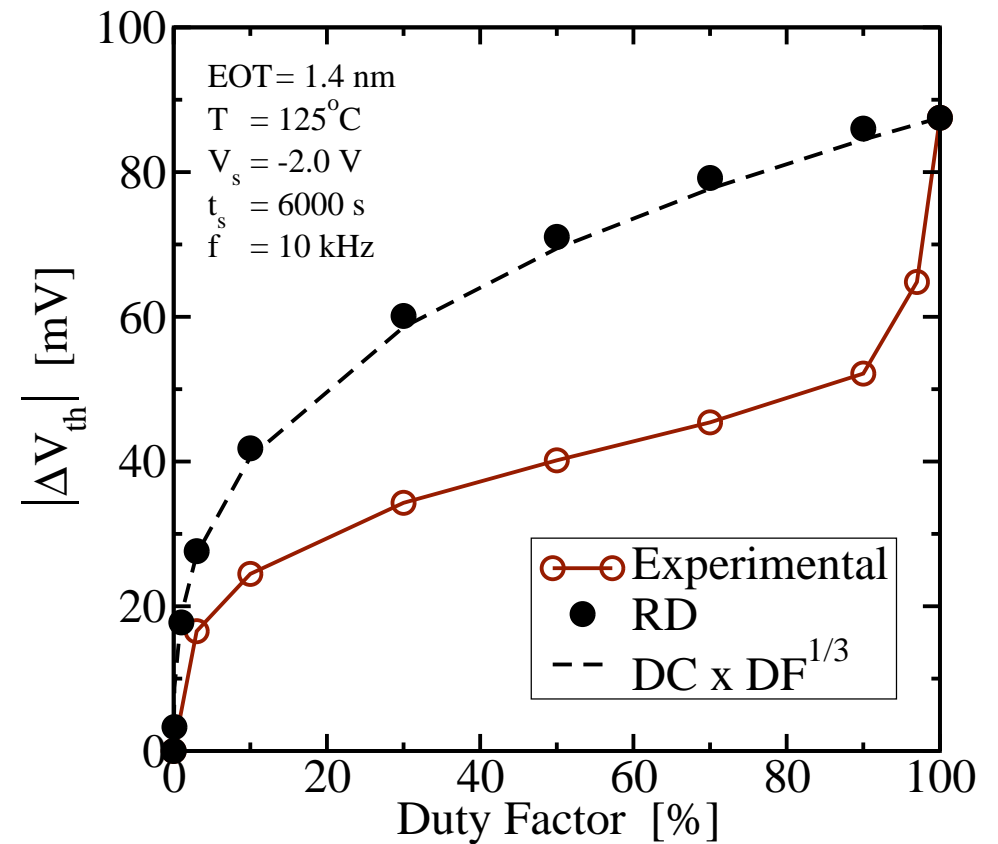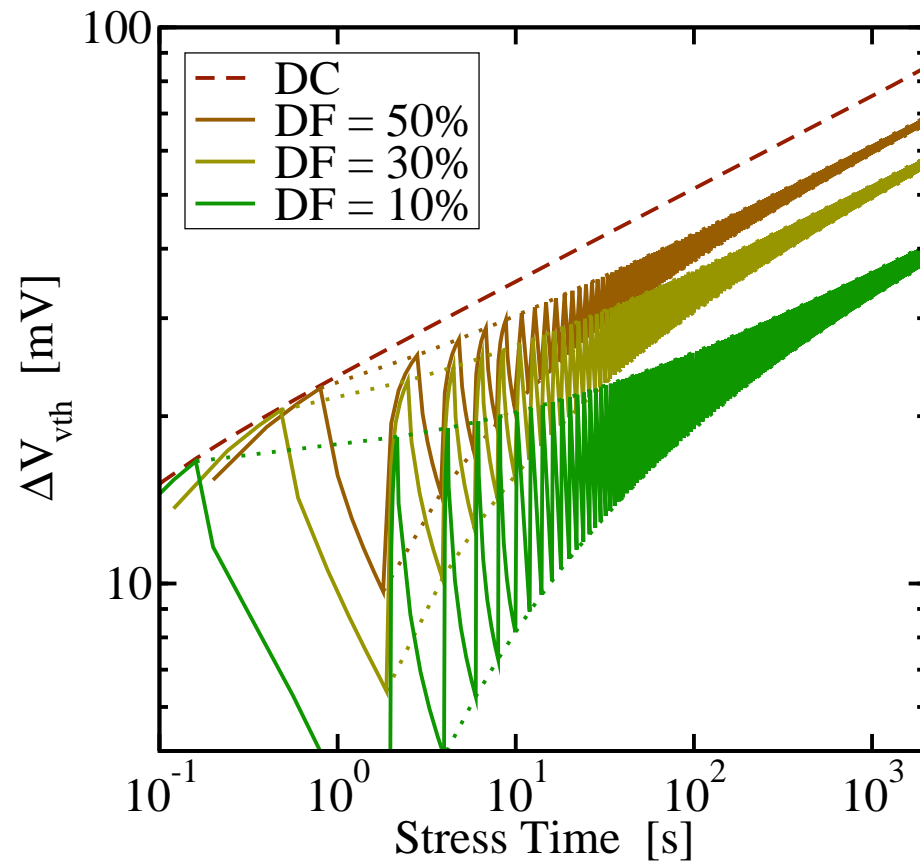
Duty Factor Dependence

Reaction-diffusion model predicts weak duty factor dependence

Weak impact around DF $\rightarrow$ 100%, contrary to experiments[1] [2] [3]

No plateau, roughly follows DF$^{1/3}$

[1] Fernandez et al., IEDM '06   [2] Grasser et al., IEDM '07   [3] Huard et al., IEDM '07

Stress: undelayed measurements

   ✗ Log-like initial behavior

   ✓ Temperature and voltage independent slope, scaling property

Stress: delayed measurements

   ✗ Temperature-dependent power-law exponent/slope (proportional to temperature?)

   ✗ Delay distorts data even at larger stress times

Recovery

   ✗ Long relaxation tails of log-like nature (more than 12 decades in time?)

   ✗ Strong bias sensitivity, particularly for positive bias

   ✓ Universal behavior (or at least in good approximation)

   ✗ Fast recoverable and slowly recovering/permanent component

Duty factor dependence

   ✗ Plateau around DF = 50%, strong sensitivity for very large and small DF

Extended Reaction-Diffusion Models

A number of extensions to basic RD theory have been proposed

Variant A: different diffusion coefficient in the oxide and the poly gate[1]

    Motivation: modern oxides are too thin: hydrogen diffusion profile meaningful?

    Would predict 'quasi-saturation', but same long-term slope during stress

Variant B: two interfaces with trap generation on both sides[2]

    Motivation: increase in SILC

    First interface releases H^0, H_2 creation at second interface

Variant C: explicit conversion between H^0 to H_2[3]

    Motivation: no instant dimerization
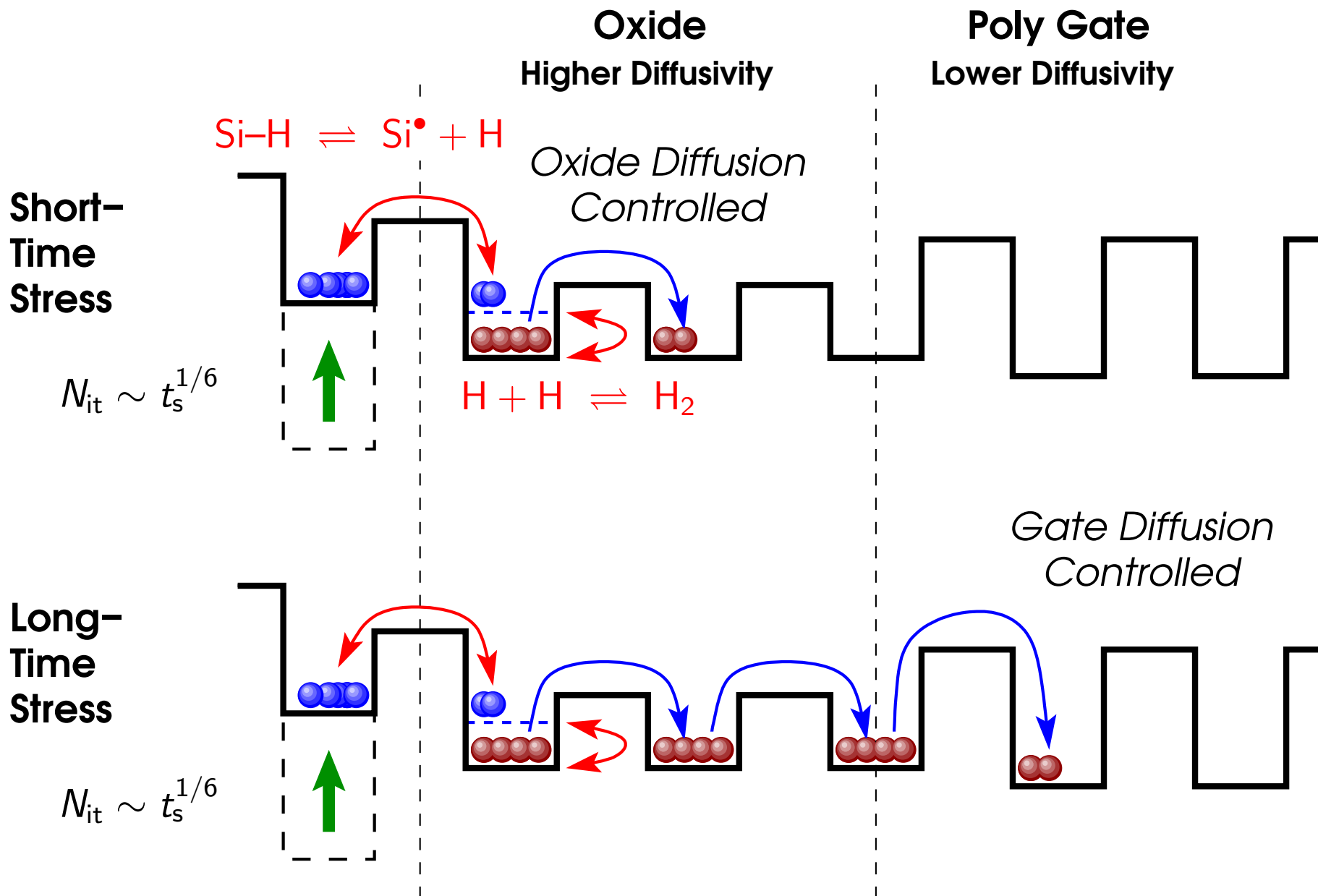
    Gives an improvement on the initial response

Variant D: consideration of the dynamic interface state occupancy[4]

    Motivation: can the occupancy follow a fast switch from $V_G = V_{stress}$ to $V_G = V_{th}$?

[1] Alam *et al.*, ECS '05   [2] Krishnan *et al.*, IEDM '05   [3] Alam *et al.*, MR '07   [4] Islam *et al.*, IEDM '07

Two-Region Reaction-Diffusion Model: Schematics – I

Variant A: different diffusion coefficient in the oxide and the poly gate[1]

Oxide
Higher Diffusivity

Poly Gate
Lower Diffusivity

$Si–H \rightleftharpoons Si^{\bullet} + H$

Short–Time Stress

Oxide Diffusion Controlled

$N_{it} \sim t_s^{1/6}$

$H + H \rightleftharpoons H_2$

Long–Time Stress

Gate Diffusion Controlled

$N_{it} \sim t_s^{1/6}$

[1] Alam *et al.*, ECS '05

Oxide

Poly Gate

Short-Time Recovery

Oxide Diffusion Controlled

$$N_{it} \sim \frac{1}{1 + \sqrt{\dfrac{t_r}{t_s}}}$$

Long-Time Recovery

Gate Diffusion Controlled

The initial stress phases are identical to the standard RD model

When H_2 diffusion front reaches the poly gate: quasi-saturation

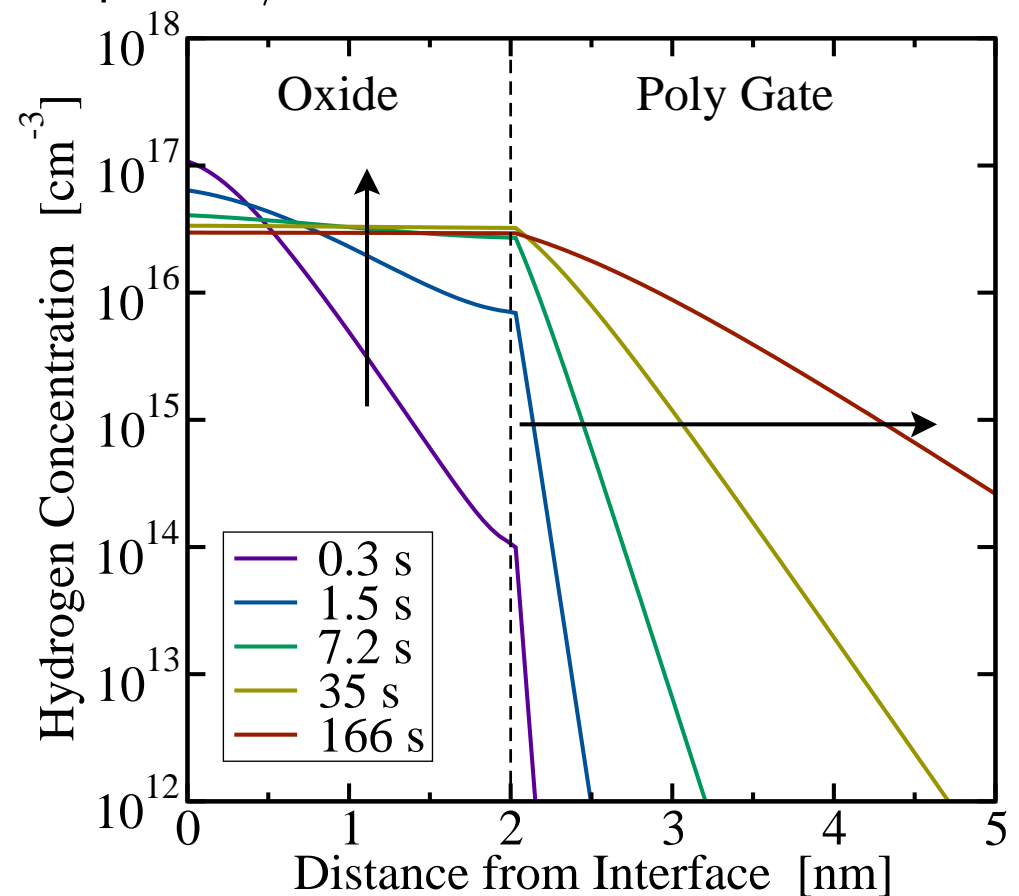For lower diffusivity in the poly compared to the oxide $\Rightarrow$ quasi-saturation

For higher diffusivity in the poly compared to the oxide $\Rightarrow$ increase in slope

However, the same H_2-RD long-term slope of 1/6 is restored [1]

Minor influence on the recovery behavior

Different recovery behavior at different stress times ('moving bumps', not universal!)

Does not speed up/slow down recovery as seen in measurement

Variant B: two interfaces with trap generation on both sides[1]

Motivation: increase in SILC has been reported after NBT stress

First interface releases H^0 which then creates H_2 and $Si^\bullet$ at the second interface

[1] Krishnan *et al.*, IEDM '05

51
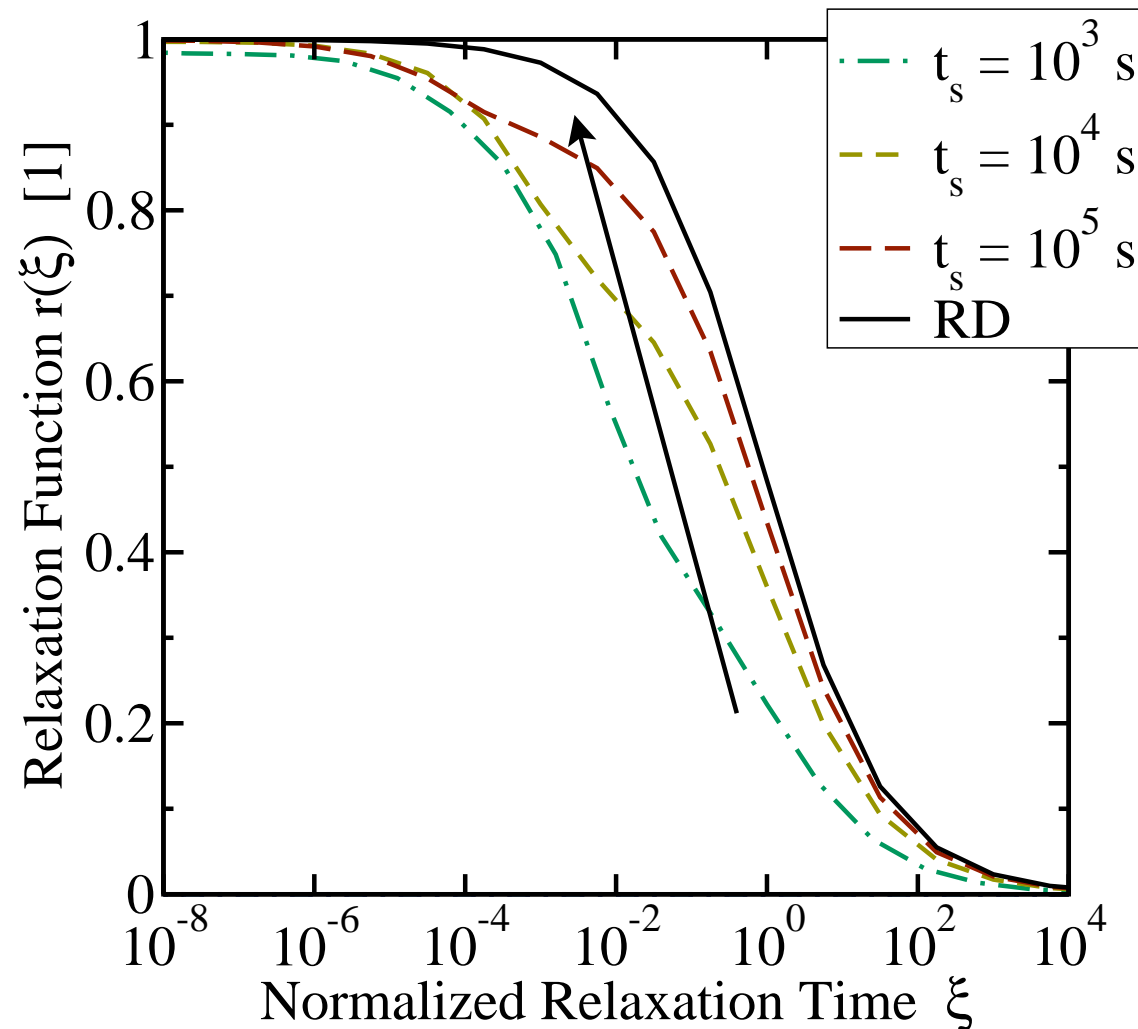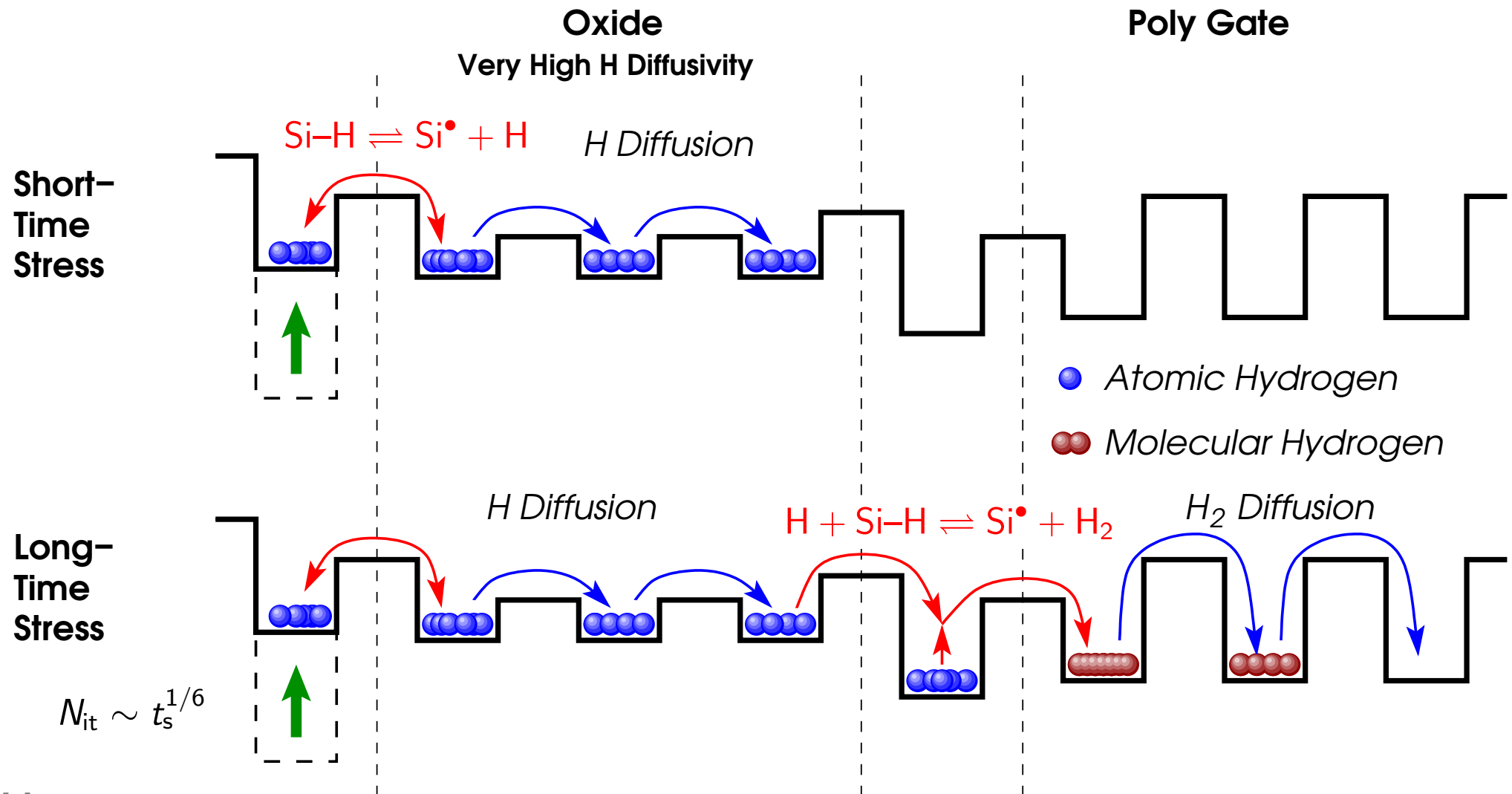
Two-Interface Reaction-Diffusion Model: Stress

Stress also results in conventional H_2-RD long-term slope $(1/6)$[1]

Quick initial build-up of H in the oxide
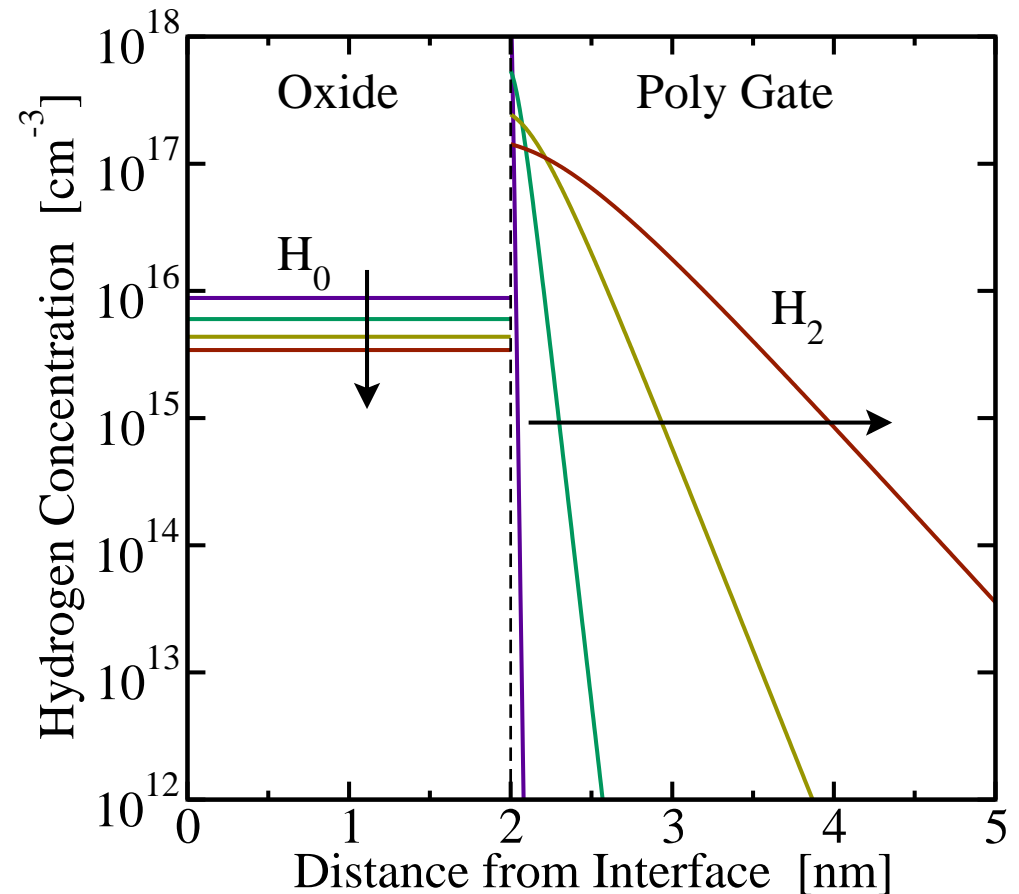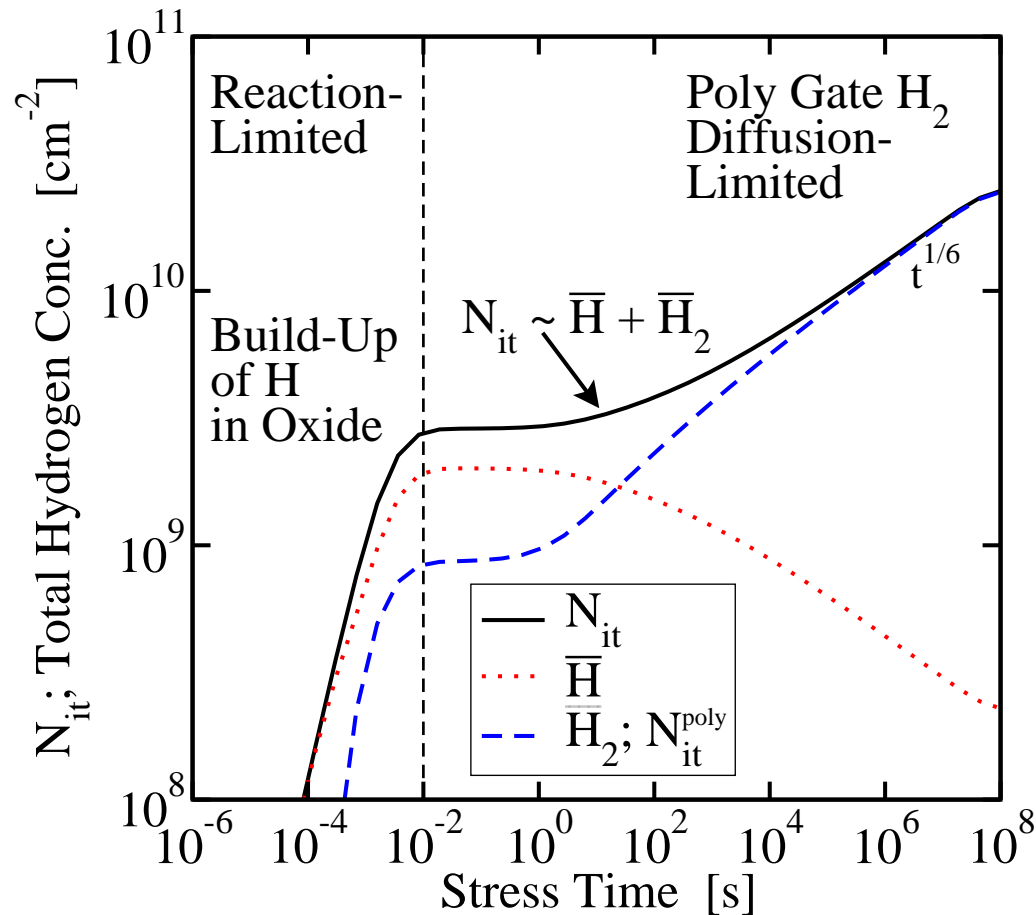
[1] Krishnan *et al.*, IEDM '05

Initial relaxation determined by (saturated) H^0 in the oxide

Initial relaxation does thus not depend on the stress time
$\Rightarrow$ *Non-universal humps, not compatible with measurements*

No improvement in the slope (no 'log(t)' behavior)
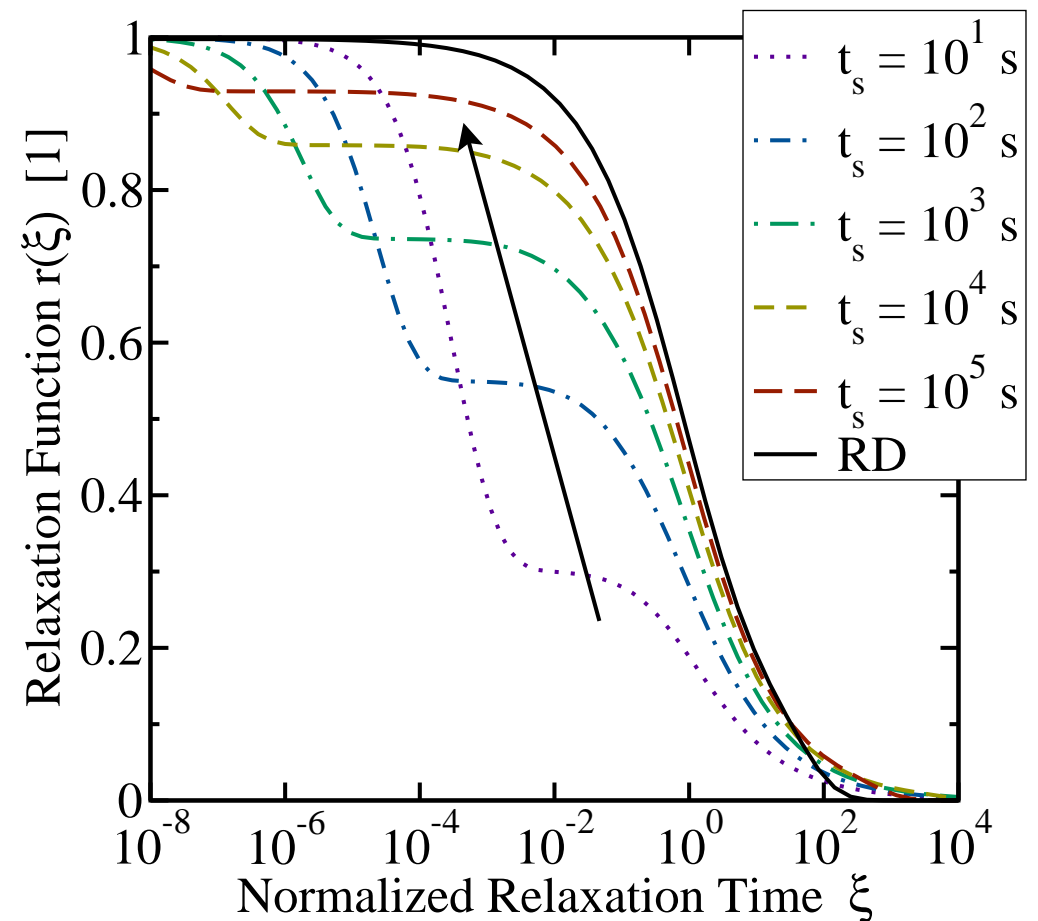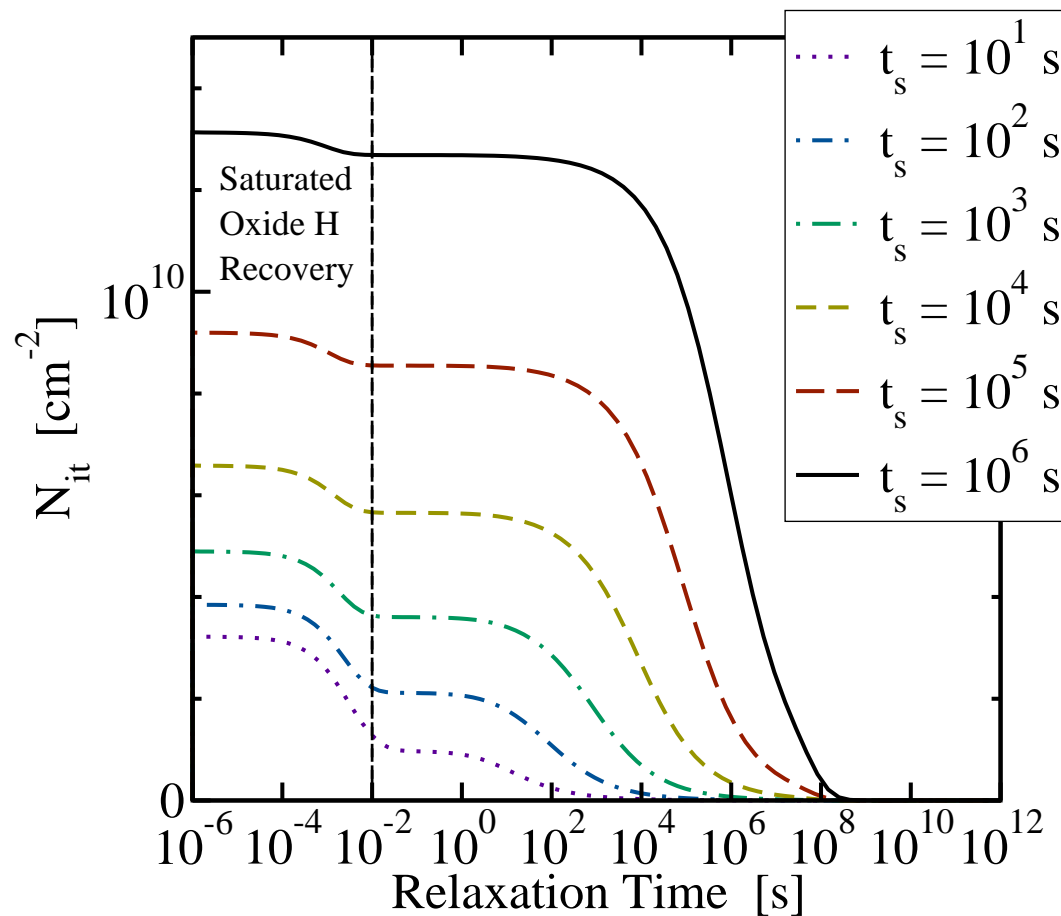
Variant C: explicit conversion between H^0 to H_2 [1]

Motivation: RD model does not fit short-time stress data

Possible explanation: no instantaneous dimerization $(H + H \rightleftharpoons H_2)$

Short-time stress data may be described by power-law with $n = 1/3$ [2]
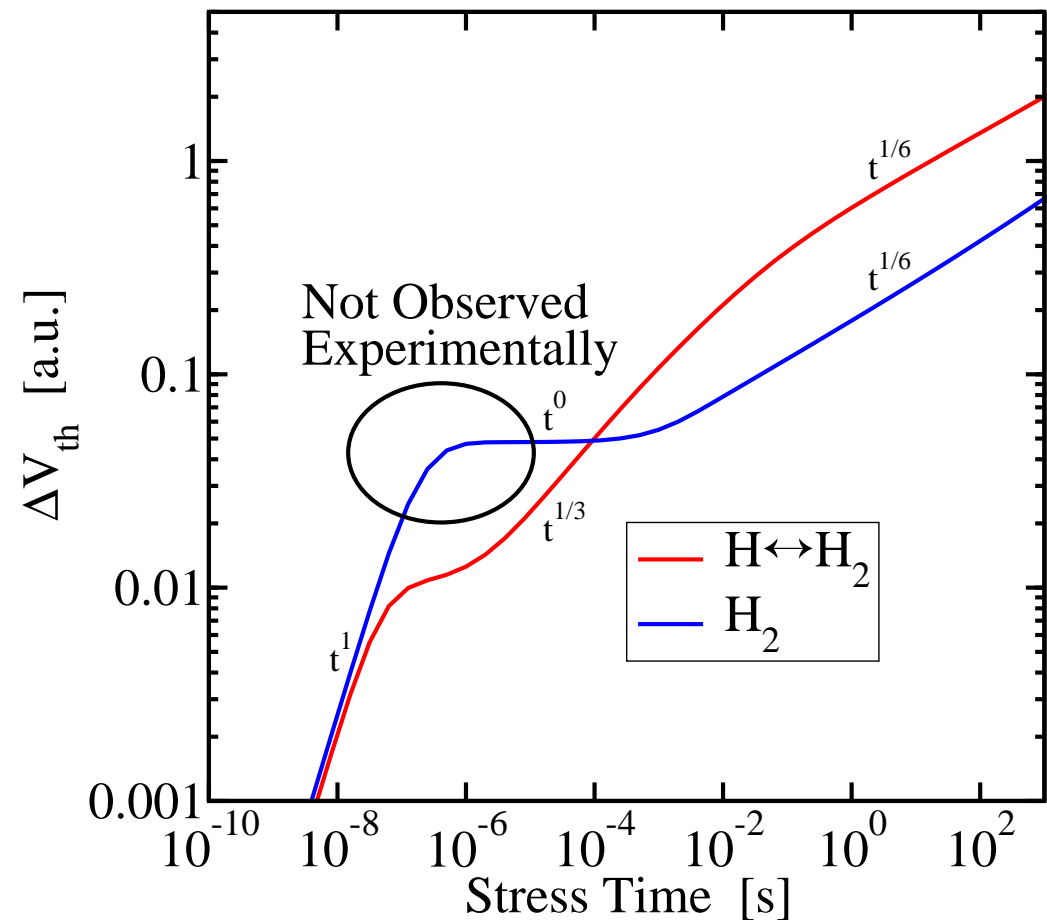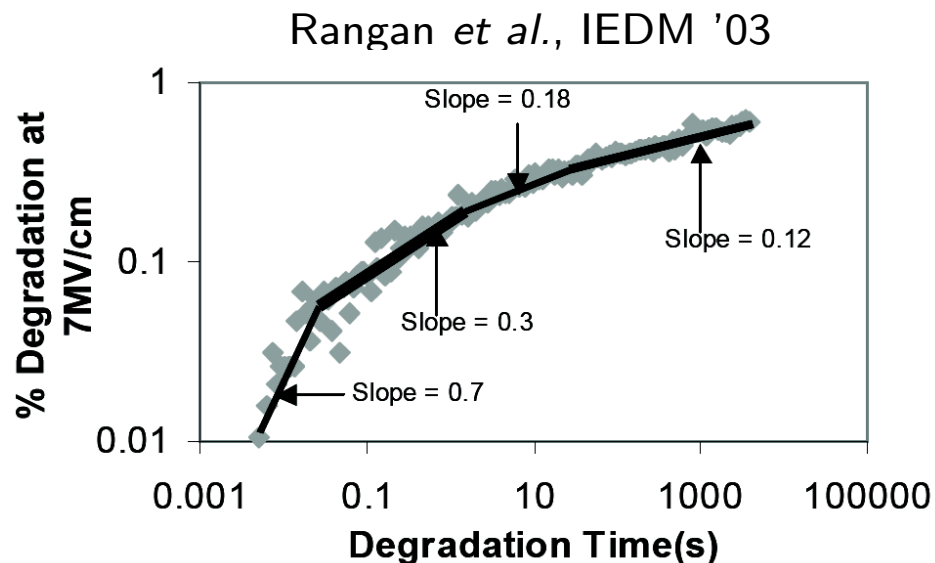
Rangan *et al.*, IEDM '03

Interface   **Oxide Bulk**

Equilibrium

$N_{it} \sim 0$

Atomic Hydrogen

Molecular Hydrogen

Reaction-Limited Stress

$N_{it} \sim t_s$

H Diffusion Barrier

H$_2$ Diffusion Barrier

H + H $\rightleftharpoons$ H$_2$
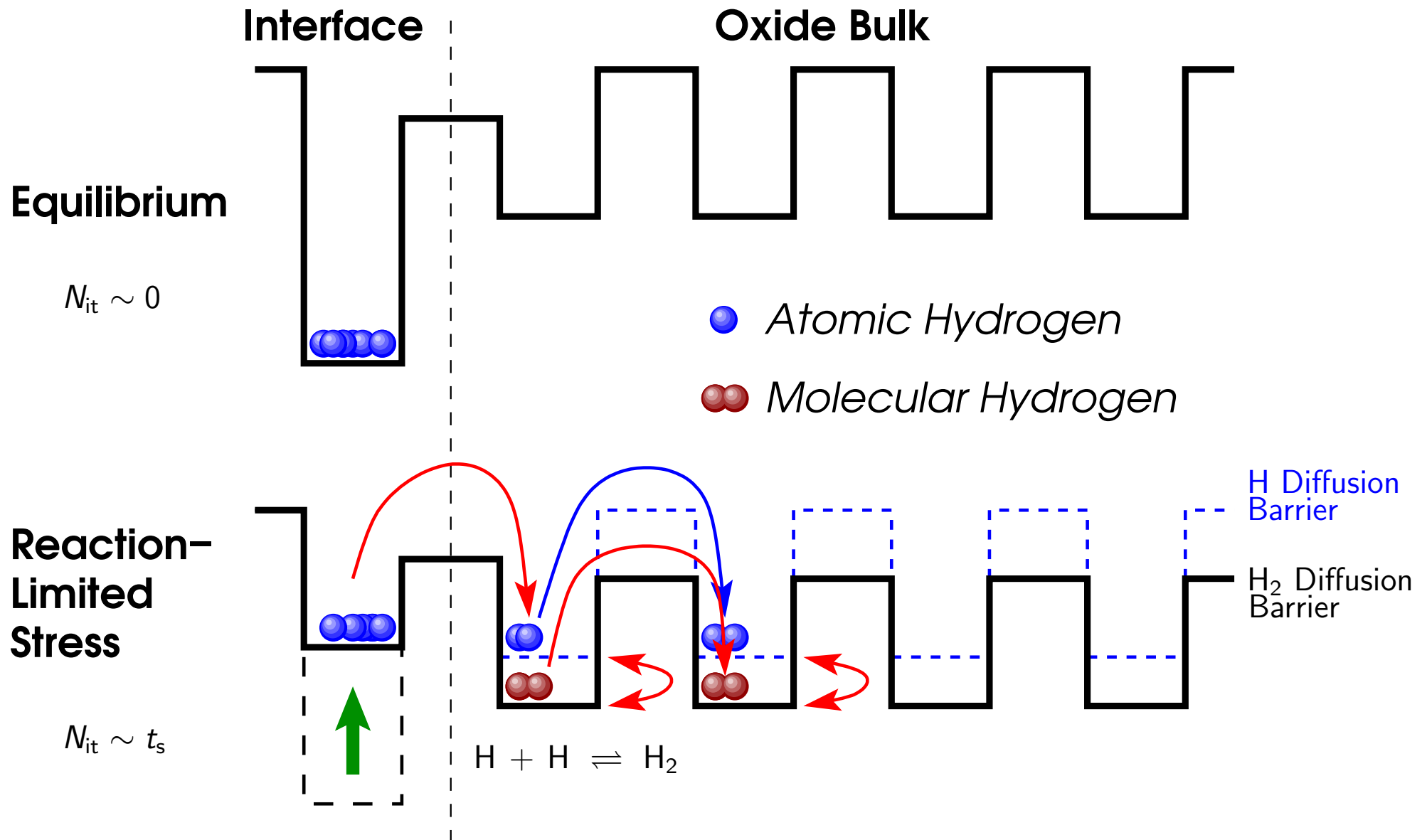
H-H$_2$ Reaction-Diffusion Model

Hydrogen diffusion profiles

Non-equilibrium at short stress times

Equilibrium at long stress times (mass action law)

$$\frac{\partial [H]}{\partial t} = k_{H_2}[H_2] - k_H[H]^2 \qquad \Rightarrow \qquad \frac{[H]^2}{[H_2]} = \text{const}$$
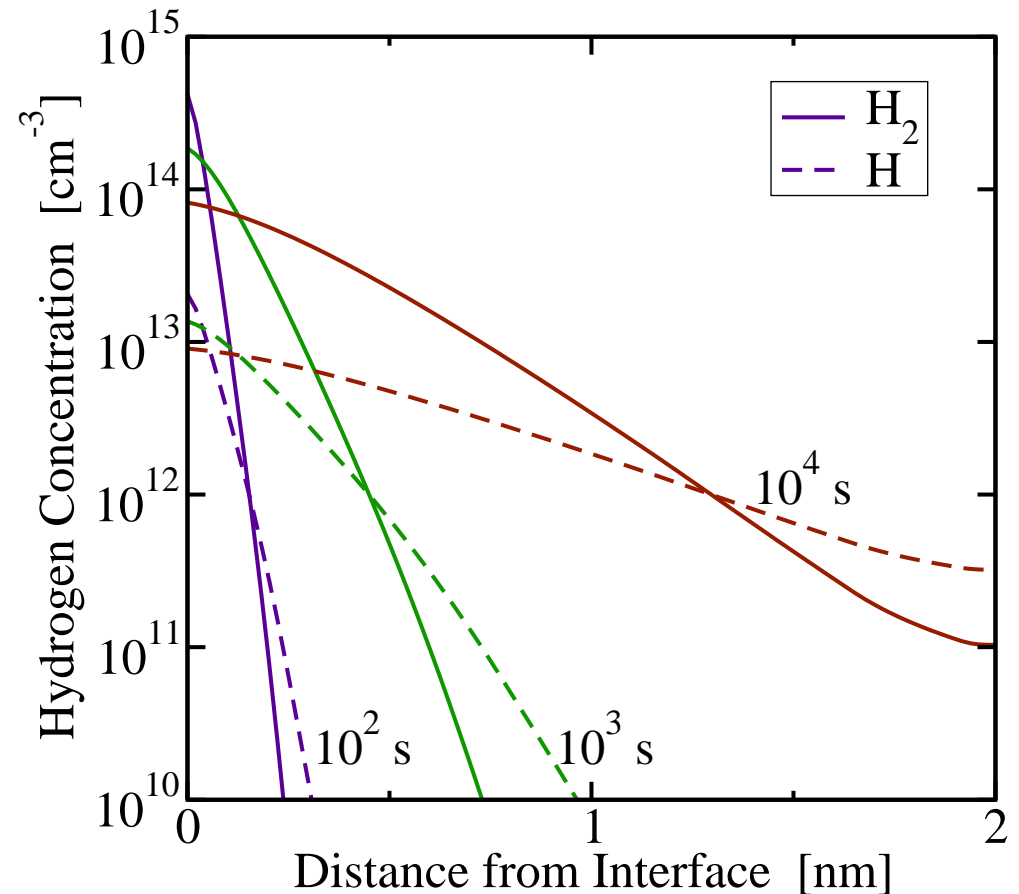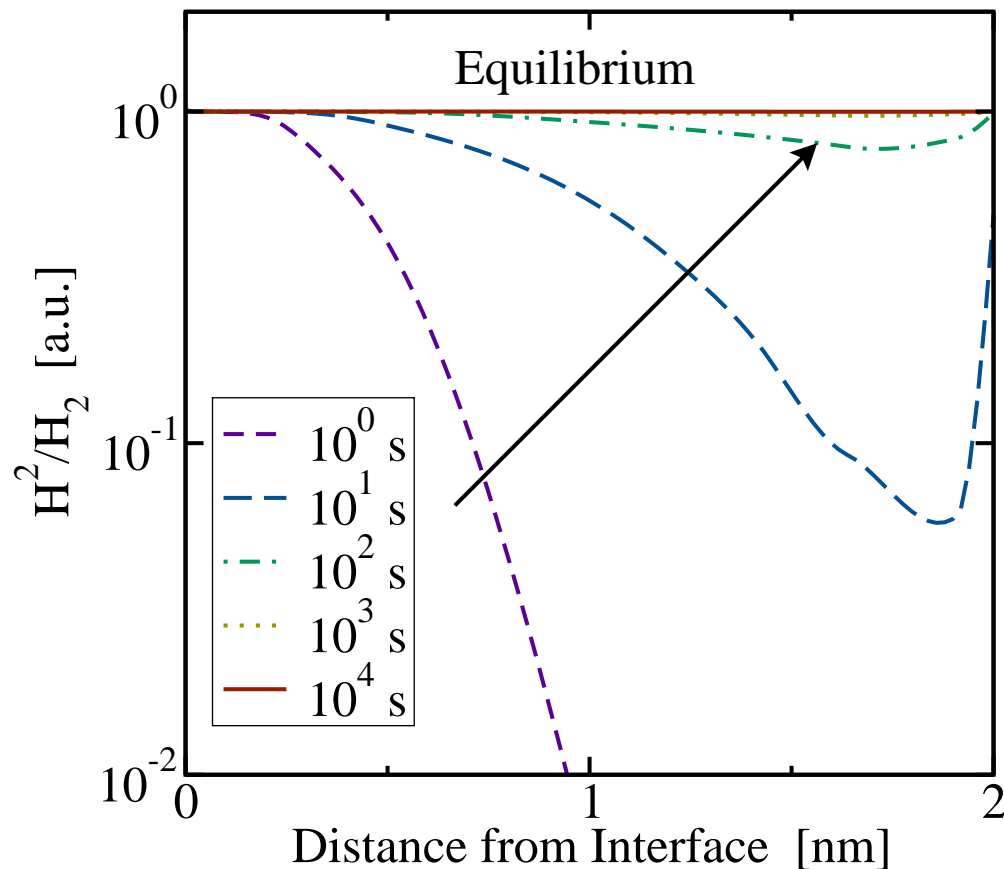
H-H$_2$ Reaction-Diffusion Model

Stress

Fit of short-term data gives slope close to 1/3

H$_2$-RD long-term slope of 1/6

Relaxation

Same as standard RD

First too slow

Then too fast

50% at $\xi = 1$ ($t_r = t_s$)
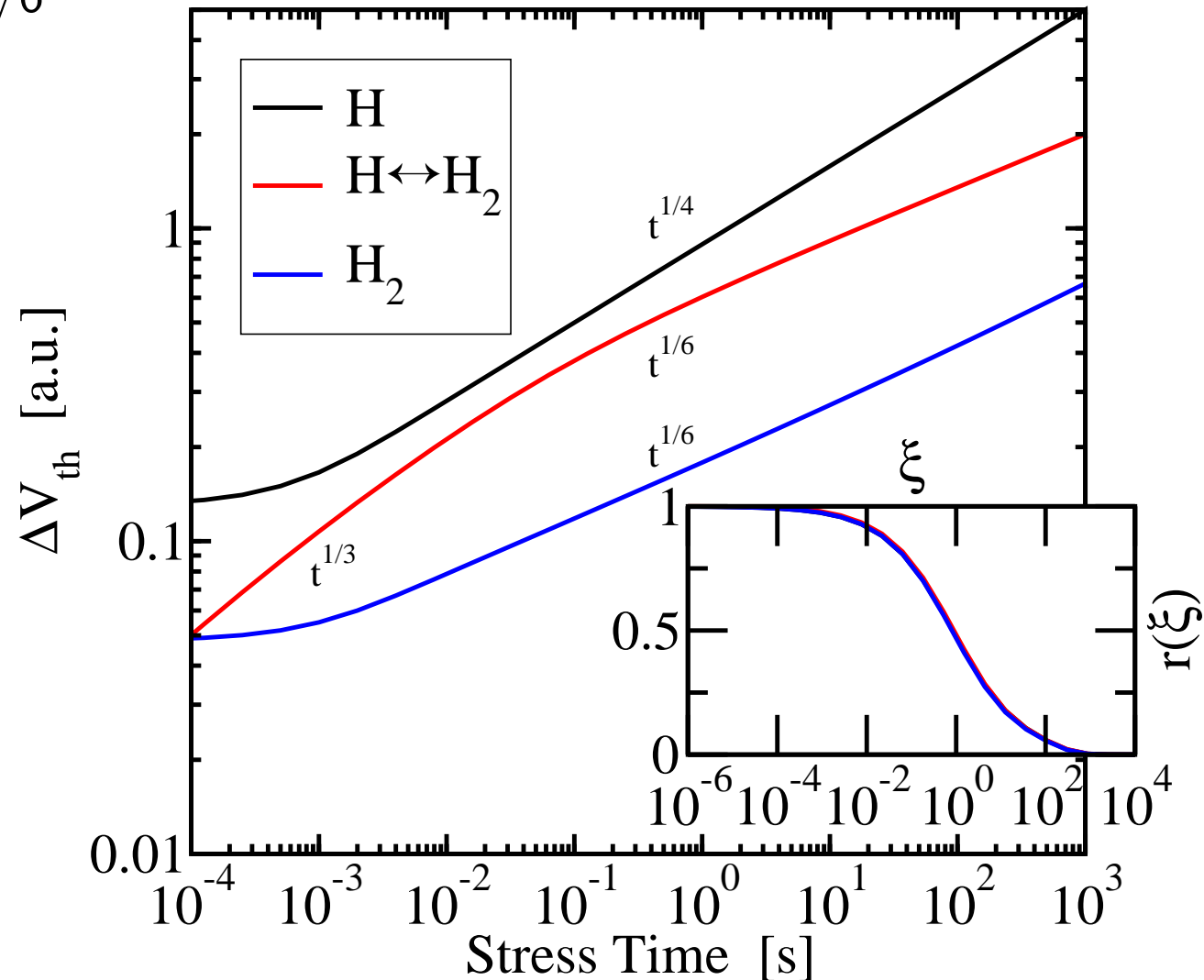
Dynamic response of the interface states

Motivation: Interface states are positively charged after stress

Stored charge may not be able to follow rapid switch from V_{stress} to V_{th}

Numerical simulation

Process described by SRH statistics

Log-like behavior

Standard values of capture cross section
Interface states are fast enough

Extremely small capture cross sections
Impact in μs regime possible

However:

Different relaxation characteristics

Not visible in measurements
Dominated by a single mechanism

Does not depend on stress time (saturated!)

How to obtain a behavior closer to measurement data?

Traps at different locations
$\Rightarrow$ *Dispersive hole-trapping model*

Figure labels: $t_s = 100$ ms, $t_s = 100$ s, $t_s = 10000$ s; $c = 10^{21}$ cm^2; $c = 2 \times 10^{16}$ cm^2; Stored Charge [eV cm^{-2}]; Relaxation Time [s]

59
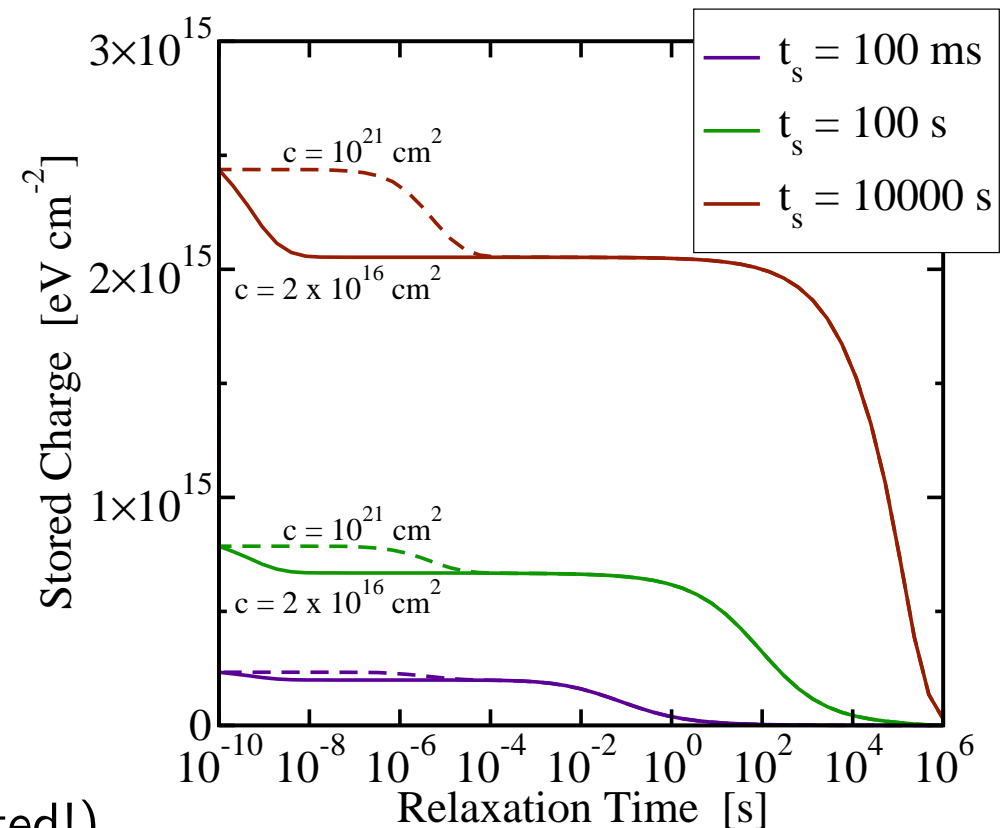
Model Checklist: Extended Reaction-Diffusion Models

Stress: undelayed measurements

   ✗ Log-like initial behavior

   ✓ Temperature and voltage independent slope, scaling property

Stress: delayed measurements

   ✗ Temperature-dependent power-law exponent/slope (proportional to temperature?)

   ✗ Delay distorts data even at larger stress times

Recovery

   ✗ Long relaxation tails of log-like nature (more than 12 decades in time?)

   ✗ Strong bias sensitivity, particularly for positive bias

   ∼ Universal behavior (or at least in good approximation)

   ✗ Fast recoverable and slowly recovering/permanent component

Duty factor dependence

   ✗ Plateau around DF = 50%, strong sensitivity for very large and small DF

Reaction-diffusion model and variants

Cannot capture long relaxation tails

Predict relaxation too late and then too fast
As a consequence, DF dependence is wrong

No bias dependence in recovery

Is the failure to capture the relaxation important?

Yes, it is fundamental!

RD models are based on a balance between forward and backward rates

$$\frac{\partial V_{th}(t_s)}{\partial t} = \text{ForwardRate} - \text{BackwardRate}$$

Forward rate alone gives power-law exponent 1 (irrespective of species)

Forward rate and backward rate give power-law exponent $1/6$ (H_2 diffusion)
Species dependence and the 'diffusion-limited regime' enter through backward rate
Backward-rate is the essence of the RD model

During stress only combined effect of backward and forward observable

Backward rate observable during relaxation
If backward rate is wrong, stress behavior questionable/coincidence

Consequences

Long relaxation tails indicate some form of dispersive mechanism

    Dispersion in that context: huge spread of time-constants (many decades)

Various dispersive mechanism have been suggested for NBTI

    Dispersive transport models [1] [2] [3]

    Dispersive rate models [4]

    Dispersive hole-trapping models [5] [6]

Can these models capture the stress phase?

Can these models capture the relaxation phase?

[1] Alam and Mahapatra, MR '05 [2] Zafar, JAP '05 [3] Kaczer, IRPS '05 [4] Huard *et al.*, MR '05

[5] Yang *et al.*, EDL '05 [6] Huard *et al.*, MR '06

Dispersive Transport and NBTI: Influence on Slope

How do traps influence hydrogen motion?

Transport becomes dispersive (not like conventional drift-diffusion)

Traps slow down transport, particles remain in traps for a 'random' amount of time
Compare the uniform 'marching of soldiers' with 'preschool kids at the zoo'

How do traps influence the NBTI slope?

Contradictory results published

Simplified dispersive transport models[1] [2] [3] [4]

Various attempts, agreement that dispersive transport decreases slope

Reaction-diffusion plus a single trap level[5]

Inclusion of traps increases slope

Dispersive multiple-trapping transport model[6]

Straight-forward application results in increased slope

To understand this discrepancy we have to look at some basics[7] ...

[1] Houssa *et al.*, JAP '04   [2] Alam and Mahapatra, MR '05   [3] Kaczer *et al.*, IRPS '05   [4] Zafar, JAP '05

[5] Chakravarthi *et al.*, IRPS '04   [6] Grasser *et al.*, SISPAD '06   [7] Grasser *et al.*, TDMR '08

Reaction-Dispersive-Diffusion Model: Schematics – I

Interface

Oxide Bulk

Dispersive Hydrogen Diffusion

Si–H $\rightleftharpoons$ Si$^\bullet$ + H_c

Initial Stress

Hydrogen Conduction Level

Hydrogen Trap Levels

'Occasional' Diffusion Events

Equilibrium

Long–Time Stress

Low Hydrogen Conc. in Conduction Level

Most Hydrogen is Trapped

64
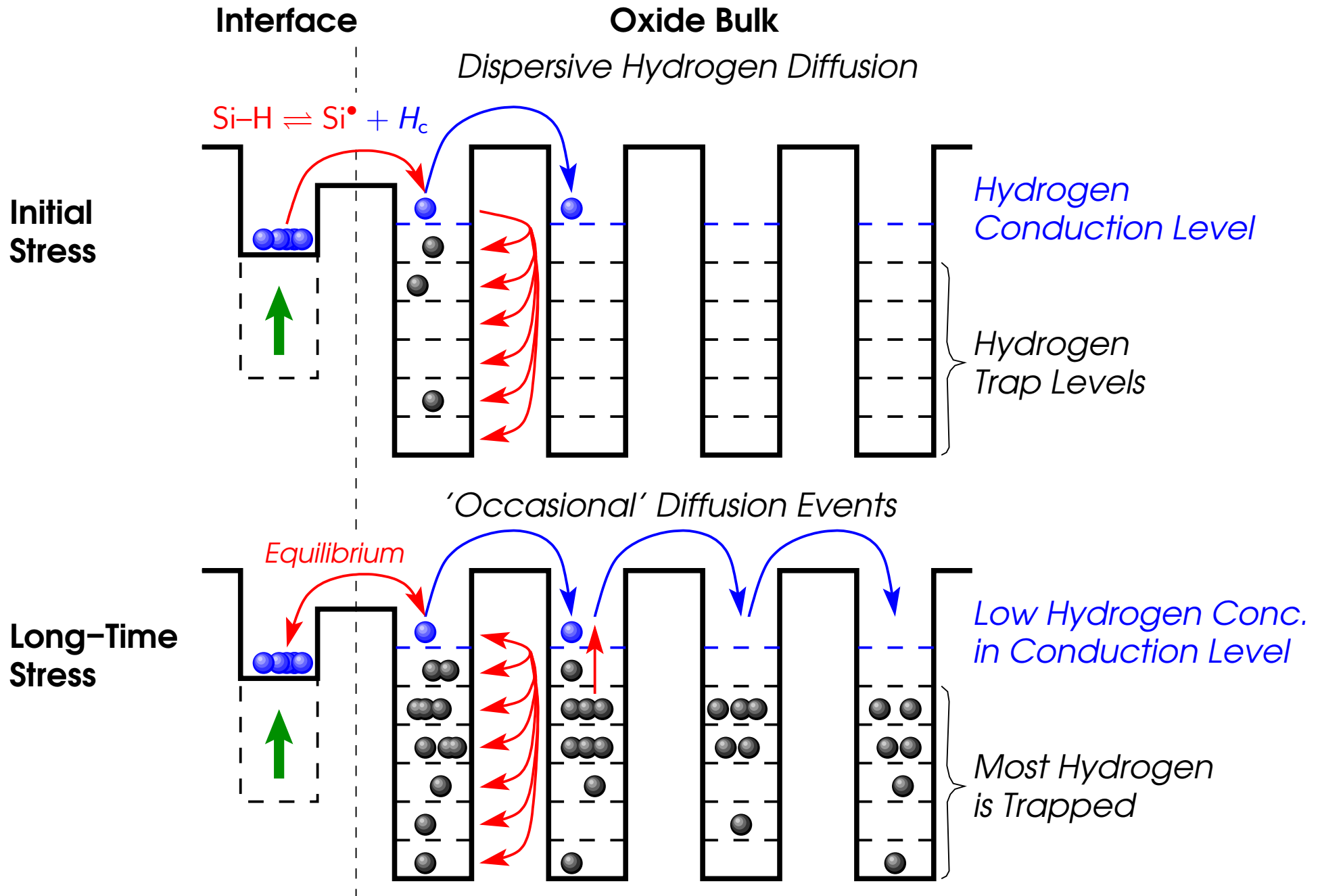
Dispersive Transport and NBTI

Most crucial aspect: coupling of transport model to interface reaction[1] [2]

Requirement: interfacial hydrogen concentration

$$\frac{\partial N_{\text{it}}}{\partial t} = k_{\text{f}}(N_0 - N_{\text{it}}) - k_{\text{r}} N_{\text{it}} H_{\text{it}}^{1/a}$$

Remember: H exists in 'free' (shallow) and deep (trapped) states!

What is H_{it}?

The information on H_{it} is given by the microscopic physics at the interface

Should/must be consistent with the dispersive transport model

[1] Grasser *et al.*, IIRW '06   [2] Grasser *et al.*, TDMR '08

Dispersive Transport

Detailed model: multiple trapping (MT) model[1]

Alternative A: extremely-nonequilibrium approximation of MT[2]

Alternative B: time-dependent diffusivity[3]

> Obtained for instance from hydrogen/deuterium SIMS profiles
>
> Simple empirical model, plugged into standard drift-diffusion equations
>
> $$D(t) = D_0(\nu_0 t)^{-(1-\alpha)}$$
>
> $\alpha = k_B T_L/E_0$ is dispersion coefficient, E_0 characteristic energy of trap DOS
>
> D_0 is the microscopic diffusivity, ν_0 the attempt frequency

Alternative C: continuous-time random walk (CTRW)[4]

All three alternatives have been used for NBTI modeling

> EN-MT[5], empirical diffusivity[6] [7], CTRW[8]
>
> The three alternatives give very similar results, but contradict MT[9] [10]

[1] E.g., Noolandi, PRB '77   [2] Arkhipov and Rudenko, Phil.Mag.B '82   [3] E.g., Nickel *et al.*, PRB '96

[4] E.g., Scher and Montroll, PRB '75   [5] Kaczer *et al.*, IRPS '05   [6] Alam and Mahapatra, MR '05

[7] Zafar, JAP '05   [8] Houssa, JAP '04   [9] Grasser *et al.*, IIRW '06   [10] Grasser *et al.*, TDMR '08

RDD Model: Boundary Conditions

The MT model differentiates between trapped and free H

Only the 'free' H_c can be used for the backward rate

Reverse rate is a two-step process: release of H_t to H_c followed by passivation

The simplified models only consider the total hydrogen concentration
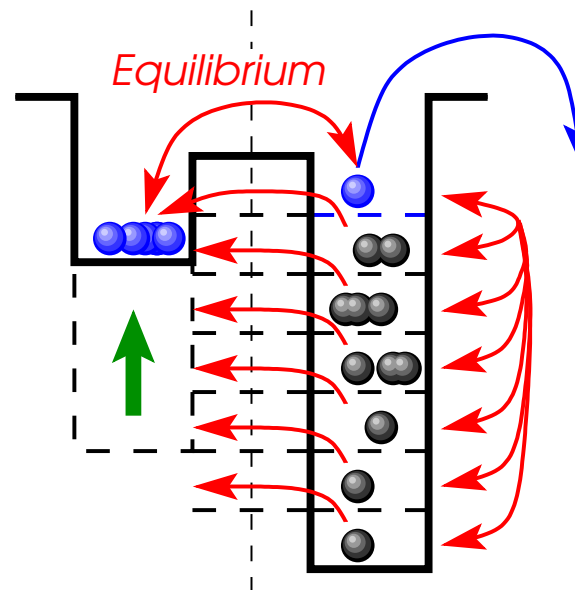
The total hydrogen concentration $H_c + H_t$ is used for the backward rate[1] [2]

Barrier for the reverse rate becomes independent of trapping level, one-step process

Only Conduction Hydrogen
$Si–H \rightleftharpoons Si^{\bullet} + H_c$
Equilibrium

Total Hydrogen
$Si–H \rightleftharpoons Si^{\bullet} + H_c + H_t$
Equilibrium
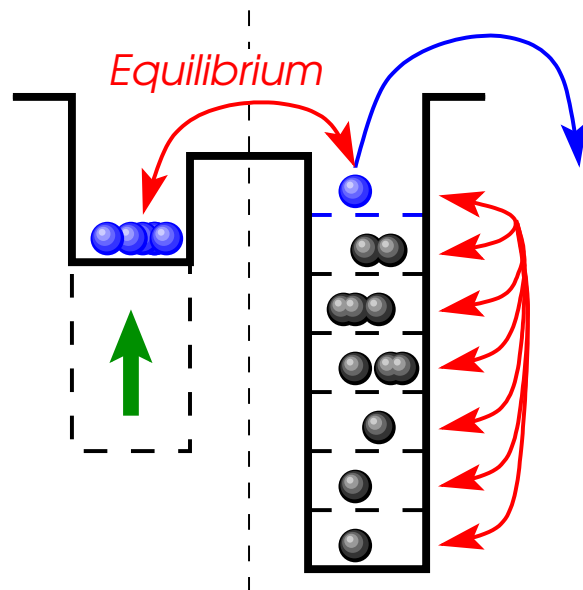
[1] Kaczer *et al.*, APL '05   [2] Grasser *et al.*, TDMR '08

How do the two boundary conditions influence the slope?

Comparison: numerical solution[1]

- Complete multiple-trapping model

Boundary condition: $H_{it} = H_{tot}$

- Hydrogen closer to interface
- Reverse reaction large
- $\Rightarrow$ slope decreases
  $$N_{it}(t) \sim t^{\alpha/4}, \quad \alpha = \frac{k_B T}{E_0}$$

Boundary condition: $H_{it} = H_c$
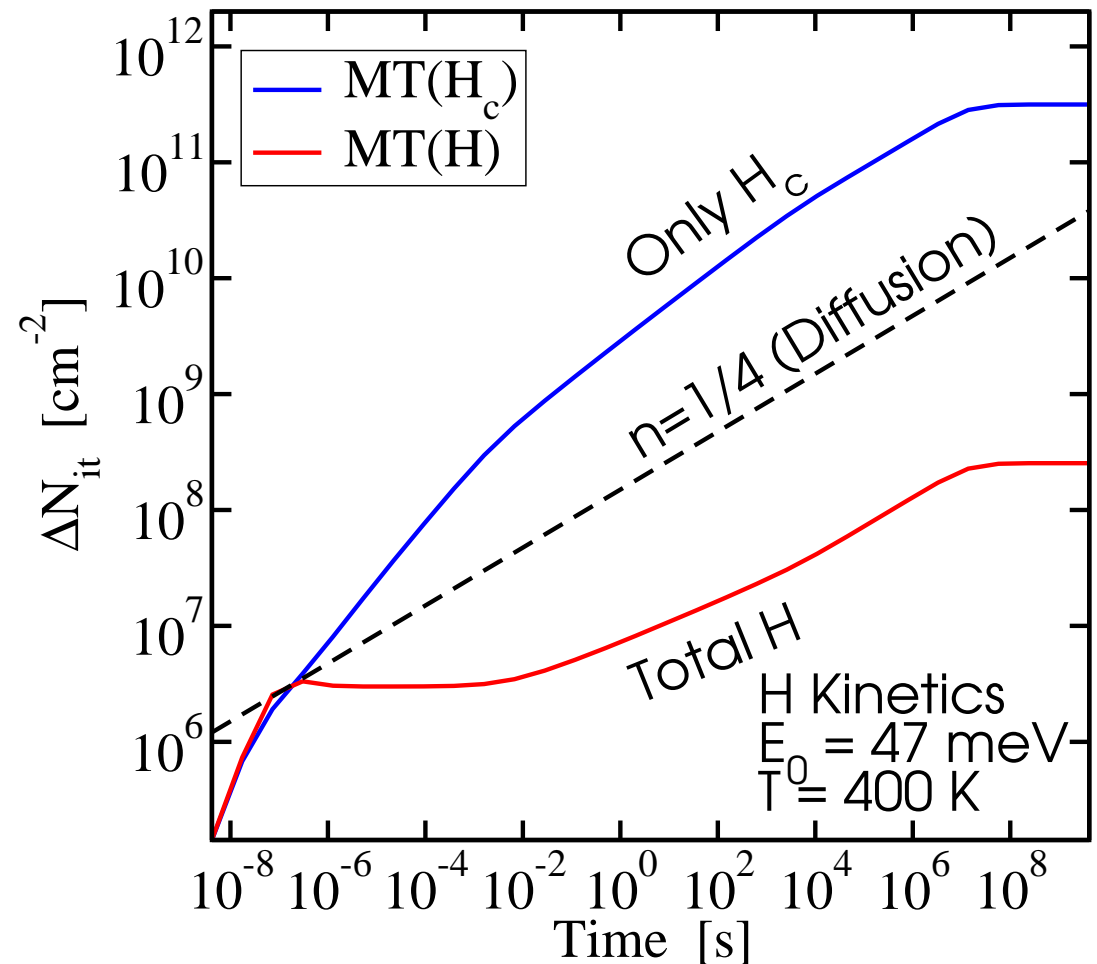
- Hydrogen quickly trapped
- Reverse reaction small
- $\Rightarrow$ slope increases
  $$N_{it}(t) \sim t^{1/2 - \alpha/4}$$

Temperature-dependent slopes!

- Contradiction to undelayed (OTF) measurements

[1] Grasser et al., TDMR '08

Relaxation predicted by reaction-dispersive-diffusion model

Boundary condition: $H_{it} = H_{tot}$

Ultra-fast initial recovery of interfacial hydrogen
Not universal

Long-time recovery dispersive-diffusion-limited

Boundary condition: $H_{it} = H_c$

Slower recovery compared to RD
Hydrogen has to be detrapped first
Long-time component of $H_{it} = H_{tot}$
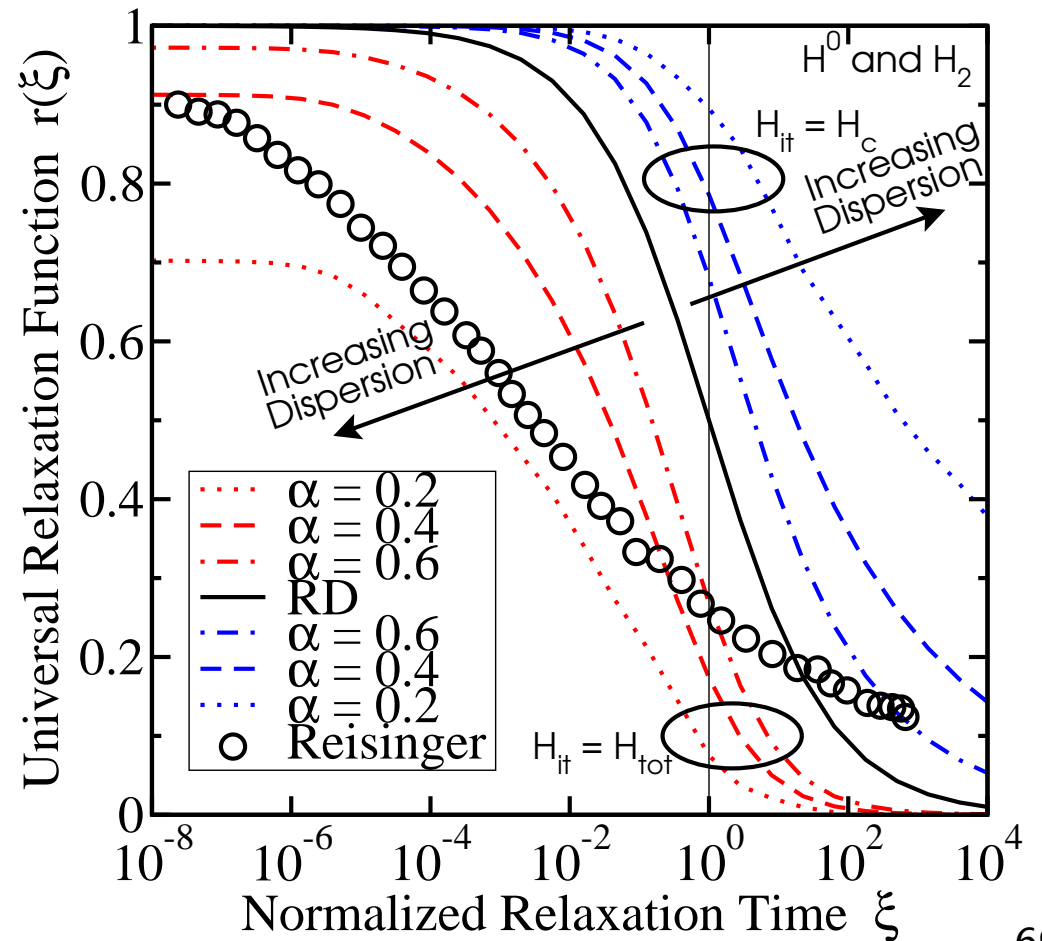*Recovery too slow*

Reaction-diffusion limit

Watershed which cannot be crossed

Essential question

Which BC is correct?
Probably neither

Model Checklist: Reaction-Dispersive-Diffusion Model

Stress: undelayed measurements

  ✗ Log-like initial behavior

  ✗ Temperature and voltage independent slope, scaling property

Stress: delayed measurements

  ✓ Temperature-dependent power-law exponent/slope (proportional to temperature?)

  ✓ Delay distorts data even at larger stress times

Recovery

  ~ Long relaxation tails of log-like nature (more than 12 decades in time?)

  ~ Strong bias sensitivity, particularly for positive bias

  ~ Universal behavior (or at least in good approximation)

  ~ Fast recoverable and slowly recovering/permanent component

Duty factor dependence

  ✗ Plateau around DF $= 50\%$, strong sensitivity for very large and small DF

$\Rightarrow$ In a nutshell: NBTI does not seem to be diffusion-limited

    At least not in the way it has been modeled so far

Reaction-Limited Model

Interface passivation/depassivation experiments indicate dispersive rates[1] [2]

Kinetics can be described using Gaussian distributions of binding energies/barriers

Creation of interface states visible in CP experiments
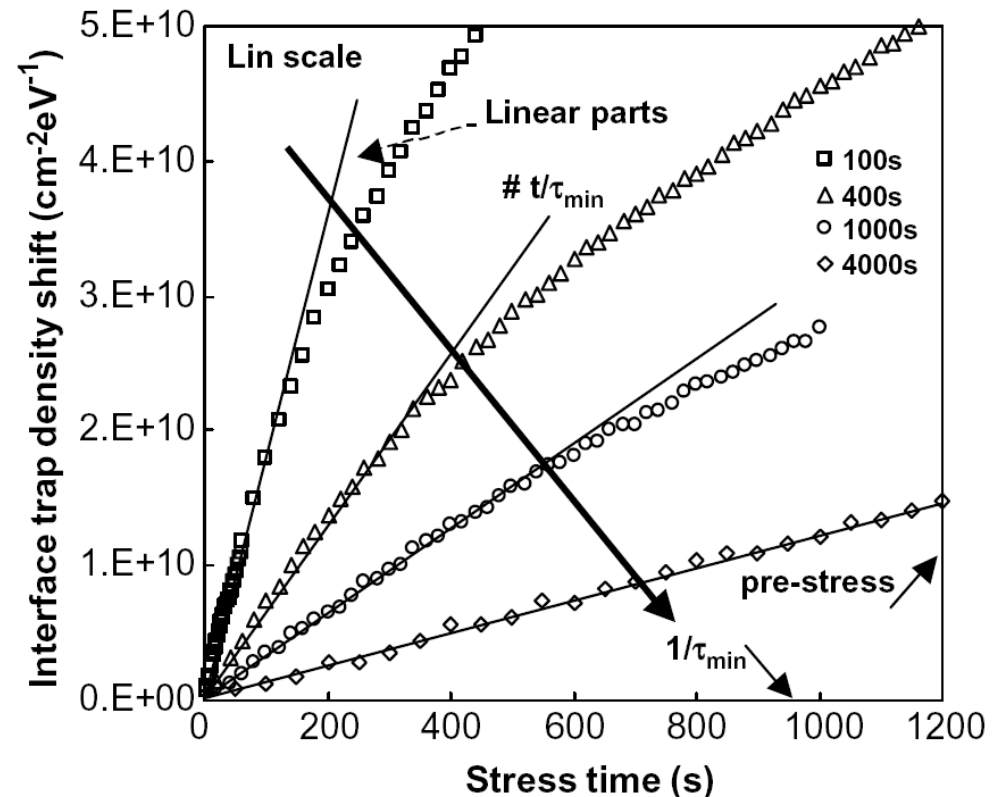
RD reaction-limited regime

$$\Delta N_{it}(t_s) = k_f \, t_s$$

According to RD theory:

ΔN_{it} *recovers after stress*

ΔN_{it} *in restress should be identical*
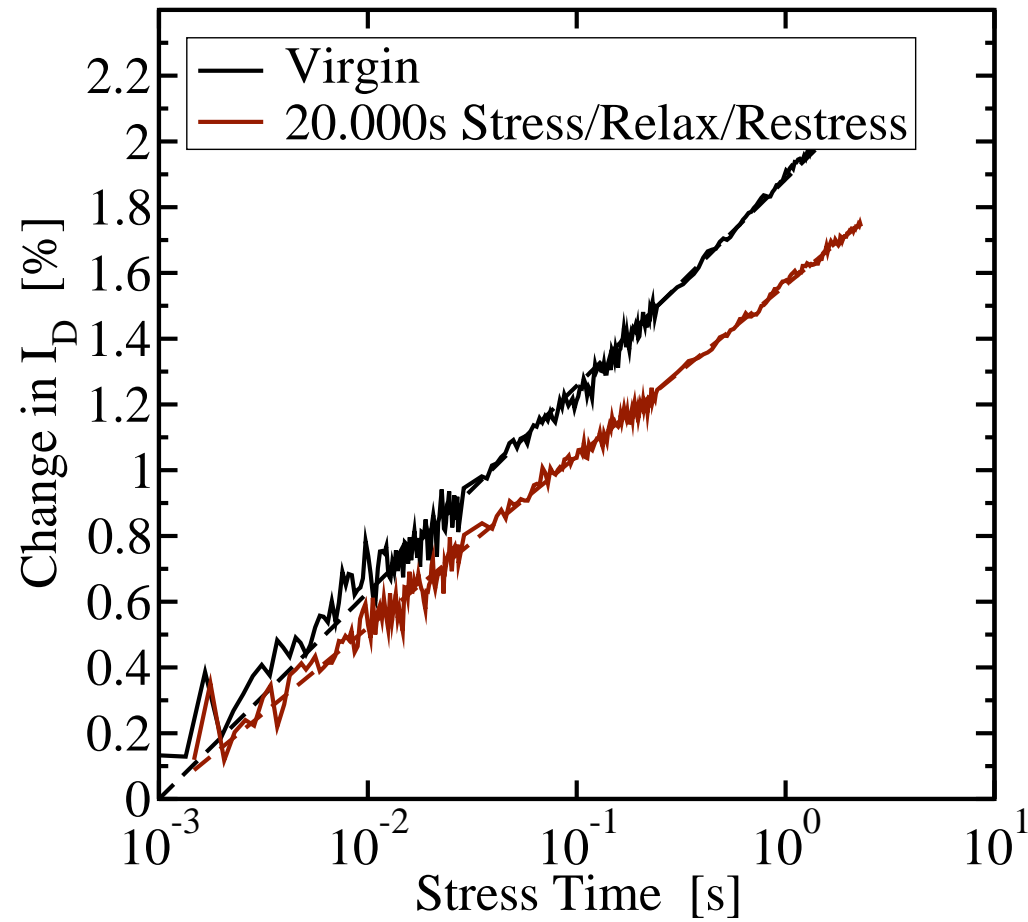
This is not the case! [3]

Huard *et al.*, MR '06

[1] Stesmans, APL '96   [2] Stesmans *et al.*, PRB '00   [3] Huard *et al.*, MR '06

Reaction-Limited Model

Effect also visible in I_D change

Possible explanation[1]

Distribution of bonding energies

Weakest bonds break first

No/weak recovery of broken bonds

RL response of a single bond

$$\frac{\Delta V_{th}(t)}{\Delta V_{th,max}} = 1 - \exp(-k_f(E_d)t)$$

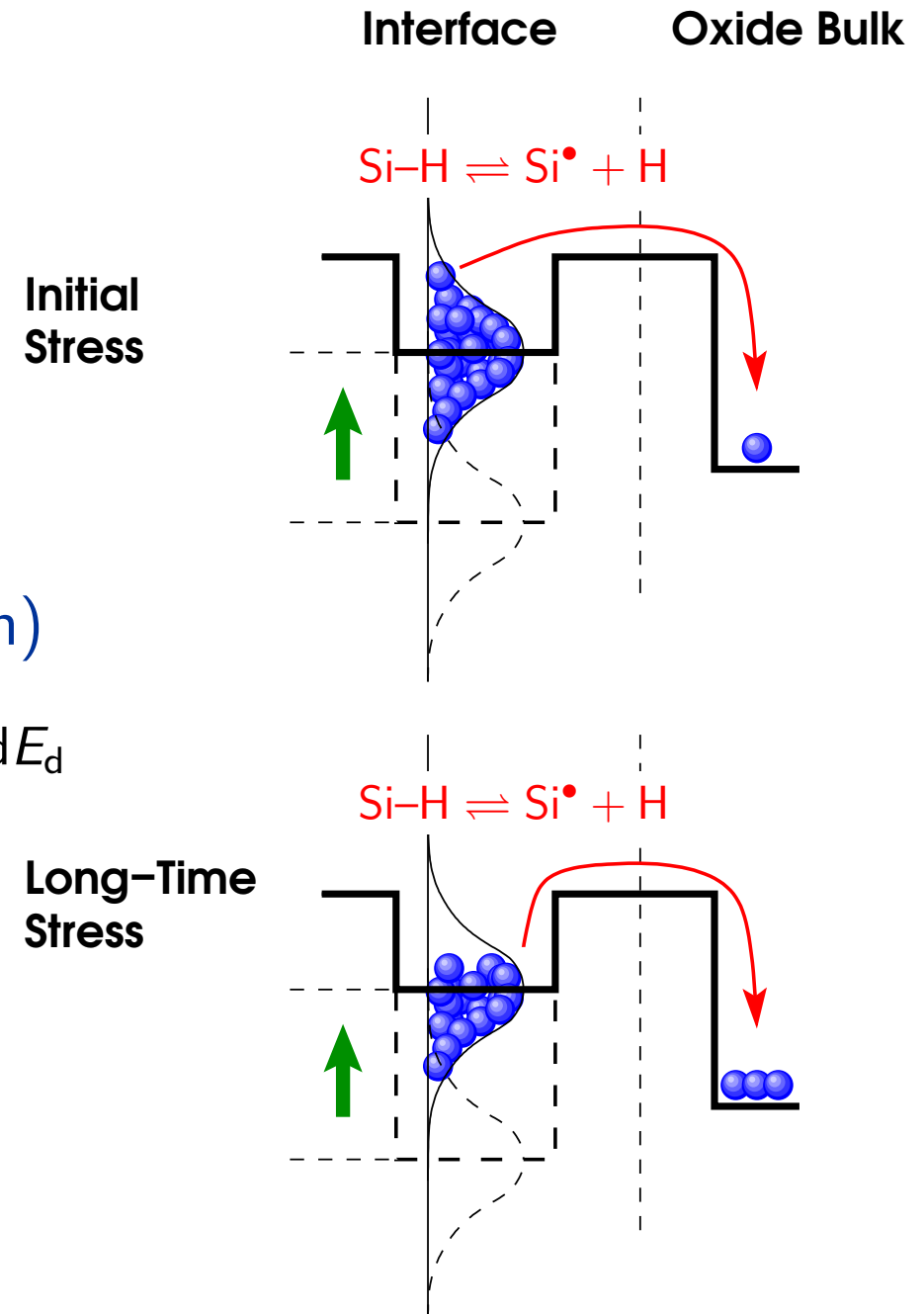**Sum over all bonds (Gaussian distribution)**

$$\frac{\Delta V_{th}(t)}{\Delta V_{th,max}} = \int g(E_d)\Big(1 - \exp(-k_f(E_d)t)\Big)dE_d$$

$$\approx \frac{1}{1 + (t/\tau)^{-\alpha}}$$

Caveat: CP highly invasive

Positive bias! Measurement delay!

Is there really no recovery in CP data? [2]
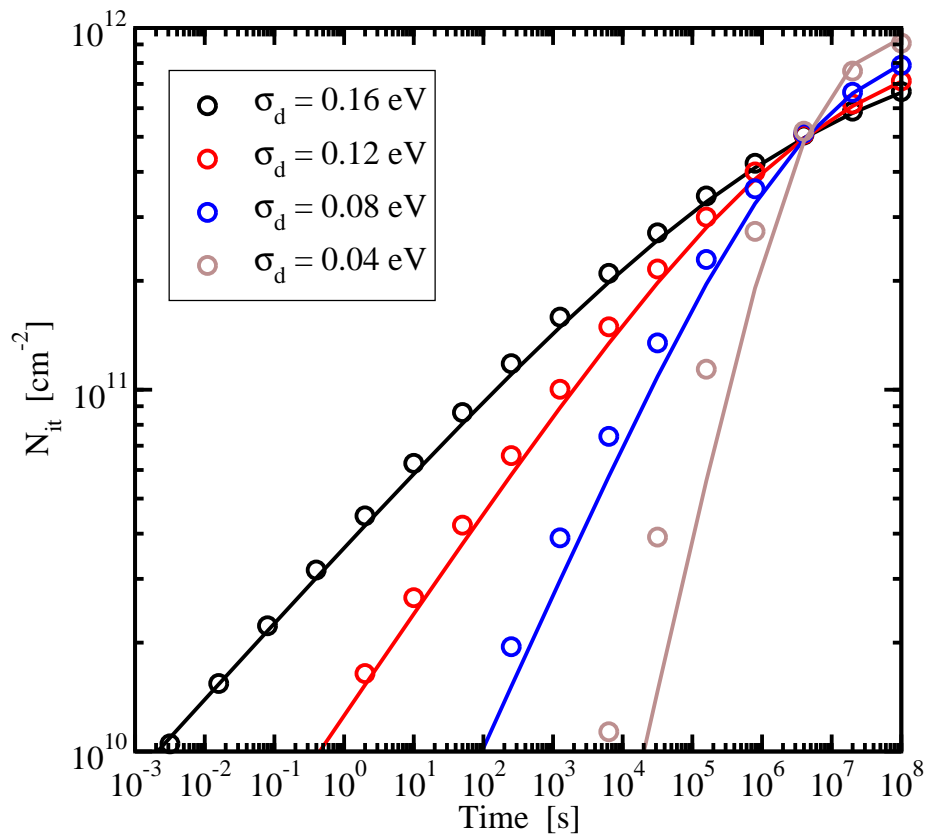
Interface   **Oxide Bulk**

$Si\text{–}H \rightleftharpoons Si^{\bullet} + H$

Initial Stress

$Si\text{–}H \rightleftharpoons Si^{\bullet} + H$

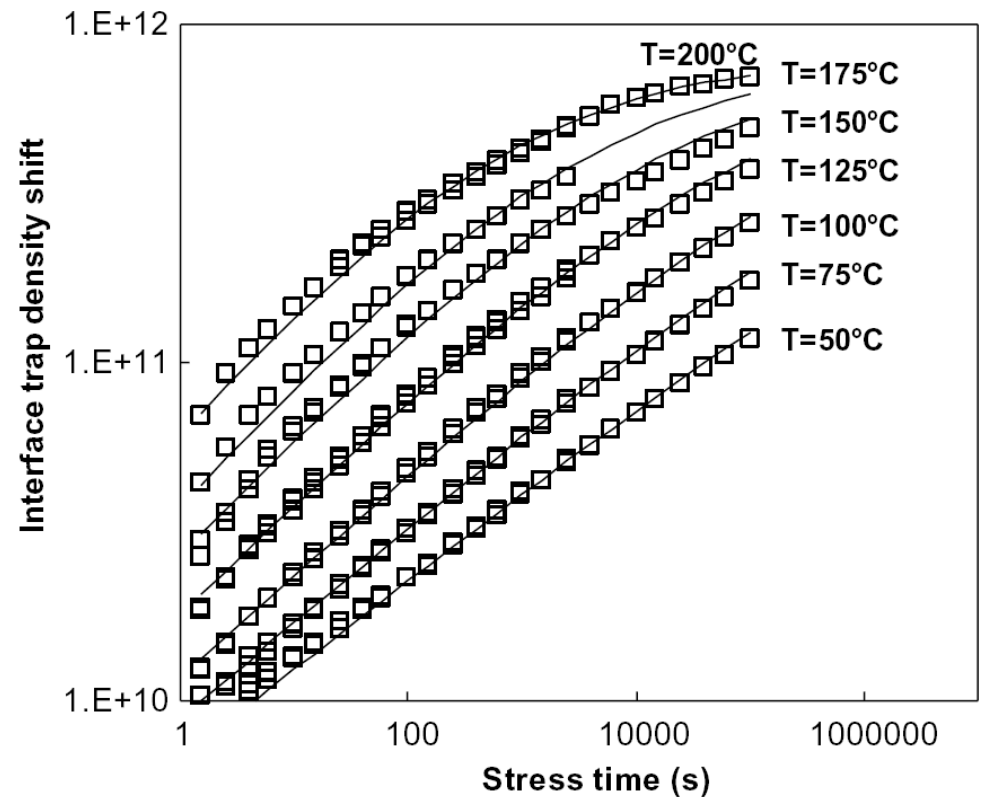**Long–Time Stress**

[1] Huard *et al.*, IEDM '07   [2] Huard *et al.*, MR '06

Excellent agreement of model to CP stress data[1]

Numerical vs. analytic solution

Analytic vs. measurements

Huard *et al.*, MR '06

[1] Huard *et al.*, MR '06

The Fast/Recoverable Component: Hole Trapping?

Detailed hole trapping model of Tewksbury[1] [2]

    Developed for 'thick' oxides of about 30 nm

    Trapping of electrons and holes in a near-interfacial layer

    Threshold-voltage shifts of $1\,\mathrm{mV}$ for NMOS, $100\,\mu\mathrm{V}$ for PMOS

    Maximum charge storage time of about $1\,\mathrm{s}$

    Much larger effect for nitrided oxides, with NMOS and PMOS being similar

Various interactions with trap levels possible via tunneling

    Conduction and valence band tunneling

    Tunneling via interface states

    Predicts log(t)-like stress and recovery behavior

Does this model explain the fast/recoverable component of NBTI?[3]

    Most controversial subject[4] [5] [6]

Alternative hole trapping models have been published, not discussed here

   Detailed physical mechanisms not given, qualitative theories[7] [8] [9] [10]

[1] Tewksbury, Ph.D. Thesis, MIT '92 [2] Tewksbury and Lee, SSC '94 [3] Huard *et al.*, MR '06

[4] Mahapatra *et al.*, IRPS '07 [5] Huard *et al.*, IEDM '07 [6] Grasser *et al.*, IRPS '08 [7] Zhang *et al.*, TED '04

[8] Ang and Wang, EDL '06 [9] Reisinger *et al.*, IRPS '06 [10] Shen *et al.*, IEDM '06

Tewksbury model[1]: charging and discharging of traps via elastic tunneling

Equilibrium   Stress   Recovery
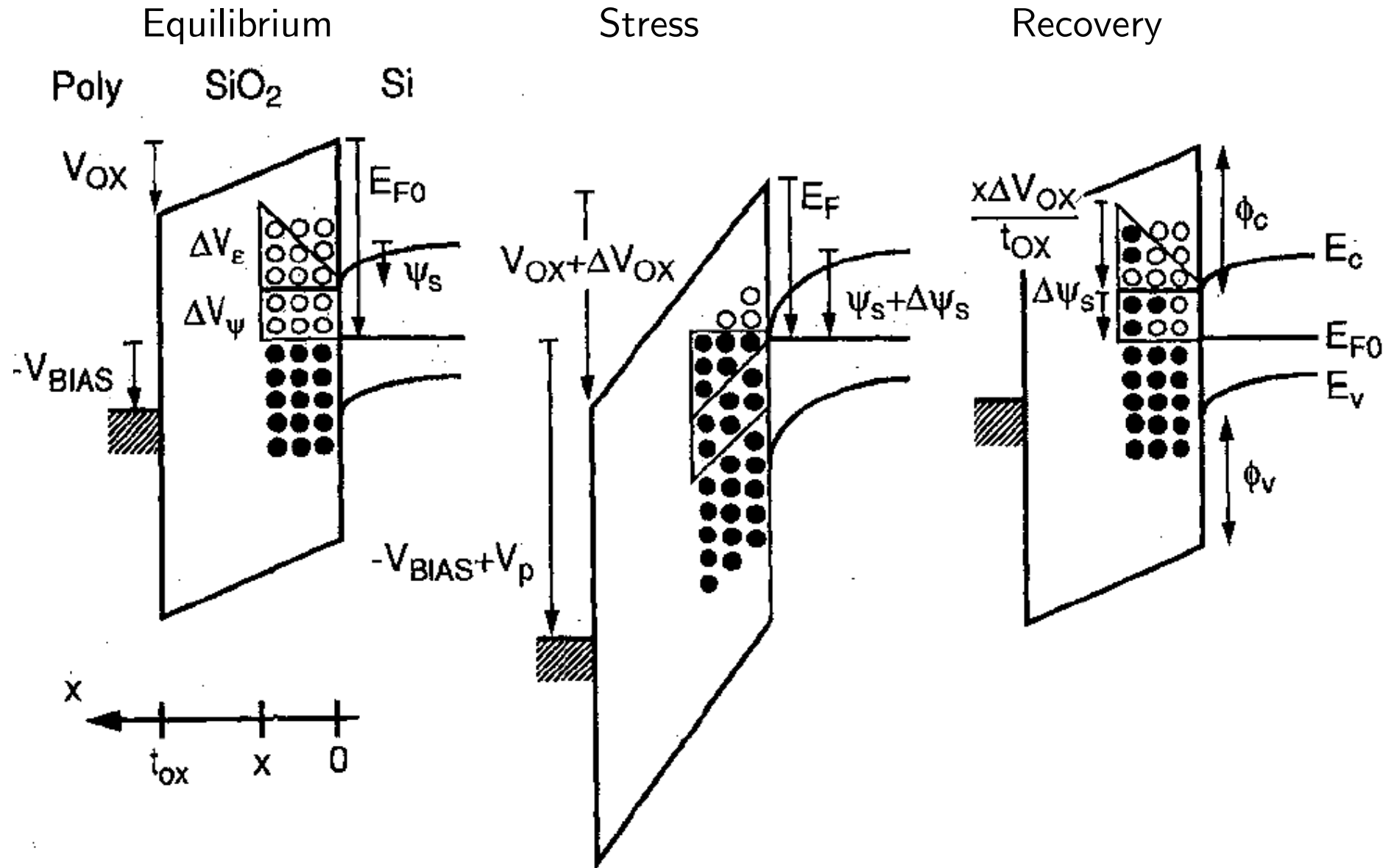
Relaxation and Duty Factor Dependence

Combination: reaction-limited creation of N_{it}, Tewksbury's hole-trapping[1]

Calibrated relaxation behavior is in excellent agreement with data[2]

Log-like behavior over many decades in time

Calibrated model gives DF dependence in excellent agreement with data

The strong impact around DF $\rightarrow$ 100% and the plateau are well reproduced
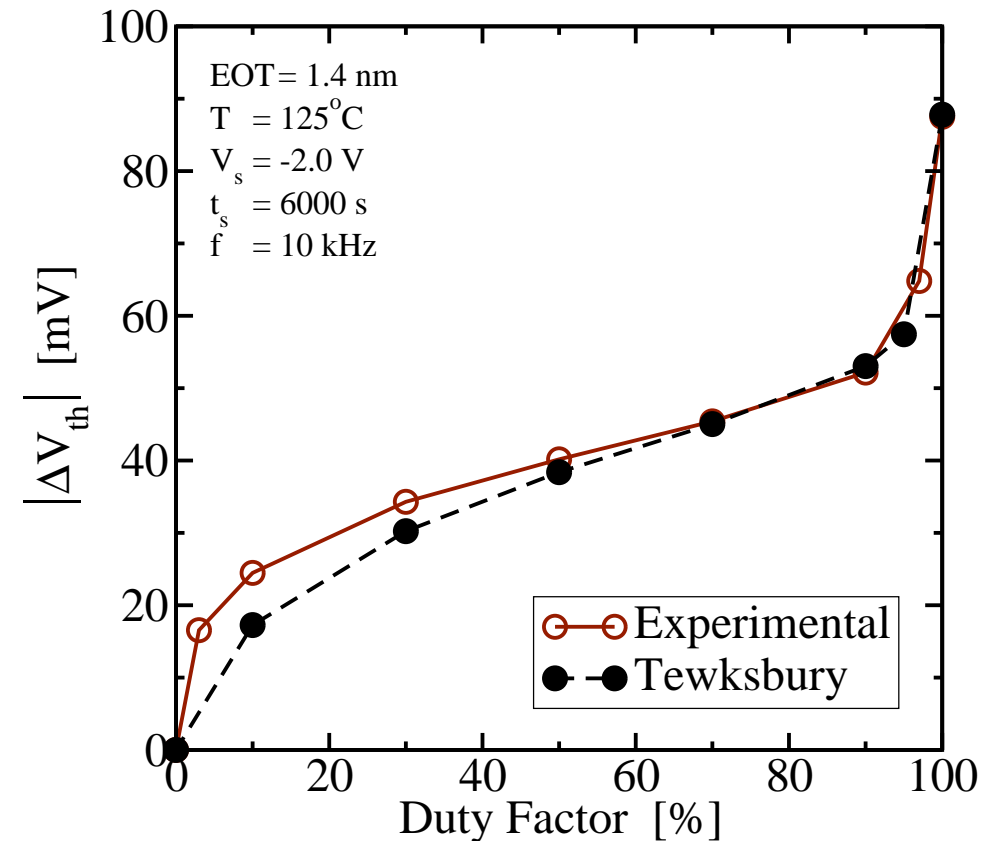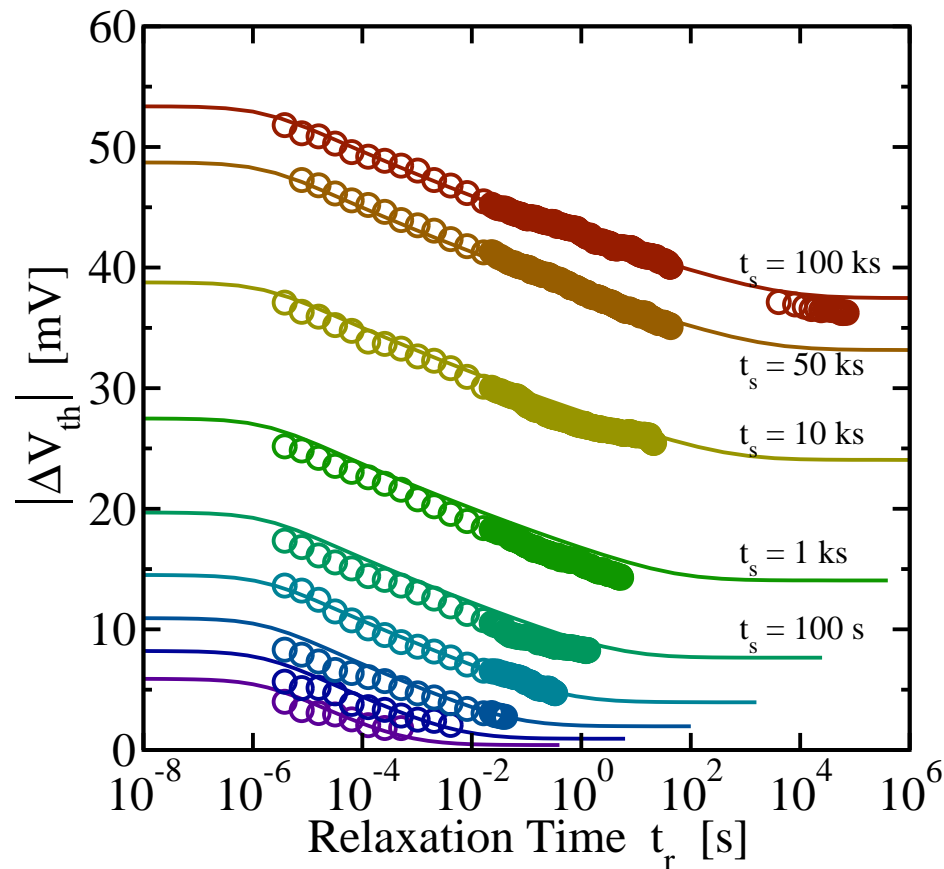
[1] As suggested by Huard *et al.*, MR '06, but using a full numerical solution   [2] Data from Grasser *et al.*, IEDM '07
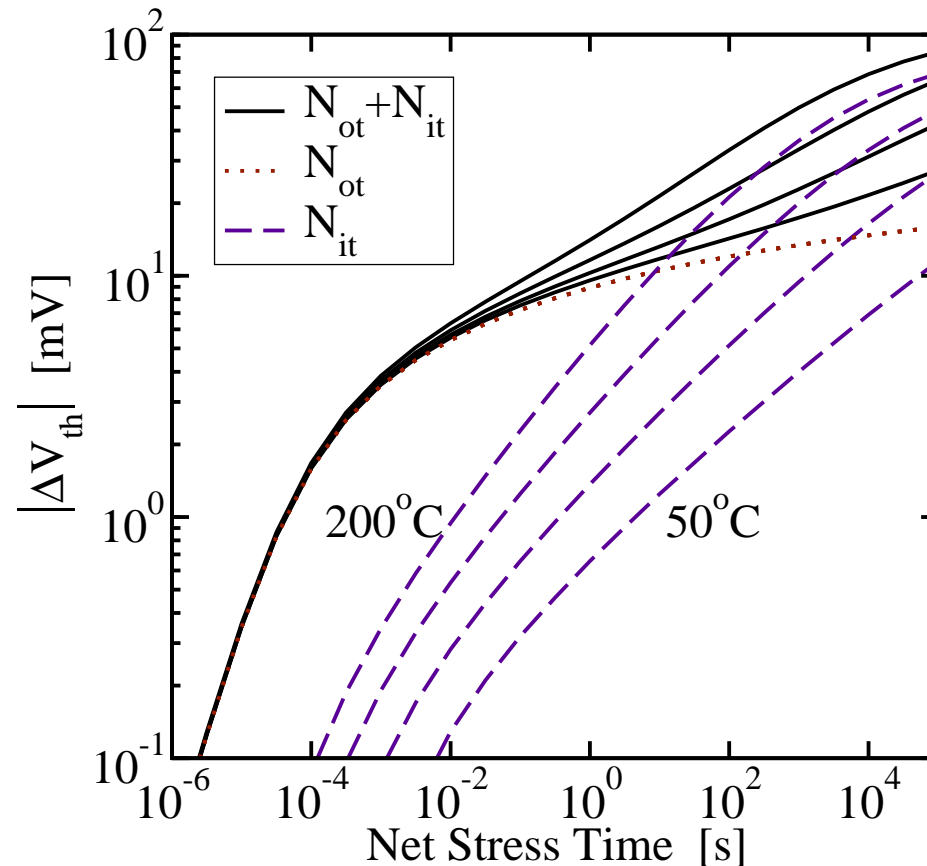
Temperature Dependence: Stress

The measured temperature dependence is not well reproduced

Reason: two independent components

Hole trapping weakly temperature-dependent, N_{it} generation strongly dependent

$$\frac{\Delta N_{it}(t_s)}{\Delta N_{it,max}} \approx \frac{1}{1 + (t_s/\tau)^{-\alpha}} \qquad \alpha = \frac{k_B\,T}{E_0}$$

The overall behavior shows an unscalable, pronounced temperature dependence
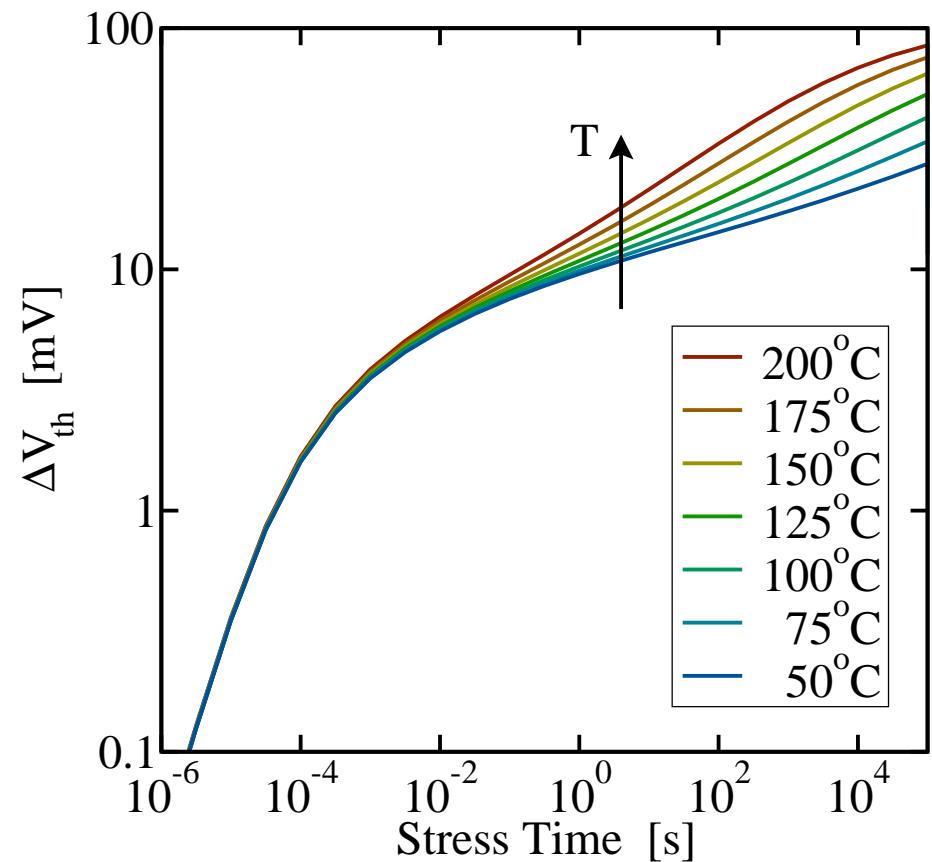
Temperature Dependence: Stress

Theory predicts different temperature acceleration at short and long times

The initial response is log-like but with a very small temperature dependence
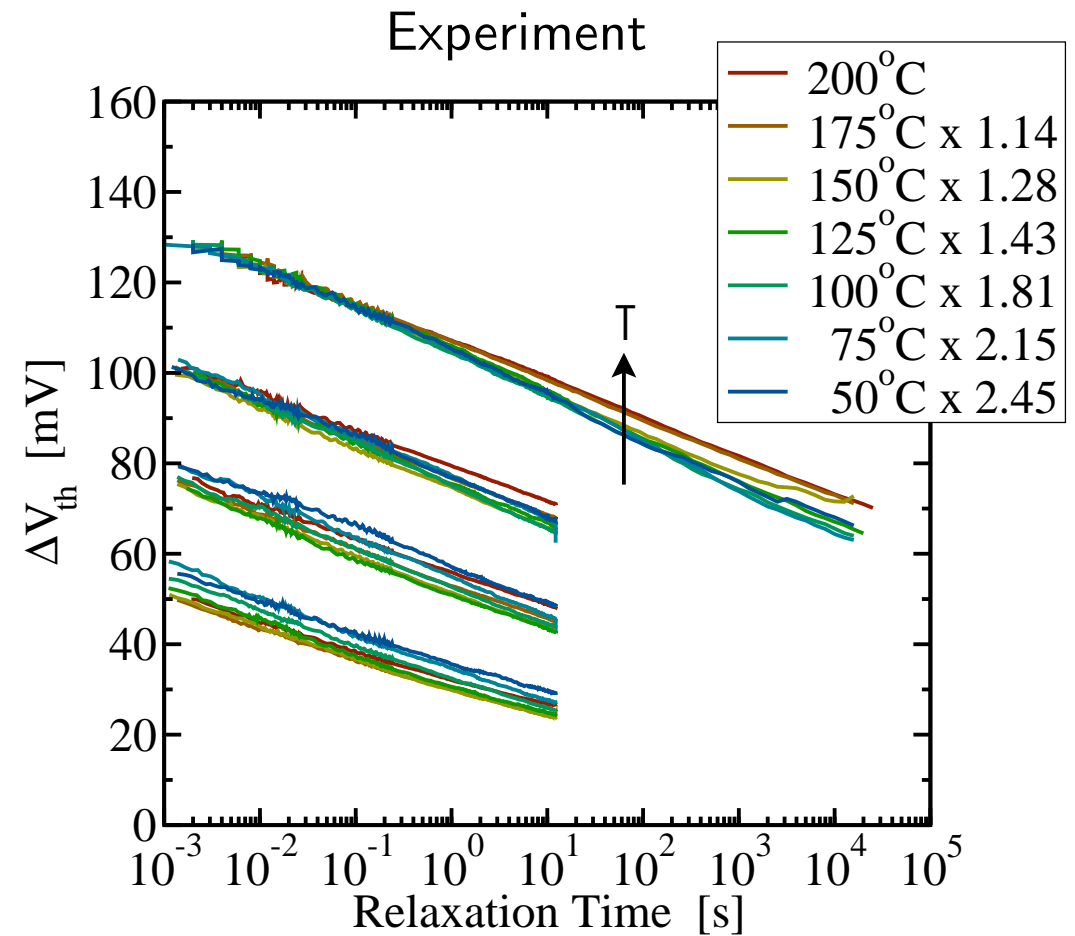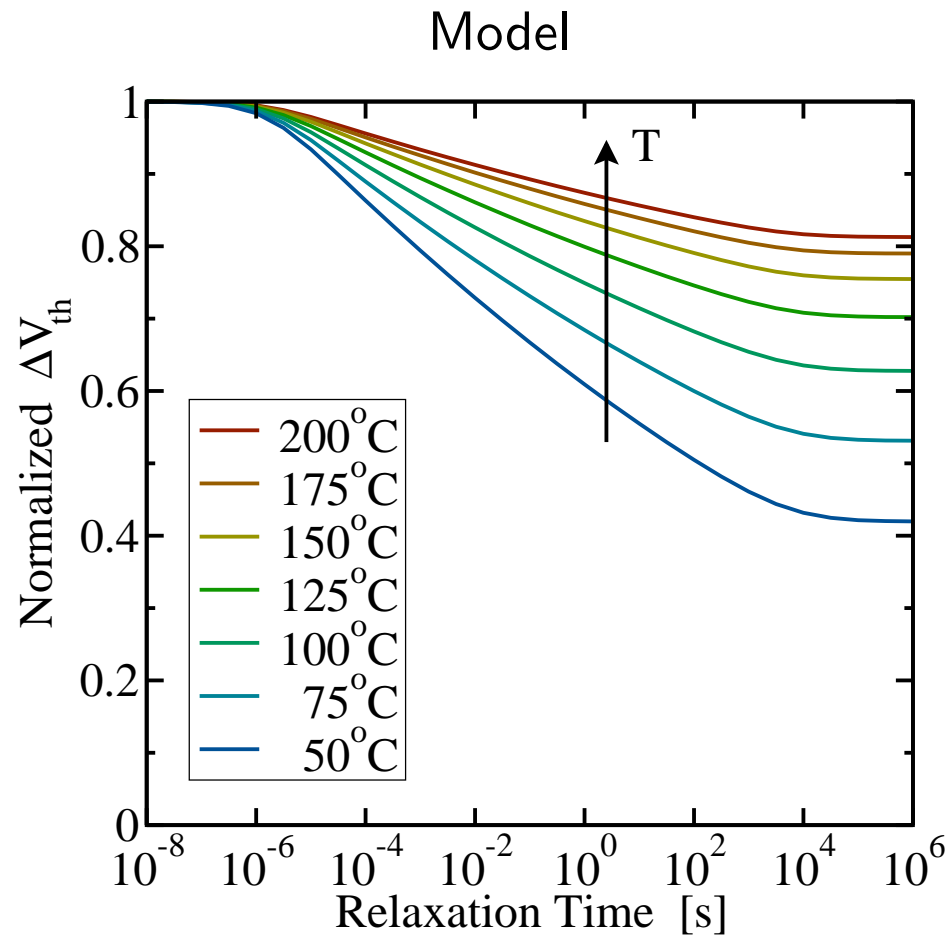
The long-time 'slopes' are strongly temperature-dependent
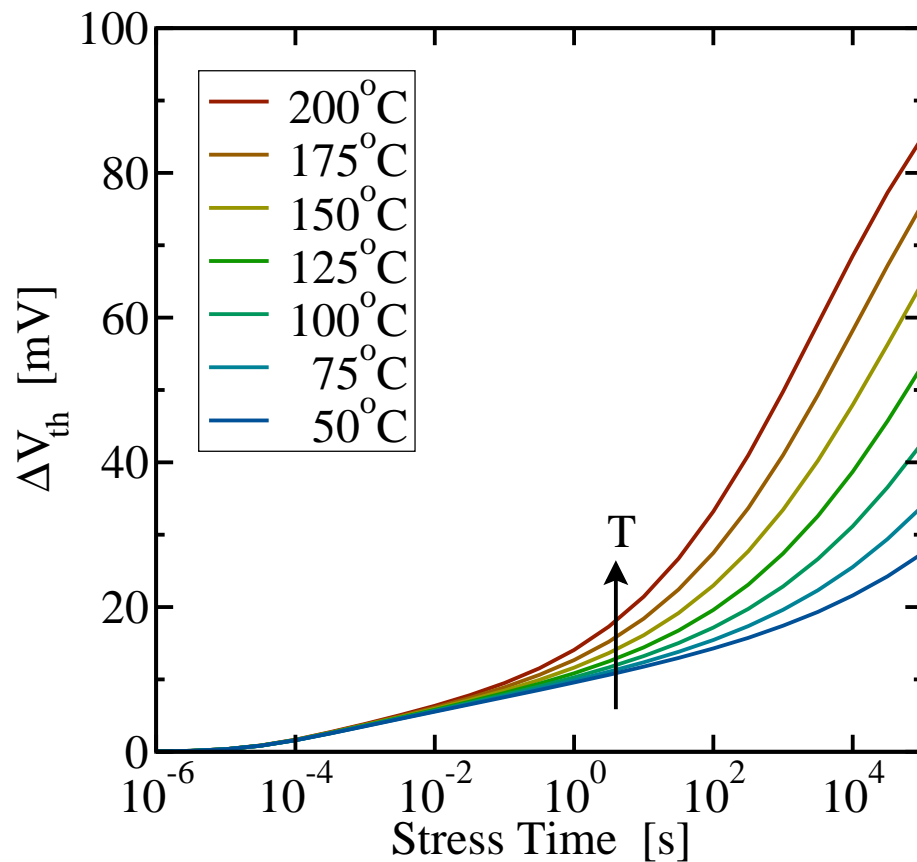
Temperature Dependence: Recovery

The strong temperature dependence of N_{it} transfers directly to relaxation

This is contrary to our experiments, which only show a weak impact of temperature

Model

Experiment

Caveats Tewksbury's Hole Trapping Model

Previous results are a (stupid?) fit to measurement data

What is the physical meaning of the calibrated model parameters?

'Original' parameters: maximum charge storage time of $\sim 1\,\mathrm{ms}$ for $t_{ox} = 2\,\mathrm{nm}$

Recall: model originally developed for 'thick' ($\sim 30\,\mathrm{nm}$) SiO_2 oxides[1]

How can a modern ultra-thin oxide sustain trapped charge for $10^6\,\mathrm{s}$? [2]

Are there some level-shifts involved? [3] [4] [5] [6]

Due to structural relaxation, trapped holes may become very difficult to anneal

[1] Tewksbury and Lee, SSC '94   [2] Islam *et al.*, '07   [3] Blöchl, PRL '99   [4] Zhang *et al.*, TED '04

[5] Ang, EDL '06   [6] Gös *et al.*, IIRW '07

Model Checklist: Combined Model

Stress: undelayed measurements

- ✓ Log-like initial behavior
- ✗ Temperature and voltage independent slope, scaling property

Stress: delayed measurements

- ∼ Temperature-dependent power-law exponent/slope (proportional to temperature?)
- ✓ Delay distorts data even at larger stress times

Recovery

- ✓ Long relaxation tails of log-like nature (more than 12 decades in time?)
- ∼ Strong bias sensitivity, particularly for positive bias
- ✓ Universal behavior (or at least in good approximation)
- ✓ Fast recoverable and slowly recovering/permanent component

Duty factor dependence

- ✓ Plateau around DF = 50%, strong sensitivity for very large and small DF

Summary of Challenges

Summary – I

Detailed analysis of published NBTI models

Models based on classic reaction-diffusion theory

    Cannot capture the complex relaxation behavior

    As a consequence, they cannot capture DF and AC stress behavior

Reaction-dispersive-diffusion models

    Allow to slow down/speed up RD recovery

    No satisfactory results obtained so far during the recovery

    Unclear coupling of dispersive transport mechanism to interface reaction

Reaction-limited models

    Can reproduce change in initial degradation after pre-stress

    Seems to agree with CP data, but CP likely introduces artifacts of its own

The recoverable component is often assumed to be trapped holes

    Inconsistent with detailed relaxation data, at least in a simple picture

    What is the detailed physical mechanism justifying trapped holes for 10^6 s?

Summary – II

A model using two independent mechanisms may be problematic

    How to combine different voltage and temperature activation?

    For what kind of sample does the scaling property hold?

No published model can explain the full complexity

    Published models capture only a subset

    $\Rightarrow$ Incomplete understanding of NBTI

Nevertheless, there is solid proof for defect creation

    A lot of data points at hydrogen as a major player

        Consequently, there is hydrogen-induced bond-breaking and hydrogen motion

    The created defects are electrically active (positively charged)

        Consequently, there is charge exchange with the substrate ('hole trapping')

What are the exact mechanism?

    Which of the above mechanisms controls the dynamics?

    What is the exact physics behind NBTI?

Are we building/have we built our models on 'proper/relevant' data?