Modeling CMOS Non-Quasi-Static Effects in a Quasi-Static Simulation Engine

Mansun Chan[1,2], Wai-Kit Lee[1], Xuemei Xi[2], Jin He[2], Ali Niknejad[2], Chenming Hu[2], and Ping K. Ko[1]

E-mail: mchan@ee.ust.hk

[1] Department of EEE, Hong Kong University of Science & Technology
[2] Department of EECS, University of California at Berkeley

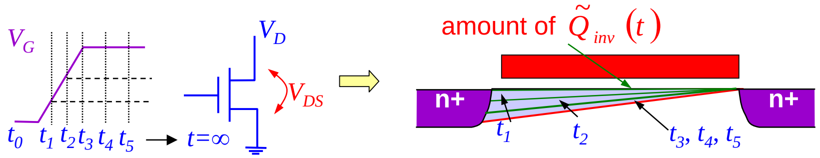
Outline

- What is the NQS problem?
- Formulation of NQS simulation
- Approximate solution for fast transient
- Frequency domain solution
- Summary

The Non-Quasi-Static Problem [1]

what happens at fast transient?

- at slow switching



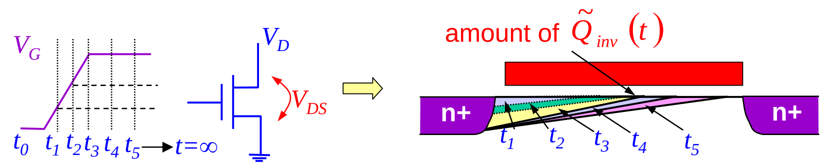
- charge at given time is known once apply voltages are given

$$\tilde{Q}(t) \approx f\left(V_D(t), V_S(t), V_G(t), V_B(t), t=\infty\right) = \tilde{Q}\left(\tilde{V}(t), t=\infty\right)$$

- quasi-static assumption is valid

The Non-Quasi-Static Problem [2]

- at fast switching



- charge at given time is unknown with given apply voltages

$$\tilde{Q}(t) = f\left(V_D(t), V_S(t), V_G(t), V_B(t), t\right) \neq \tilde{Q}\left(\tilde{V}(t), t=\infty\right)$$

- quasi-static assumption is invalid

The Non-Quasi-Static Problem [3]

what is quasi-static assumption?

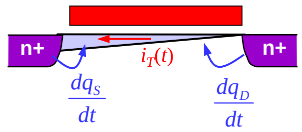
- most precisely defined by Yannis Tsividis
  - let $V_D(t)$, $V_S(t)$, $V_G(t)$, $V_B(t)$ be the varying terminal voltages; then at any position, the charges per unit area at any time t are assumed identical to those that would be found if DC voltages $V_D=V_D(t)$, $V_S=V_S(t)$, $V_G=V_G(t)$, $V_B=V_B(t)$ were used instead
  - that is: we assume charge is a memoryless state variable
- why quasi-static assumption?
  - SPICE does not keep track of the time, and only guesses of voltages are provided to solve the linear iterative equation

$$\tilde{g}^m \tilde{V}^{(m+1)} = -\tilde{I}^m + \tilde{g}^m \tilde{V}^m$$

Approach of QS Approximation

calculation of drain current

- compose of a transport current plus a charging current



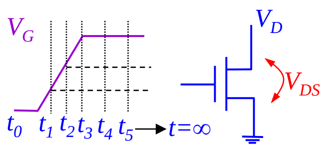
$$\Rightarrow\ i_D = i_T(t) + \frac{dq_D}{dt}(t)$$

- 2 problems:
  - how to calculate the transport current?
  - we know how to calculate the channel charge (to some degree), but don't know how to partition the channel charge to the source and drain

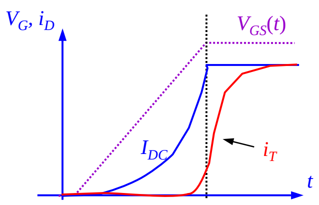
Approximating Transport Current

- assume the transport current is the same as the DC current at the given voltage (QS approximation)

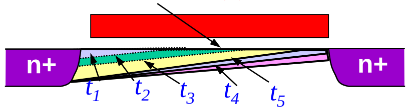
$$i_T(t) \approx I\left(\tilde{V}(t), t=\infty\right) = I_{DC}\left(\tilde{V}(t)\right)$$



- problem:
  - check current at different time



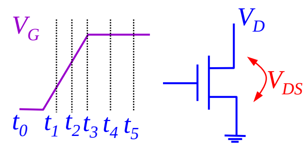
amount of $\tilde{Q}_{inv}(t)$



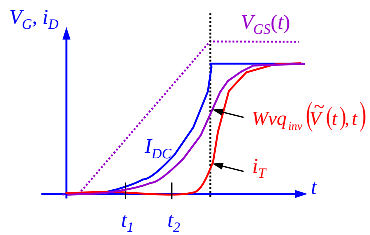
Another Attempt of Approximation

- use transient charge to calculate the transport current

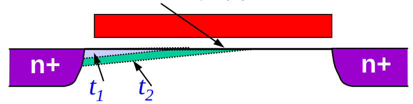
$$i_T(t) \approx Wvq_{inv}\left(\tilde{V}(t), t\right)\Big|_{\text{some location } y}$$



- problem:
  - knowing the amount of charge alone is insufficient, and the spatial distribution is also required



amount of $\tilde{Q}_{inv}(t)$



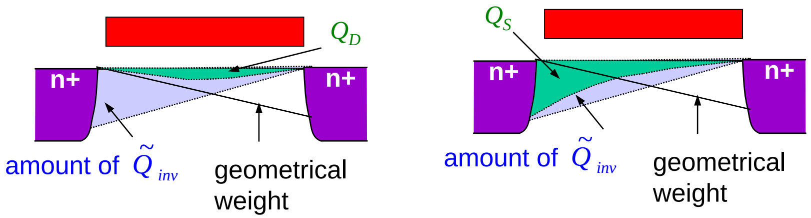
- location information also not available in SPICE

Charge Partition Scheme [1]

Charging current: Fixed charge partition

- 40/60 charge partition
  - based on geometrical weighted integral

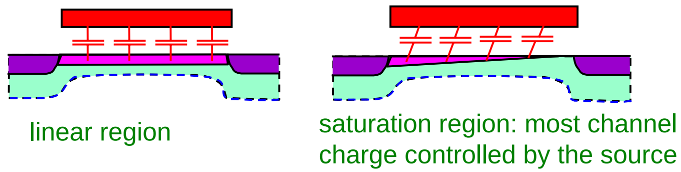
$$Q_D = \int_0^L \frac{y}{L} q_i(y)dy \quad \text{and} \quad Q_S = \int_0^L \left(1 - \frac{y}{L}\right) q_i(y)dy$$



Charge Partition Scheme [2]

- 0/100 charge partition
  - missing part of 40/60 charge partition

  - based on pinch-off approximation

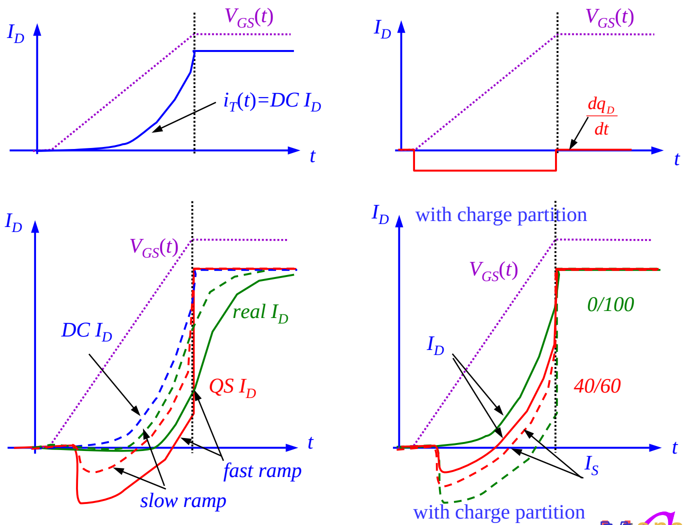


- 50/50 charge partition
  - sort of arbitrary and seldom used these days

Simulation result with Charge Partition

Results of QS approximations

- individual component
- final output



Measuring the Charge Partition [1]

How to measure the NQS charge partition

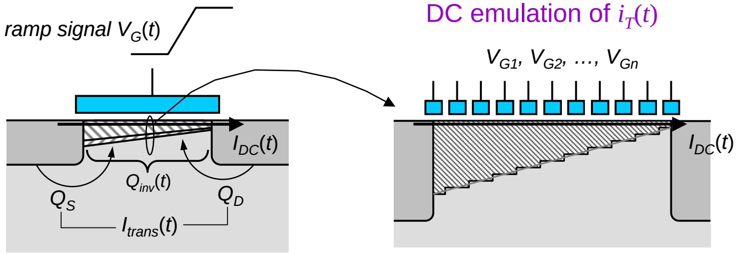
- can be measured by the drain and source current

$$\begin{cases} i_D = i_T(t) + \frac{dq_D}{dt}(t) = i_T(t) + x\frac{dq_{inv}}{dt}(t) \\ i_S = -i_T(t) + \frac{dq_S}{dt}(t) = -i_T(t) + (1-x)\frac{dq_{inv}}{dt}(t) \end{cases}$$

- the 2 equations are linearly dependent
- need to find $-i_T(t)$ in order to find the value of x

Measuring the Charge Partition [2]

By using 2-D simulator



$$i_D = i_T(t_p) + x\frac{dq_D}{dt}(t_p) \qquad i_D = i_T(t_p) = I_{DC}$$

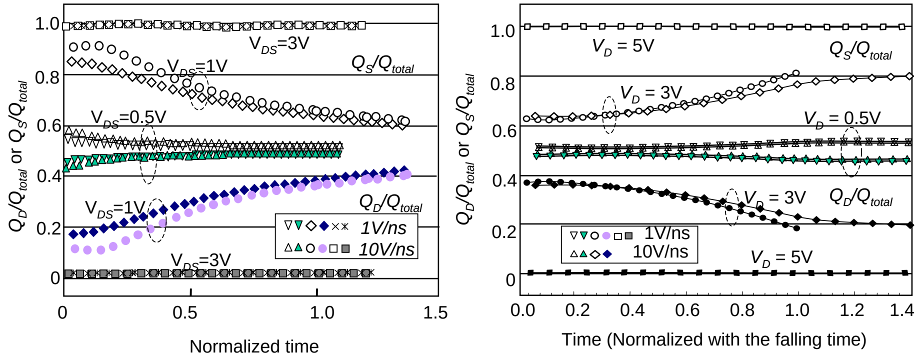
$$i_S = -i_T(t_p) + (1-x)\frac{dq_S}{dt}(t_p) \qquad i_S = i_T(t)$$

- drain/source current ratio can be found by subtracting current in the left figure by current in the right figure

Actual Channel Charge Partition

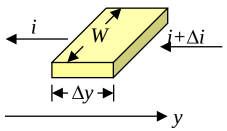
Turn-on and Turn-off charge partition ratio

- Charge partition is dynamic dependent on ramp rate



Exact Solution of MOSFET

Consider strong inversion only



- continuity equation $\frac{\partial i(y,t)}{\partial y} = W\frac{\partial q_i(y,t)}{\partial t}$
- drift diffusion $i(y,t) = \mu W q_i(y,t)\frac{\partial v_{CB}(y,t)}{\partial t}$
- charge control equation

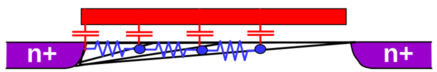
$$q_i(y,t) = C_{ox}\left(v_{GB}(t) - V_{FB} - \phi_0 - v_{CB}(y,t) - \gamma\sqrt{\phi_0 + v_{CB}(y,t)}\right)$$

- terminal currents are given by $i_D(t) = i(L,t)$ and $i_S(t) = i(0,t)$
- total inversion charge is given by $Q_{inv} = \int_0^L q_i(y,t)dy$
- question: how to incorporate this into a compact model?

Approximate Compact Model [1]

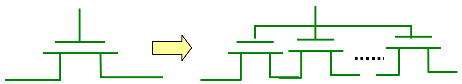
By using distributed network

- using more internal nodes to keep track of the timing



- need charge and I-V expression at each internal node

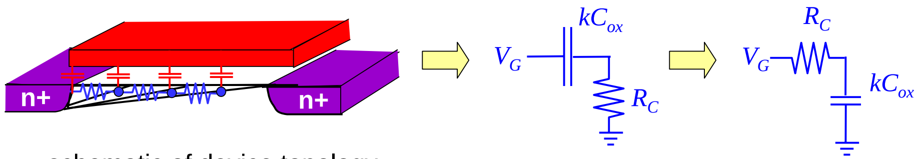
Common implementation



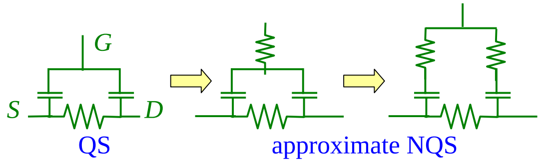
- transport current remain Quasi-static
- need a correct charge partition model

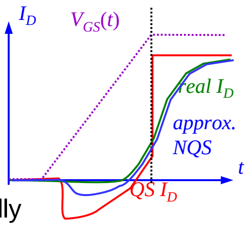
Approximate Compact Model [2]

Further simplification



- schematic of device topology

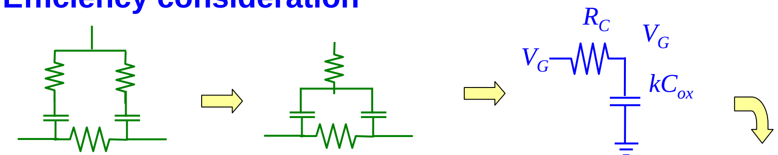




- successfully includes the delay, but remains physically incorrect QS transport current and charge partition scheme

Implementation of NQS Model in BSIM

Efficiency consideration



requires 2 additional nodes

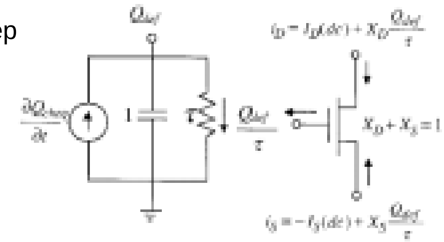
inflexible to include charge partition

Using external circuit

- A state variable Q_{def} is created to keep track of the deficit charge and the charging current is calculated by

$$\frac{dQ_{def}(t)}{dt} = \frac{dQ_{cheq}(t)}{dt} - \frac{dQ_{ch}(t)}{dt}$$

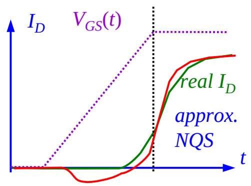
and $$\frac{dQ_{ch}(t)}{dt} \approx \frac{Q_{def}(t)}{\tau}$$



Limitation in BSIM's NQS Approach

What can be achieved using the BSIM NQS model?

- Correct delay in digital switching can be predicted



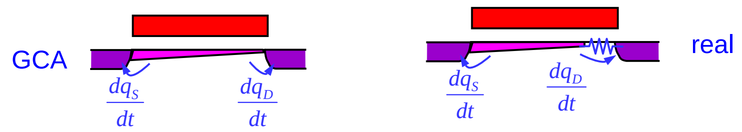
Limitation of the BSIM NQS approach

- The use of $I_D(dc)$ is physically incorrect
- Charge partition should be dependent on the drain voltage (linear region is 50/50), but initial implementation in BSIM does not allow it
- Need to partition the deficit charge to the gate and substrate, which is not done in BSIM
- Can predict the delay in digital switching (for speed estimation) but no guarantee the waveform is correct

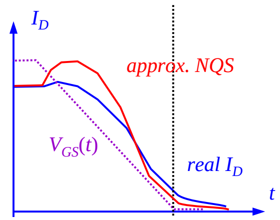
Effect of Velocity Saturation

Limitation of the gradual channel approximation

- The gradual channel approximation assume the channel is in-contact with the drain terminal at "pinch-off"
- Velocity saturation introduce extract resistance at the drain, as current cannot travel faster than v_{sat} to the drain



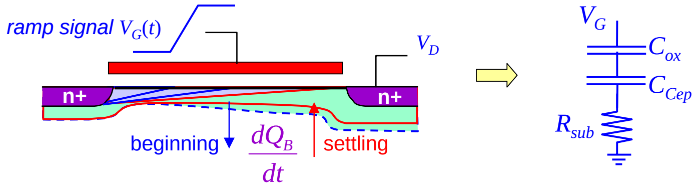
- Observation during turn-off, an overly large drain current overshoot is observed in the approximate model



Action of Depletion Charge

At the drain region during turn-on

- The drain region at high drain voltage can enter deep-depletion before the charge from the source arrived

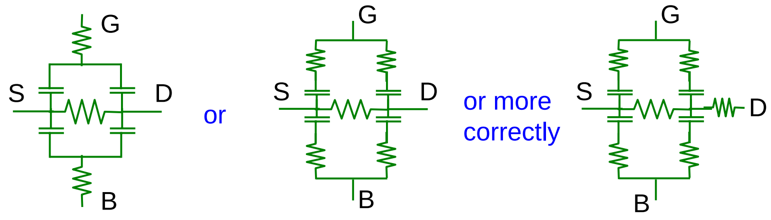


- Initially, depletion charge extended beyond the equilibrium through a current path controlled by the substrate resistance
- Once the carriers from the source reached the drain, the depletion is converted back to the inversion charge
- Not only affecting I_{sub}, but also create a dynamic effect in the threshold voltage or surface potential

Modification of Approximate NQS Model

Required enhancement in approximate model

- need one more node to include the action of substrate charge



- Actual implementation is more complicated than the above circuit as the distributed resistance from the drain to the channel is required to allow the coupling between the gate and substrate

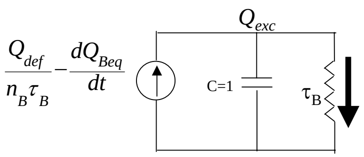
Unpublished BSIM Enhancement

- A new variable Q_{exc} has been defined to keep track of the excess substrate charge from equilibrium

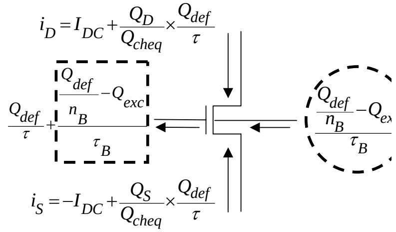
$$\frac{dQ_{exc}}{dt} = \frac{dQ_B}{dt} - \frac{dQ_{Beq}}{dt}$$

where $\frac{dQ_B}{dt} = \left.\frac{dQ_B}{dt}\right|_{in} + \left.\frac{dQ_B}{dt}\right|_{out} \approx \frac{\frac{Q_{def}}{n_B} - Q_{exc}}{\tau_B}$

therefore $\frac{dQ_{exc}}{dt} \approx \frac{\frac{Q_{def}}{n_B} - Q_{exc}}{\tau_B} - \frac{dQ_{Beq}}{dt}$



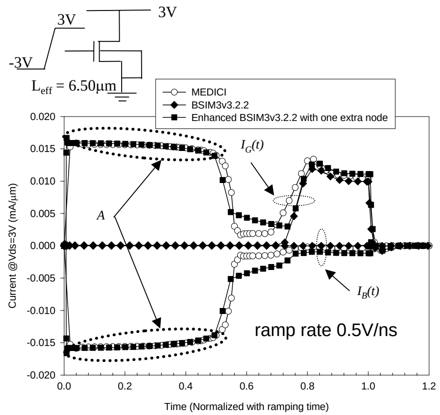
- The substrate current can be calculated by the another sub-circuit as shown

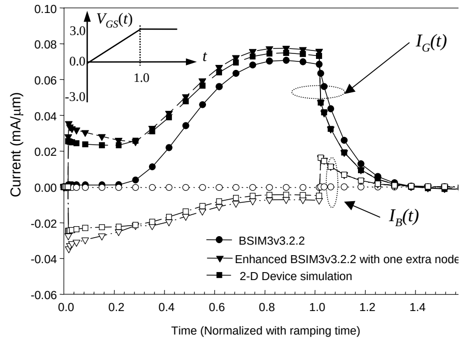


Gate and Substrate Currents

Result of the BSIM enhancement

- Turn-on of a 6.50μm NMOS transistor with input ramp rate 0.5V/ns

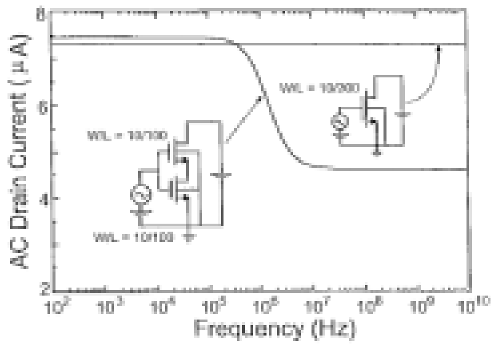




ramp rate 5V/ns

What About Small Signal AC Simulation?

Pointed out by Tsividis



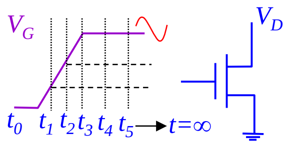
- This is a mis-conception of NQS effect and this behavior can be modeled well by adding a pole by the inclusion of gate resistance as in the BSIM model

NQS Effect in AC Simulation?

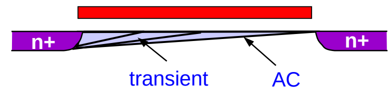
Recall definition of QS assumption

$$\tilde{Q}(t) = f\left(\tilde{V}(t), t\right) \approx f\left(\tilde{V}(t), t = \infty\right)$$

- in AC, all node voltages are equal to the voltages at t=∞, thus the above equation becomes an equality



$$\tilde{Q}(t = \infty) = f\left(\tilde{V}(t = \infty), t = \infty\right)$$

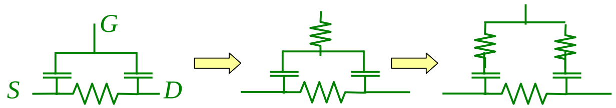


- Strictly speaking, NQS refers to Large Signal non-equilibrium
- AC small signal simulation is always static
- NQS effects may become important when large signal simulation is used to general high frequency response by using Fourier Transforms

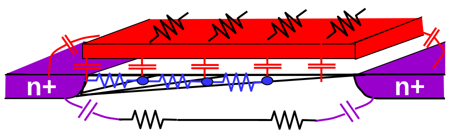
Modeling MOSFET for HF AC Simulation

Real problem:

- How to model the distributed RLC network after reaching equilibrium!!
- It is really quasi-static
- The distributed RC approach works pretty well



- More internal nodes will help



- Question: Can we have a separate large signal and small signal model with different number of nodes?

Result of Distributed RC Network Model

How Serious is the NQS Problem?

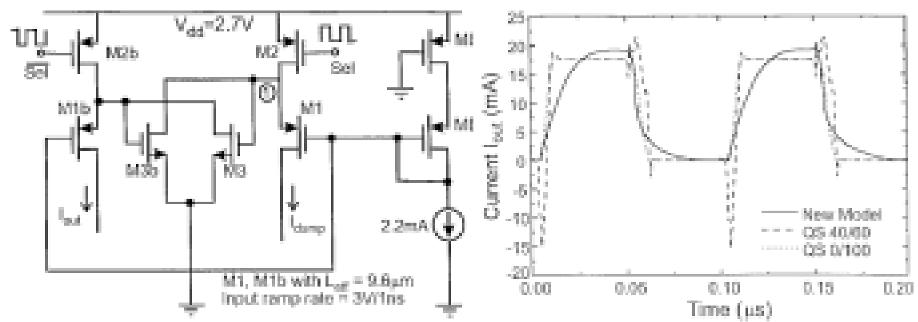
For Digital Circuits

- never happens in self driven switching circuits (most circuits)
- only happens when a small (fast) device is driving a large (slow) device
- conclusion: can be ignored in most digital circuits

For High Frequency Circuits

- Only in large signal switching circuits (like oscillator) that use Fourier Transform to obtain the frequency response
- QS approximation works pretty well up to about $0.05f_T$
- with increasing f_T in nano-CMOS, *NQS* problem maybe over worried

Example Circuit that Require NQS Model



- An A/D current source with $0.6\mu m$ technology
- The delay is correctly predicted, but no guarantee on waveform

Summary

- The current compact modeling approach are inherently Quasi-Static
- NQS effects is a result of large signal non-equilibrium situation during switching
- Most Compact NQS model uses the distribute RC network to emulate the NQS switching, which provide some degree of similarity despite incorrect RC elements used
- Frequency domain simulation is always Static, and no NQS event in a strict sense
- The NQS problem is somehow over exaggerated