Compact Thermal Models for Estimation of Temperature-dependent Power/Performance in FinFET Technology

Presenter: Kunhyuk Kang

Aditya Bansal, Mesut Meterelliyoz, Siddharth Singh*,
Jung Hwan Choi, Jayathi Murthy§, Kaushik Roy

School of Electrical and Computer Engineering, Purdue University,
§Department of Mechanical Engineering, Purdue University
West Lafayette, IN 47907
*Department of Electrical Engineering, Osmania University, India

Outline

- Introduction
- Power/Performance vs. Temp.
- Compact Thermal Models
- Generating Temperature Map
- What next?

Thermal Issues

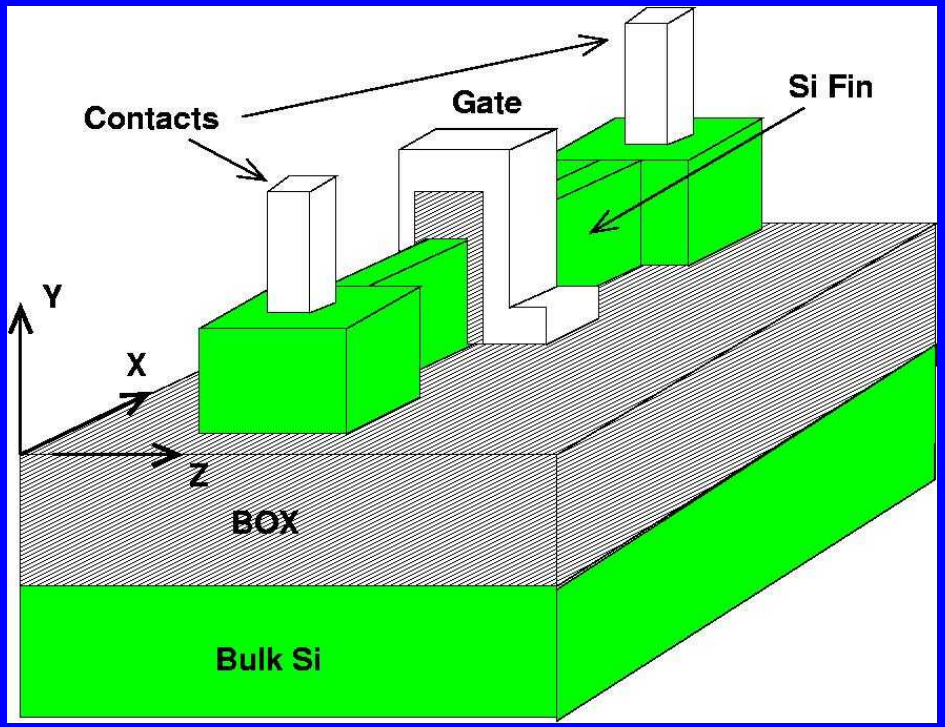
- High chip temperatures and gradients have serious consequences for transistor and circuit performance
- Every 10°C rise decreases drive current by 4% and increases interconnect delay by 5%
- Leakage power depends exponentially on temperature
- Cooling devices does not support a detail granularity
- Objective is to directly integrate thermal analysis with circuit design

Why Double-Gate SOI over Bulk-MOSFETs?

Electrical properties:

- Improved Short Channel Effects
- Undoped body: no random dopant fluctuations
- Higher on-current (higher mobility and two gates)
- Relatively less gate tunneling leakage (relaxed constraint on oxide thickness)
- More feasible to fabricate with the advent of FinFET technology

FinFET Structure

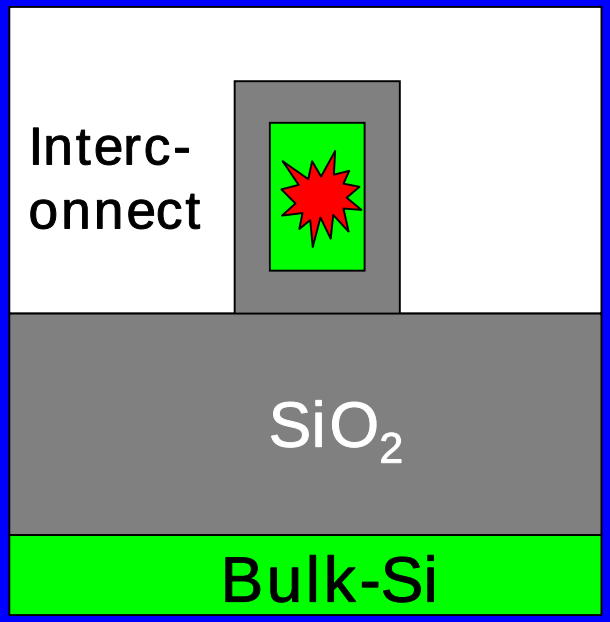


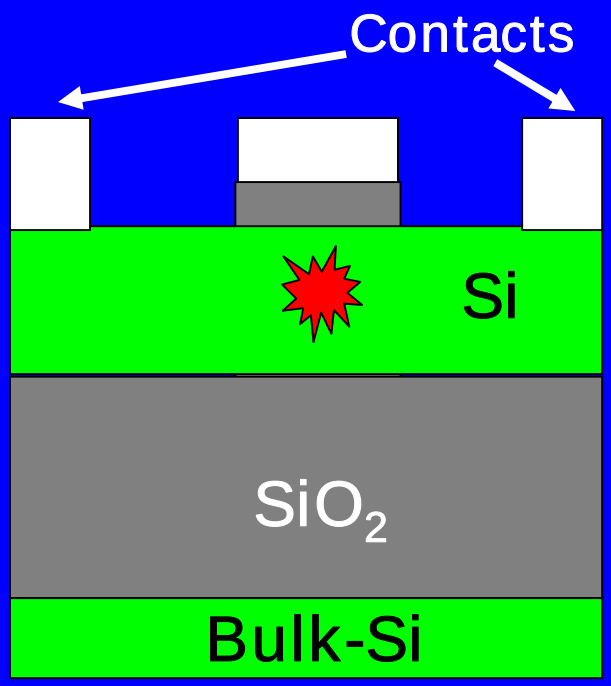
- Gate insulator all around the channel
- Increased surface to volume ratio (thermal boundary resistances of various materials)
- Device layer is separated from bulk-Si

Thermal issues in FinFETs

In Vertical plane:

Heat dissipates through SiO_2 and interconnects





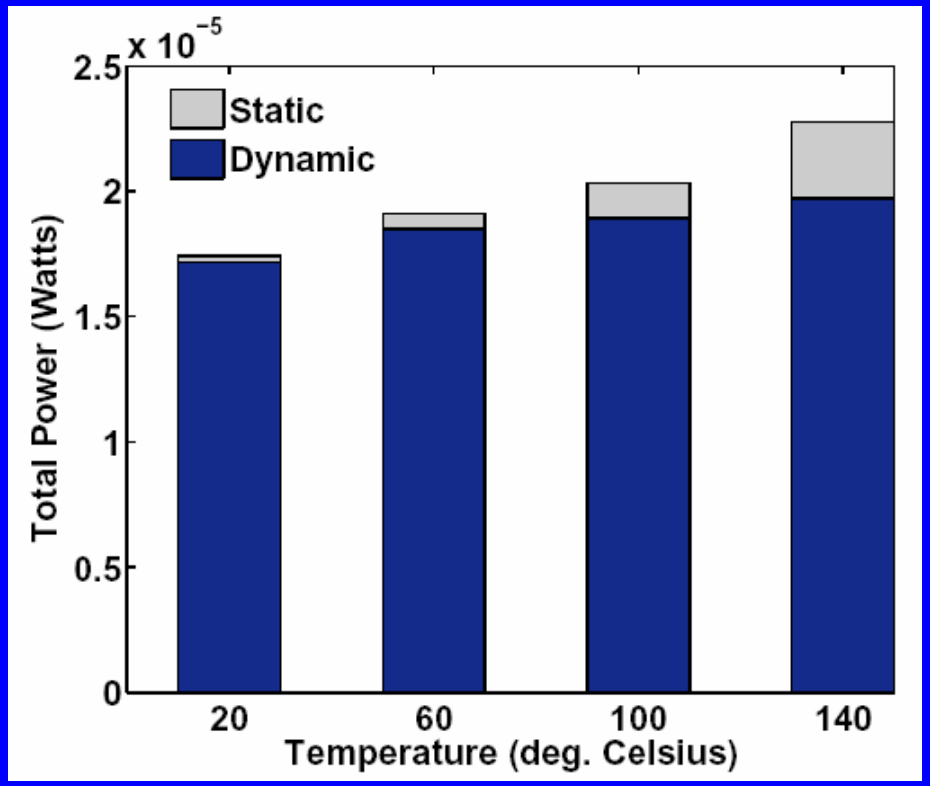
In Horizontal plane:

Heat dissipates through contacts and interconnects

Outline

- Introduction
- Power/Performance vs. Temp.
- Compact Thermal Models
- Generating Temperature Map
- What next?

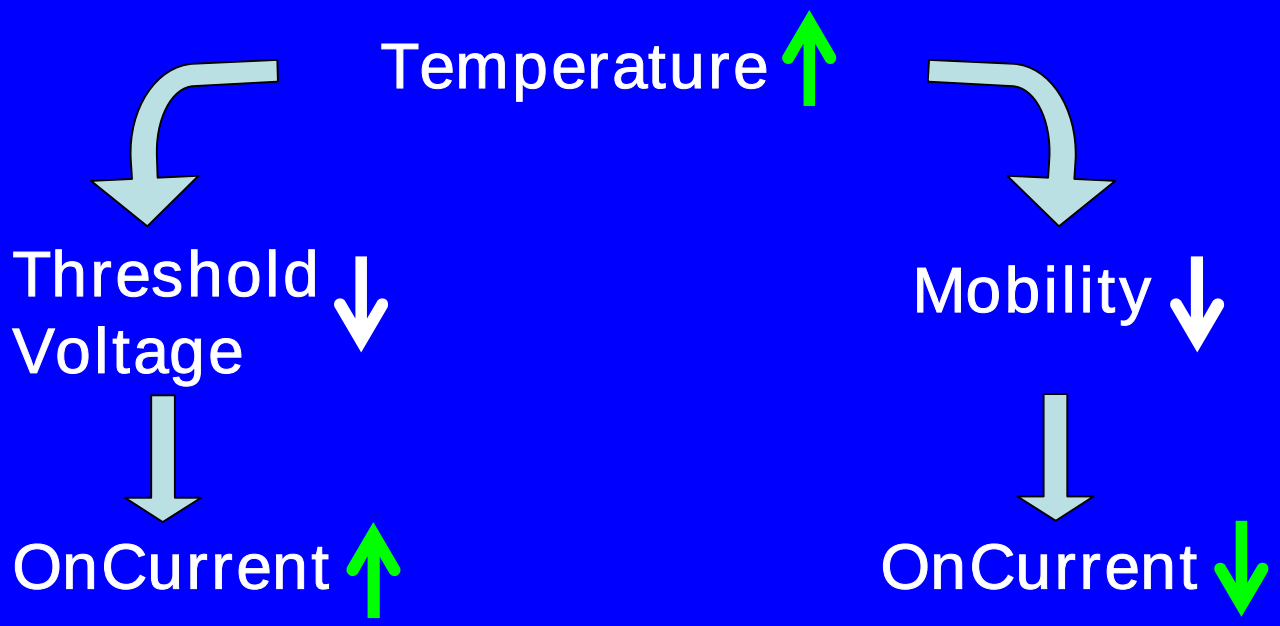
Power vs. Temperature



- Static power (subthreshold leakage) increases exponentially with temperature

- Short-circuit component of dynamic power increases because of increased sub-t slope and delayed output transition

Performance vs. Temperature



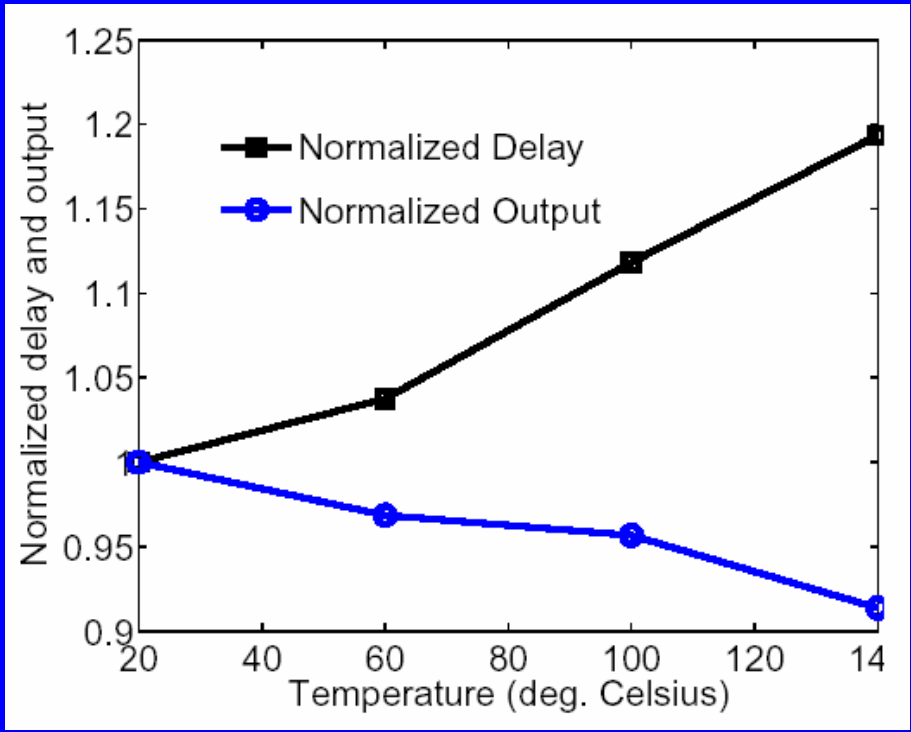
Delay?

(Relative Sensitivities)

Performance vs. Temperature

For 120 C rise in temperature:

- Intrinsic delay increases 20%
- Output Swing decreases 9%

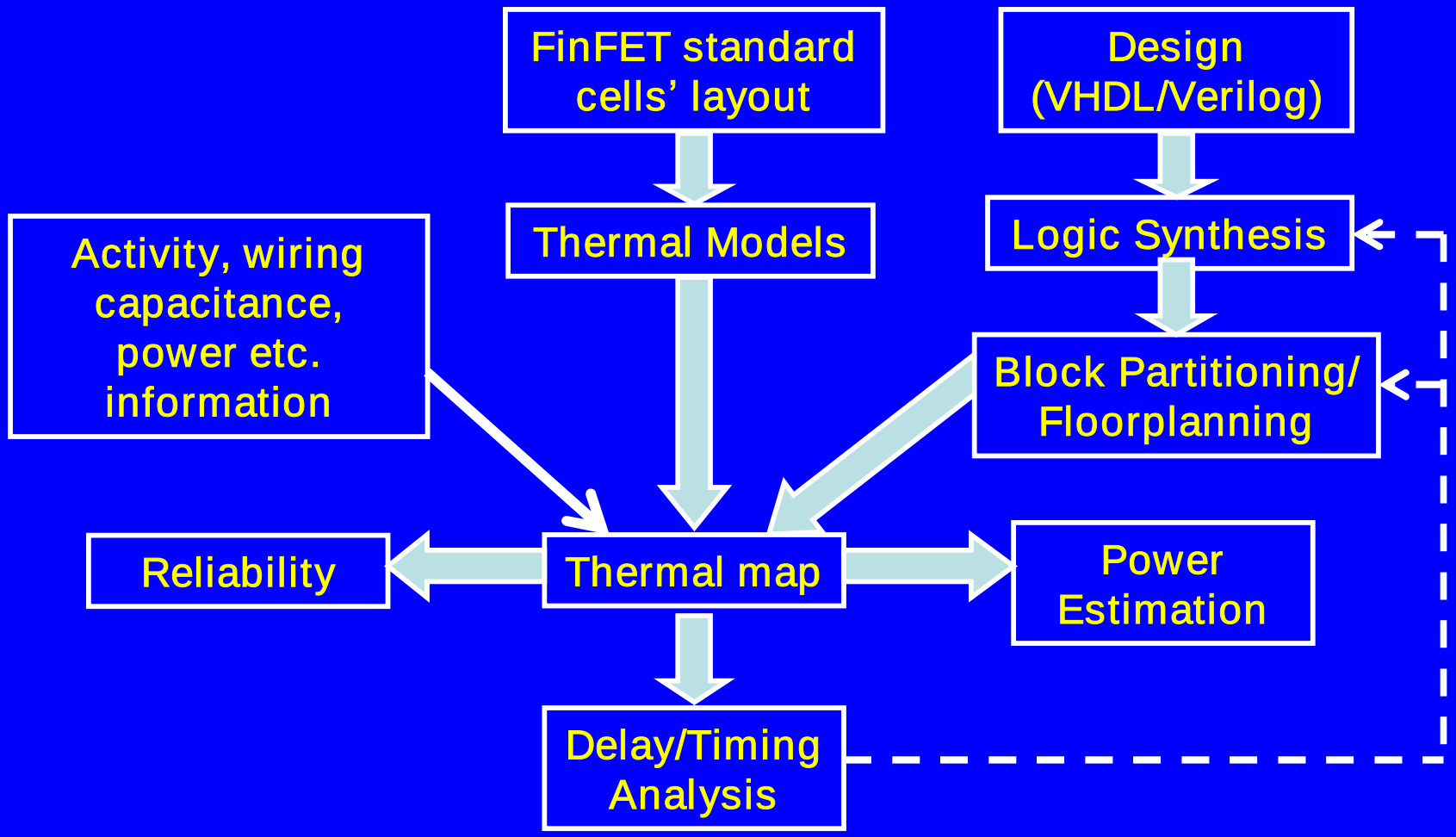


3 NOR Stage Ring Oscillator

Why Thermal Modeling?

- Electro-thermal co-design of transistors and circuits is necessary
- A typical chip may have hundreds of millions of transistors => direct thermal simulation is impractical
- Need for compact thermal models to predict power/performance with temperature

Thermal Aware Design



Outline

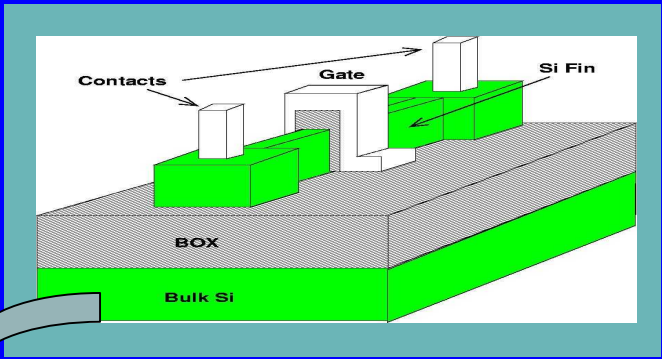
- Introduction
- Power/Performance vs. Temp.
- Compact Thermal Models
- Generating Temperature Map
- What next?

FinFET Cell Layout

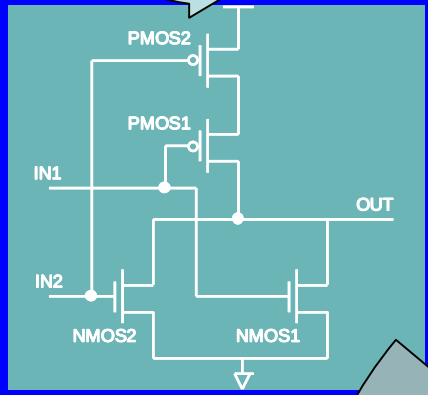
- Technology node: 28nm ITRS 2004
- Layouts of INV and 2-input NAND and NOR gates
- Metal 1 layer is included for horizontal heat flow

jym1

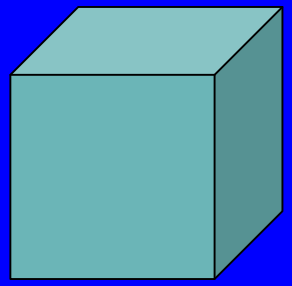
Gate Level Compact Models



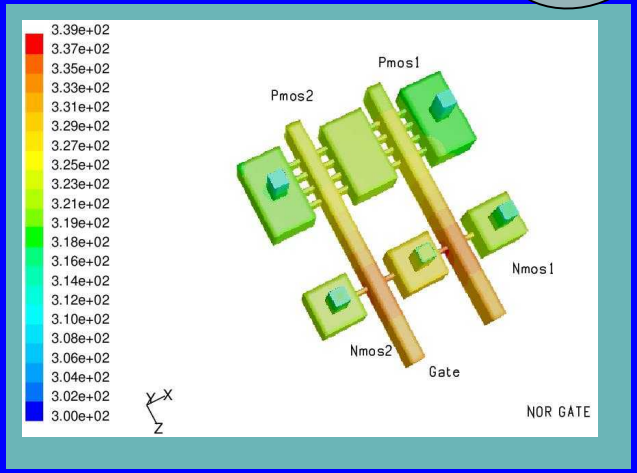
Transistor Level: FinFET



NOR gate composed of PMOS and NMOS FinFETs



Compact thermal model of NOR gate



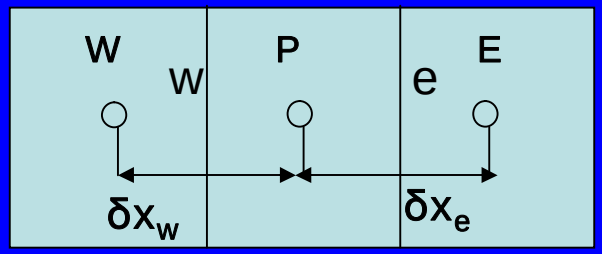
Detailed thermal model of NOR gate

jym1 Finfets or other transistors are used in creating the gate-level circuit, i.e., NOR gates, NAND gates and inverters, which is components used in designing circuits. Detailed thermal models are created for the gate, in this case a NOR gate. The thermal model can be made at any level of sophistication -- either Fourier or BTE or a mix. The heat source in the thermal model is at present obtained from a TAURUS drift-diffusion device simulation, but there too, more complicated BTE models for electrical simulation can be used if necessary.

jym, 2005/12/03

Compact Modeling Methodology

- Create detailed Fourier thermal models using CFD code FLUENT
  - Device layer, some metallization, SiO_2
- Finite Volume Method (Example for 1-D)



Consider Diffusion Equation:

$$\frac{d}{dx}\left(\Gamma \frac{dT}{dx}\right) + q = 0$$

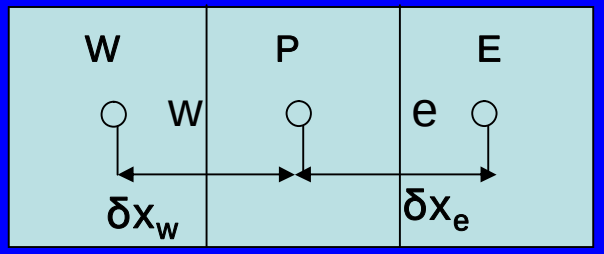
Γ: Diffusion Coefficient
q: Heat Generated

Compact Modeling Methodology (cont'd)

Step 1: Integrate over Control Volume

Step 2: Make linear profile assumption between cell centroids for T

Step 3: Find the algebraic equation



$$T_P = \frac{a_E}{a_P} T_E + \frac{a_W}{a_P} T_W + b$$

$$a_E = \Gamma_e / \delta x_e$$

$$a_W = \Gamma_w / \delta x_w$$

$$a_P = a_E + a_W$$

$$b = q\Delta x$$

Compact Modeling Methodology (cont'd)

- For each gate, avg. temp. and heat flow rate on 6 boundary surfaces of calculation domain expressed in terms of temperature boundary conditions and heat generation:

$$T_{avg} = \sum_{f=1}^{6} a_f T_f + a_0 \alpha q_0$$

$$q_{bj} = \sum_{f=1}^{6} b_{fj} T_f + b_{0j} \alpha q_0 \qquad j=1,2,...6$$

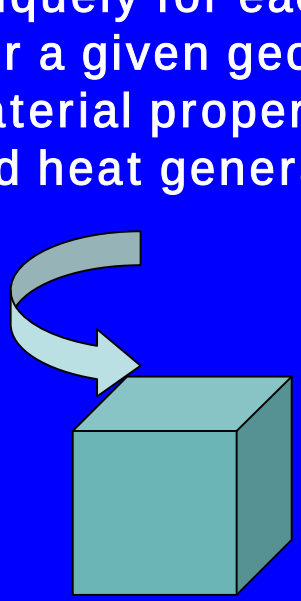
Compact Modeling Methodology (cont'd)

- Compute temperature field for seven sets of temperature boundary conditions to find coefficients in compact model (a_f, a_o, b_{fj}, b_{0j})
- Similar methodology used for substrate cells containing SiO_2
- Can also use the same idea to characterize metallization cells, bulk Si cells and bulk CMOS devices

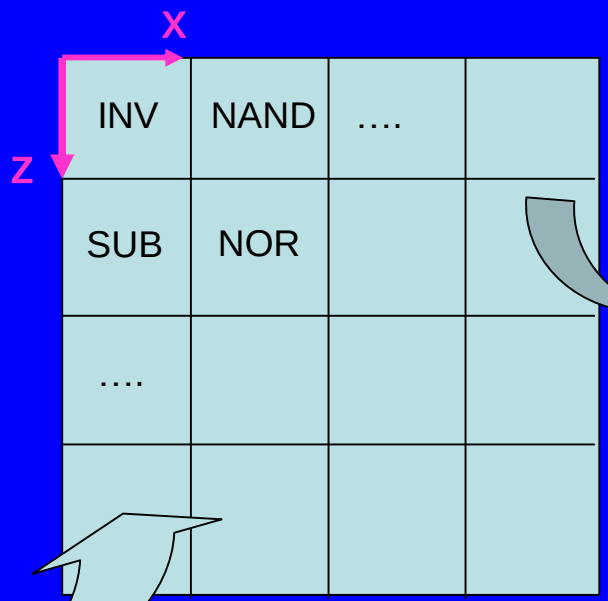
jym2

Multiscale Simulation

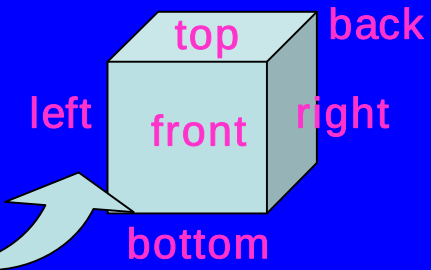
Determine coefficients uniquely for each cell (for a given geometry material properties and heat generation)

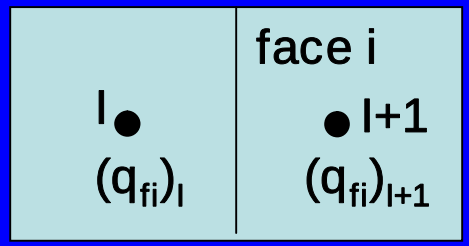


Compact model



Floorplan for the circuit





Cell and face heat balances yield coupled equation set for cell temperatures

jym2

Compact models of the gates are included in a floorplan. Each thermal compact model relates the cell temperature to that of its six faces, its heat generation rate and the gate's activity level. By writing heat balances at the faces, we can obtain a coupled eqn set for all the cell temperatures in the domain. As of now, our floor plan has only one layer of cells, but we can easily include several layers, and put in compact models for spot coolers, metallization layers etc. We can also recursively create compact models of each floor plan, all the way to the chip level if desired. Conversely, once we have the temperature prediction on the floow plan, we can go down to the transistor level and predict its performance more accurately in any region of the circuit.

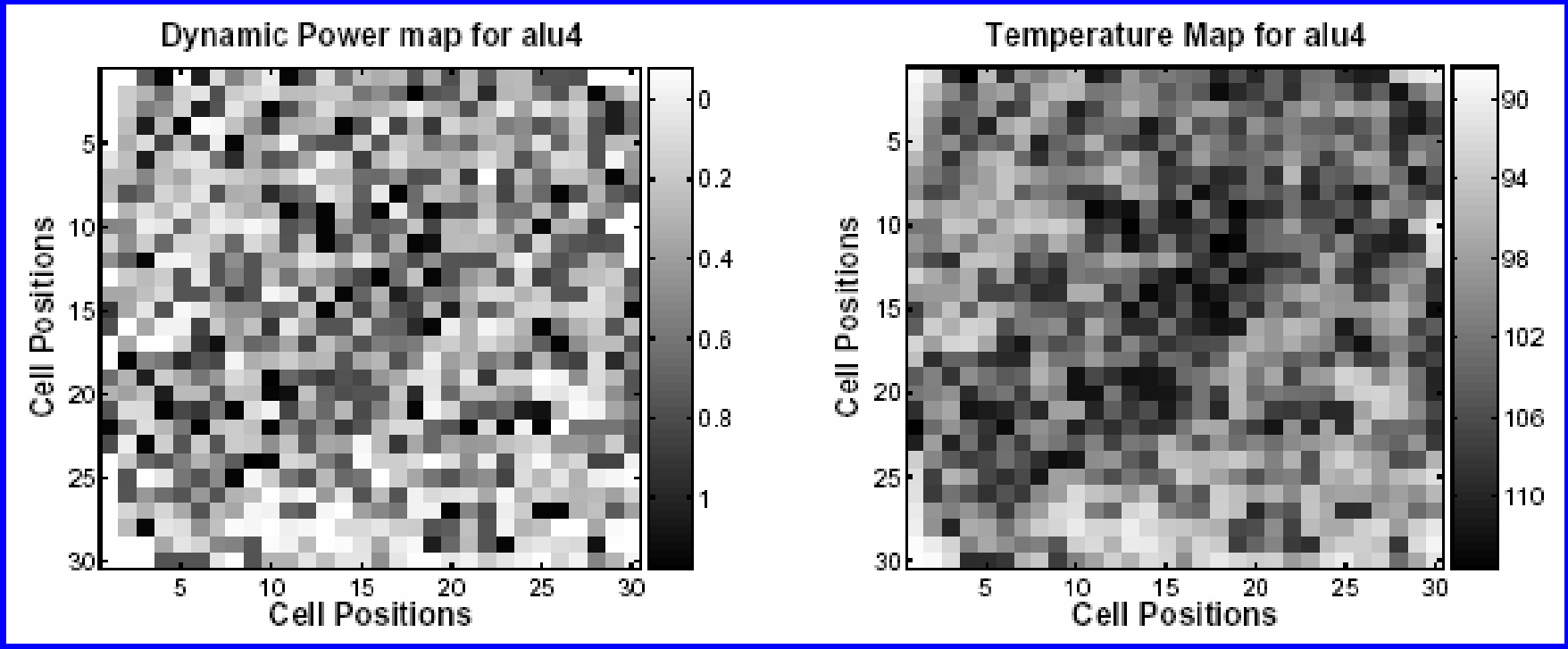
jym, 2005/12/03

Outline

- Introduction
- Power/Performance vs. Temp.
- Compact Thermal Models
- Generating Temperature Map
- What next?

Dynamic Power and Temperature Maps

Benchmark Circuit: alu4

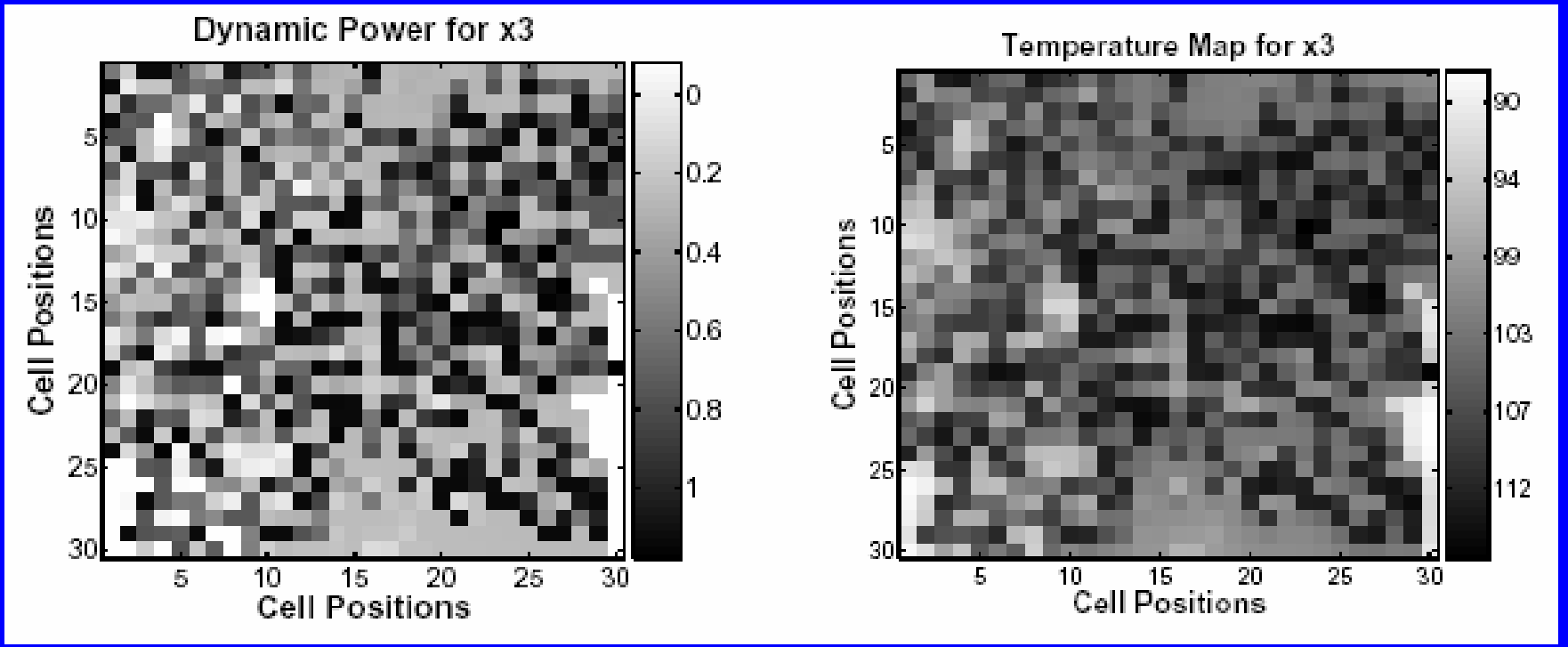


Max. temp. difference is 20 C

Dynamic power map closely corresponds to temperature maps because of inefficient horizontal heat flow

Dynamic Power and Temperature Maps

Benchmark Circuit: x3



So far…

- Standard cell level compact thermal models are generated
- In FinFETs, cell level granularity is imperative because of less efficient horizontal temperature distribution
- Planar grid floorplans are generated
- Cell library needs to be extended for realistic applications

What next?

- Heat flow in higher metallization layers and Si-Bulk need to be included
- In FinFET circuits, rigorous temperature dependent timing analysis is required
- Circuit level techniques such as sizing need to be employed to alleviate thermal issues in FinFETs
- Recursive use of present methodology to develop a multiscale thermal modeling capability from device scale to chip scale

Thank you!
Q & A

Contact author: Aditya Bansal (bansal@ecn.purdue.edu)