Design/Automation Lab
University of Michigan Engineering

Analyzing Electrical Effects of RTA-driven Local Anneal Temperature Variation

Vivek Joshi, Kanak Agarwal*, Dennis Sylvester, David Blaauw

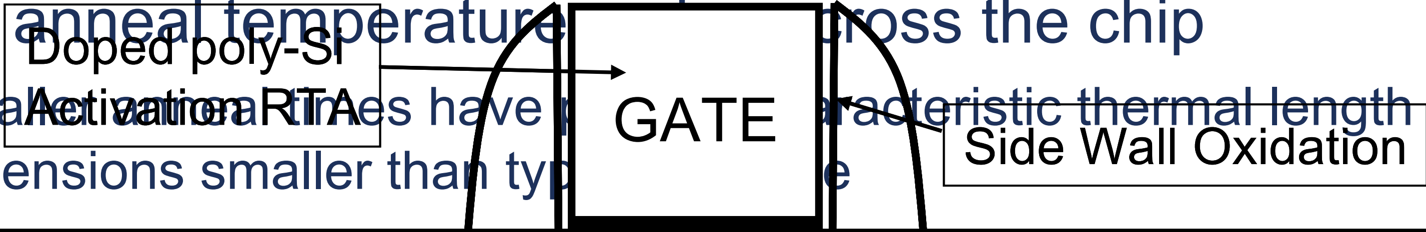
Electrical Engineering & Computer Science
University of Michigan, Ann Arbor
**IBM Research, Austin, TX.*

Outline

- Introduction
- Simulation results for an isolated device
- Modeling local anneal temperature dependent performance/leakage variation
- Chip-level anneal temperature variation analysis
- Results for 45nm and 65nm experimental chips
- Conclusion and Future Work

Introduction

- RTP plays an important role in advanced fabrication process
  - Employed for fabrication steps that require a low thermal budget
- Local anneal temperature … cross the chip
  - Smal… times have … aracteristic thermal length to dimensions smaller than typ… e
  - A… rn de…
- Ann… leak…
  - Hi… ance, le…
- This … aware anal…

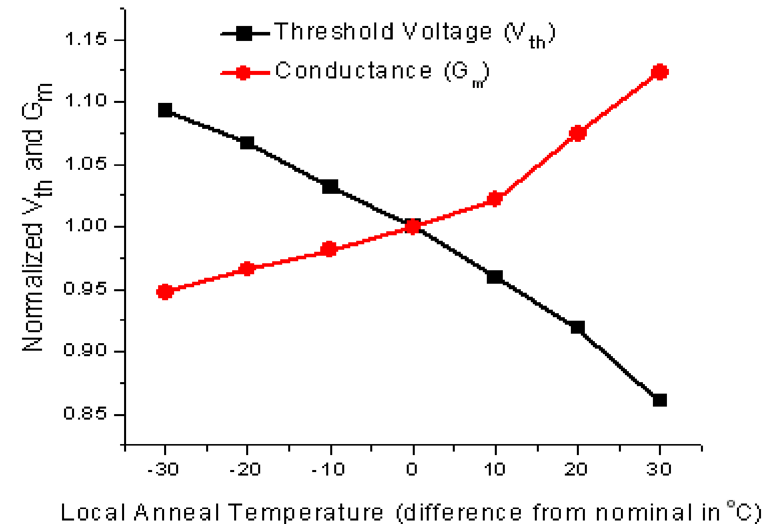


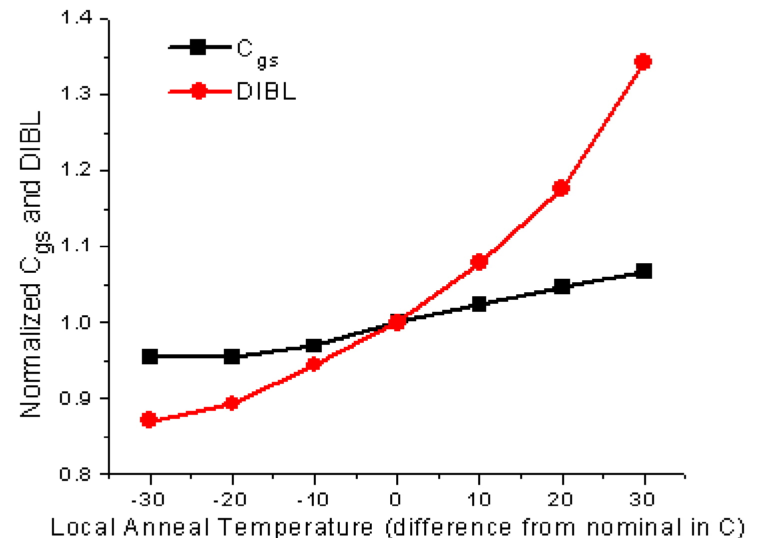
Introduction

- Temperature based variation in V_{th} and R_{ext} correlated
  - Higher anneal temperature reduces V_{th} and R_{ext}
    - Short channel effects, increased gate overlap of S/D
    - Compensation of halo doping
    - Higher dopant activation
  - Correlation leads to pronounced effect on leakage, performance
  - Since characteristic thermal length is large, entire circuit blocks may be systematically faster or slower depending on layout position
- I. Ahasan et. al. showed up to ~20% variation in ring oscillator frequency due to anneal temperature variation in 65nm technology (VLSI Symp. Tech. Digest 2006)
- Limited past work on RTA-driven local anneal temperature variation aware analysis of leakage/performance

Device level variation analysis

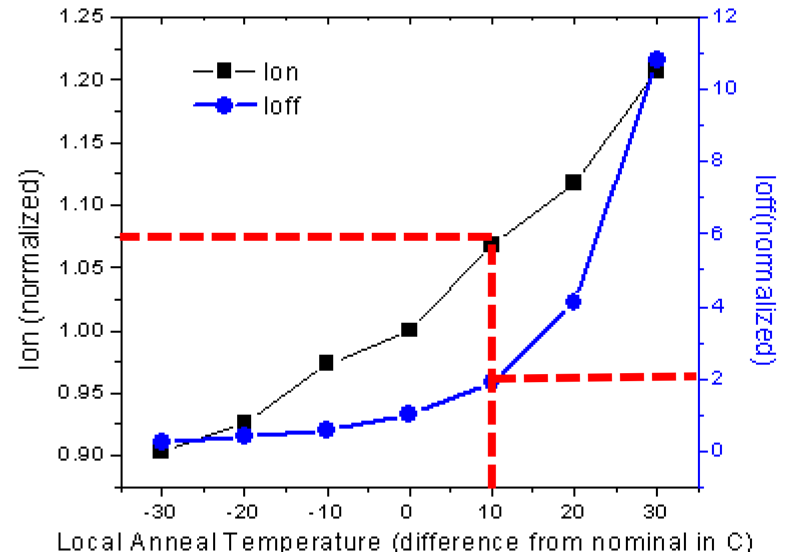
- V_{th} decreases with temperature
  - Short channel effects and compensation of halo doping
- G_m increases with temperature
  - Higher dopant activation and increased gate overlap of S/D
- C_{gs} and DIBL increase with temperature
  - Increased G/D and G/S overlap due to increased dopant diffusion





Device level variation analysis

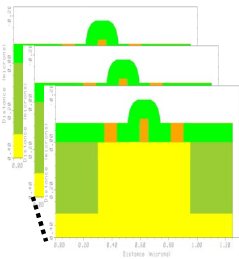
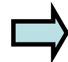
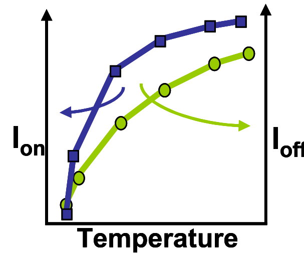
- TCAD simulations show high dependence of on/off currents on temperature
- Local anneal temperature variation of 10°C results in ~7.5% change in Ion and ~2X change in Ioff for 45nm PMOS device

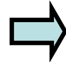


- This temperature dependence can cause significant layout dependent variation in performance/leakage across chip
- These plots also suggest polynomial fits can be used to model temperature dependence of on/off currents

RTA aware performance/leakage analysis

- Model Ion/Ioff temperature dependence using TCAD tools (Tsuprem4)
- Solve for local anneal temperature distribution
  - Set up differential equations describing heat flow
  - Discretize chip area into rectangular grids
  - Approximate partial derivatives as finite differences

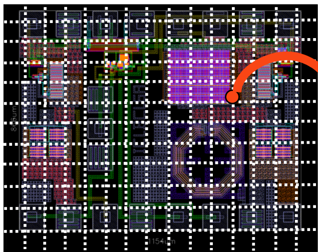






$I_{on,mult}(T) = f_1(T)$

$I_{off,mult}(T) = f_2(T)$

TCAD based modeling of temperature dependence



T(x,y) = T

* Layout Definition File
……………..
……………..
- U1 NAND3 + PLACED (a b) N;
……………..
……………..
T(U1) = T(a,b)

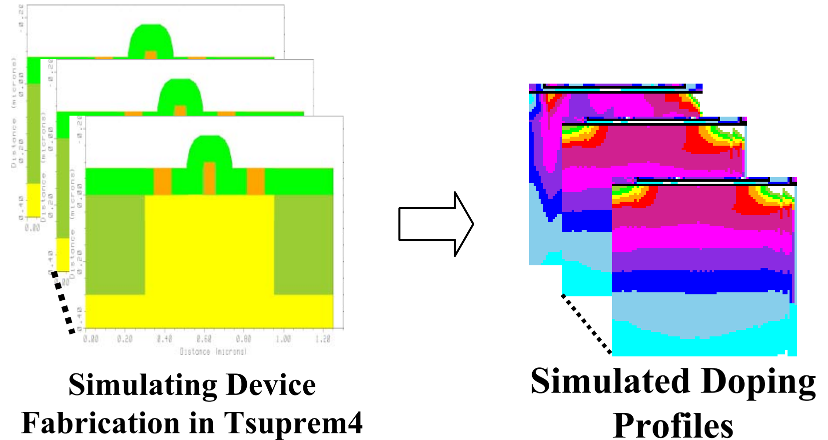
Determination of local anneal temperature based on position

*Timing/Leakage Simulation
……………..
……………..
.
$I_{on}(U1) = I_{on}(NAND3)^* f_1(T(U1))$
$I_{off}(U1) = I_{off}(NAND3)^* f_2(T(U1))$
……………..
……………..

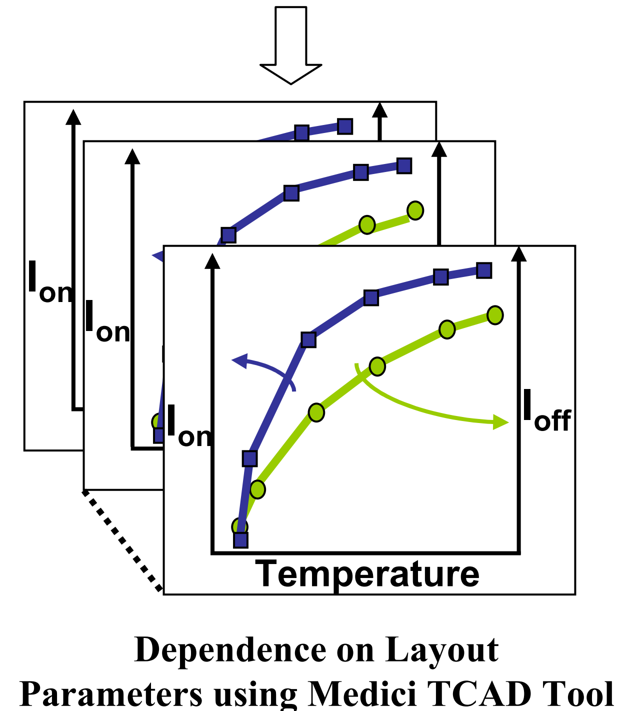
RTA aware timing/ leakage simulation

Simulation flow for TCAD based analysis

- Tsuprem4 for simulating device fabrication
- Simulated device data (doping, stress, etc.) imported into Medici and solve for current values
- Current values were matched to 45nm technology data
- Based on the simulation results, analyze the dependence of leakage/performance on temperature


Simulating Device Fabrication in Tsuprem4

Simulated Doping Profiles



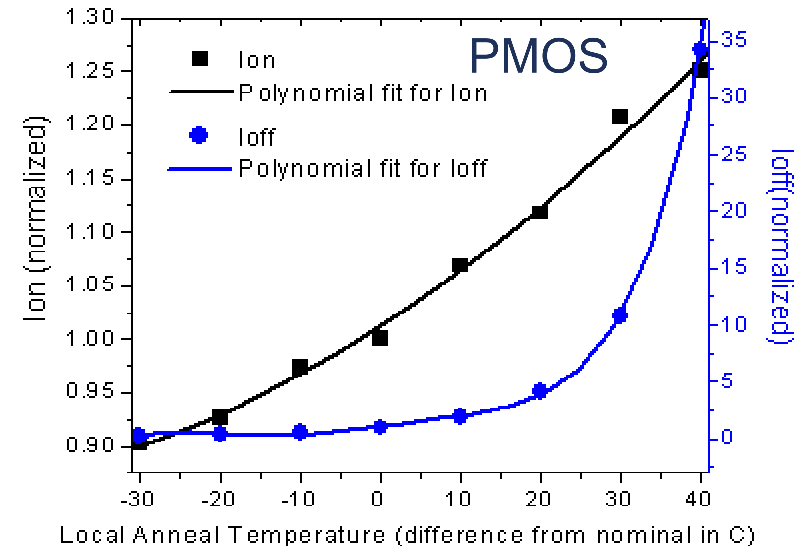
Dependence on Layout Parameters using Medici TCAD Tool

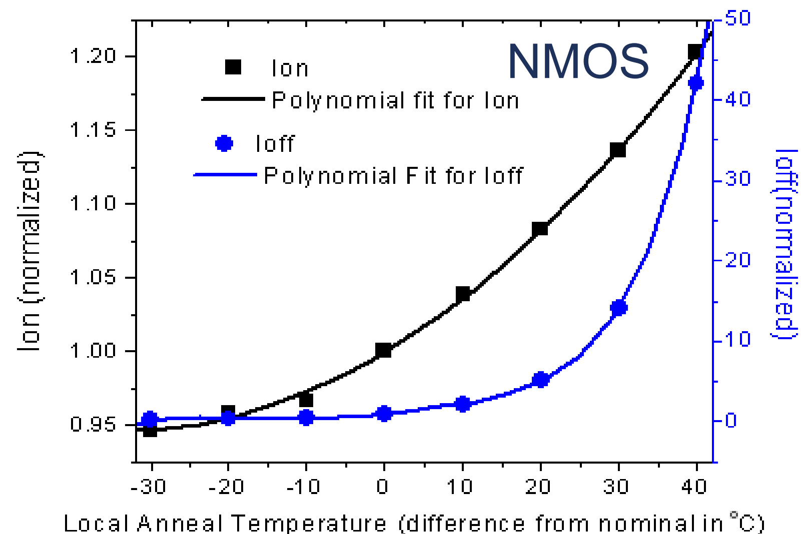
Modeling Ion/Ioff variation with temperature

- Approach 1: Model the effect of anneal temperature variation on basic device properties (V_{th}, G_m, C_{gs}, etc.)
  - Use these models to modify SPICE models, and characterize standard cell library for different values of temperature
  - Given a temperature interpolate between known characterized values
- Approach 2: Model the effect in terms on Ion, Ioff multipliers
  - Characterize standard cell library once for the nominal value of anneal temperature
  - Given a temperature use the modeled function to generate a multiplier
- We use the second approach in this work
  - More accurate for delay/leakage analysis
  - Lower characterization cost

Device level variation analysis

- Use TCAD setup to generate on/off current dependence on local anneal temperature
- Fit polynomial models to normalized values
  - These function represent multipliers for nominal values of on/off currents
- Used MATLAB to obtain accurate polynomial fits





Chip level anneal temperature analysis

- Set up differential equations for heat flow
- Discretize chip area into rectangular grid, and approximate partial derivatives
- Assume radiation is primary mode of heat transfer – neglect conduction, convection
- Assume negligible temperature variation across chip thickness
  - Characteristic time duration of RTP is large enough for temperature to reach steady state
  - Partial derivatives wrt wafer thickness assumed to be zero
- In steady state heat balance can be written as

$$dk\nabla^2 T(x,y) = -(P_{ABS}(x,y) - P_{EMI}(x,y))$$

d = wafer thickness, k = thermal conductivity, P = absorbed/emitted power

Chip level anneal temperature analysis

- Absorbed, emitted power is layout pattern dependent due to layout dependence of optical properties

$$P_{ABS}(x,y) = \alpha(x,y,T)P_{incident} \qquad P_{EMI}(x,y) = \varepsilon(x,y,T)\sigma T^4(x,y)$$

$P_{incident}$ = incident heater power, σ = Stefan-Boltzmann Constant,

α = effective absorptivity, ε = effective emissivity

- Emissivity/absorptivity depend on optical properties of the layout region, temperature and wavelength of incident radiation
- Discritize the chip area into rectangular grids of dimensions Δx and Δy
- Obtain discritized node equations for each node

$$\nabla^2 T(x,y) = \frac{\partial^2}{\partial x}T(x,y) + \frac{\partial^2}{\partial y}T(x,y)$$

Chip level anneal temperature analysis

- Approximate the expression for second spatial derivative in terms of node temperatures under descritization

$$\frac{\partial^2 T_{a,b}}{\partial x} \approx \frac{\frac{T_{a-1,b} - T_{a,b}}{\Delta x} - \frac{T_{a,b} - T_{a+1,b}}{\Delta x}}{\Delta x} \approx \frac{T_{a-1,b} - 2T_{a,b} + T_{a+1,b}}{(\Delta x)^2}$$

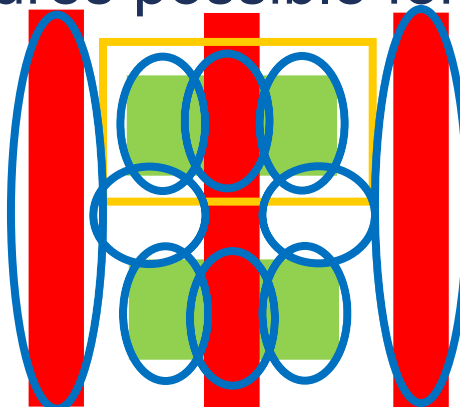
- Finally obtain the following system of non linear equations (one equation for each node)

$$\frac{T_{a-1,b} - 2T_{a,b} + T_{a+1,b}}{(\Delta x)^2} + \frac{T_{a,b-1} - 2T_{a,b} + T_{a,b+1}}{(\Delta y)^2} =$$

$$-(\alpha_{a,b}(T_{a,b})P_{incident} - \varepsilon_{a,b}(T_{a,b})\sigma T^4{}_{a,b})$$

Chip level anneal temperature analysis

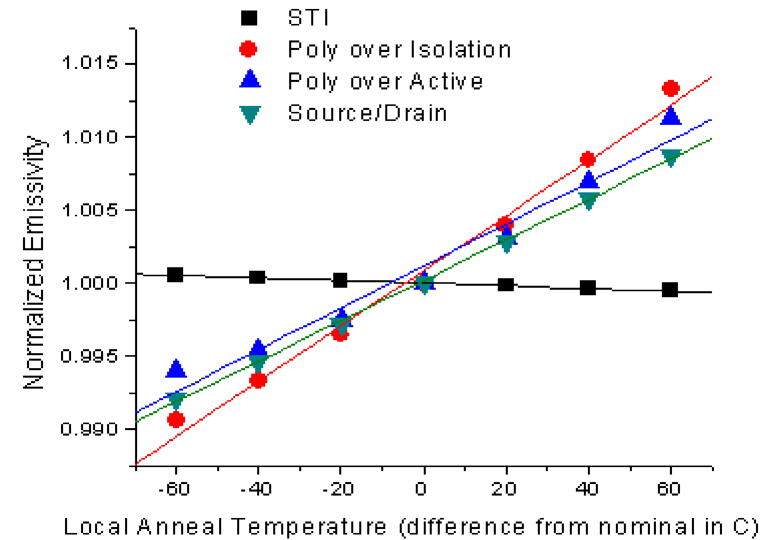
- Five different kinds of layer structures possible for RTA
  - N+ source/drain
  - P+ source/drain
  - Polysilicon over isolation
  - Polysilicon over transistor
  - Shallow trench isolation (STI)

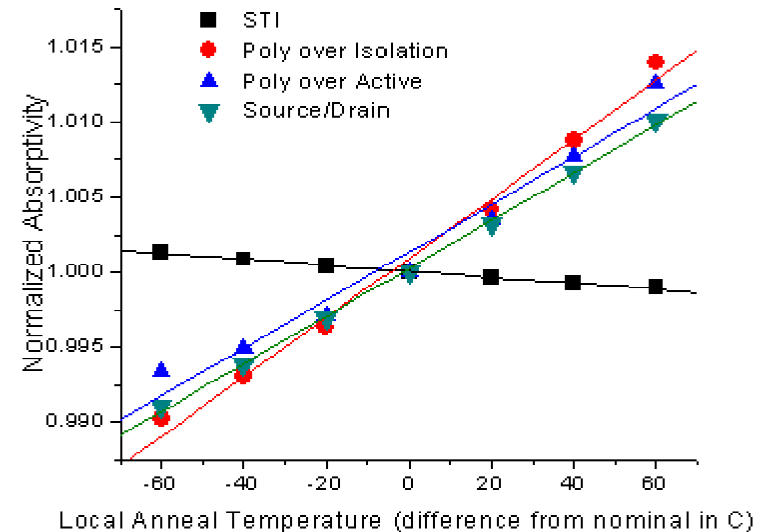


- Interfering optical reflections at layer interfaces cause layer structure dependence of optical properties
- Use RadPro to determine temperature dependent absorptivity and emissivity for each layer structure
- For each grid point take the layout density based average of optical properties for different layer structures

Device level variation analysis

- Optical property for a given grid point is a function of temperature
- Fit linear functions for different layer structures
- Extract density data for each node
- Take density based average to obtain final function for each grid point

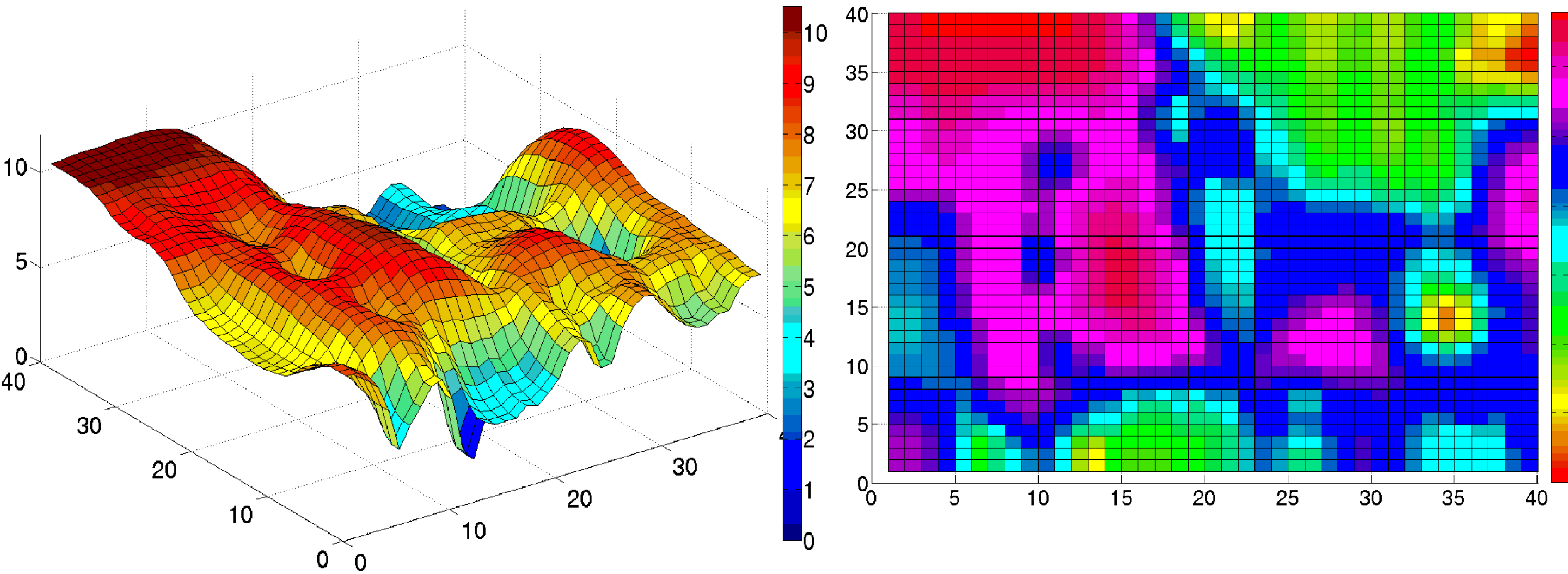




Experimental Setup

- Divide the chip area into 40X40 rectangular grid
  - Use Calibre based script to extract area density for different layer structures
  - Calculate average to determine optical properties for each node
- Use MATLAB for solving the non-linear system of equations to determine 40X40 temperature map
  - Trust region based technique used
- TCAD setup calibrated to the 45nm technology used for experiments
- TCAD based on/off current multipliers used to determine the electrical impact
  - Current numbers reported are average of NMOS and PMOS currents

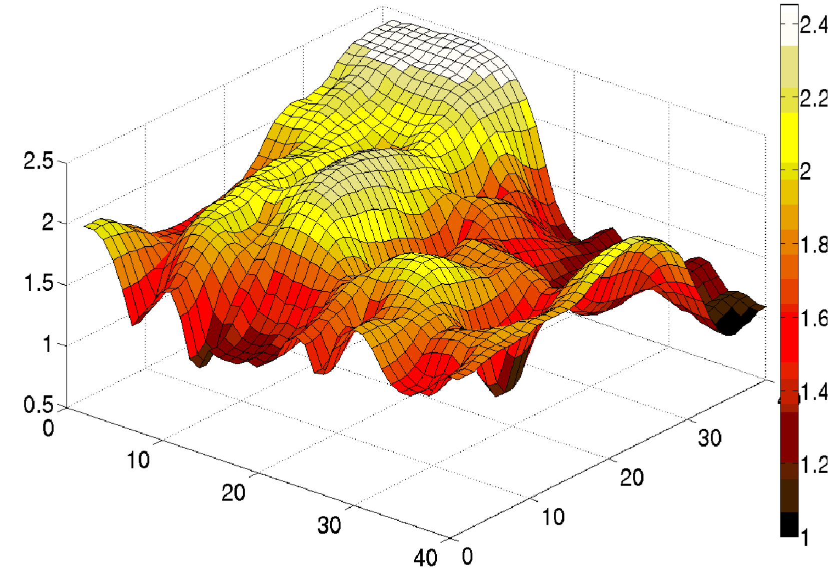
Results – 45nm test chip



- Local anneal temperature variation of up to ~10.5°C
- On current variation of up to 6.8%
- Variation in inverter delay ~7.3%

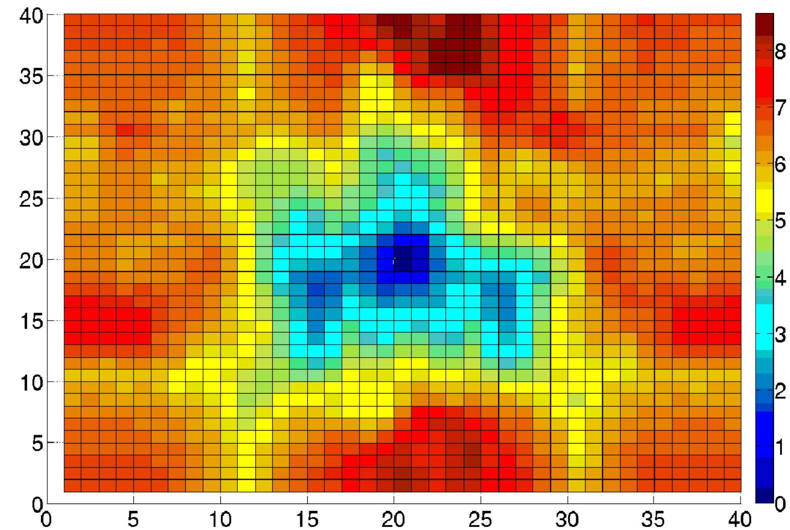
Results – 45nm test chip

- Device leakage at the highest temperature node is ~2.45X the lowest leakage point
- Temperature/current distribution shows some correlation with STI density distribution
  - STI has lowest reflectivity
    - Lower reflectivity translates into high absorptivity
    - High STI density means high temperature
  - Long range effects related to characteristic thermal length make the correlation less exact
  - Need for a more exact solution
- Appreciable variation in currents due to local anneal temperature variation



Results – 65nm test chip

- 40X40 rectangular grid used
- Temperature map shows variation of up to 8.5^{o}C
  - Slightly smaller than 45nm test case
  - It can cause reasonable impact on leakage/performance



- Temperature distribution shows some correlation with STI density distribution
- Establishes that there is a need for RTA variation aware analysis for timing/leakage for accurate simulation

Conclusions and Future Work

- We propose a new local anneal temperature variation aware performance and leakage analysis framework
  - Embodies transistor level models for anneal temperature sensitivity
  - Solve for chip level anneal temperature distribution by dividing the wafer surface into rectangular grids
  - Employ TCAD based device level models for drive current (Ion) and leakage current (Ioff) dependence on anneal temperature variation
- Temperature variations of up to 10.5^{o}C for 45nm test chip
  - 6.8% variation in on current, and 2.45X variation in leakage
  - Establishes the need for such a framework
- Ongoing and future work
  - Use this framework to drive layout optimization to reduce RTA-induced variation

Thank You