Interconnect Optimization for Deep-Submicron and Giga-Hertz ICs

Lei He

he@ece.wisc.edu

http://eda.ece.wisc.edu

Outline

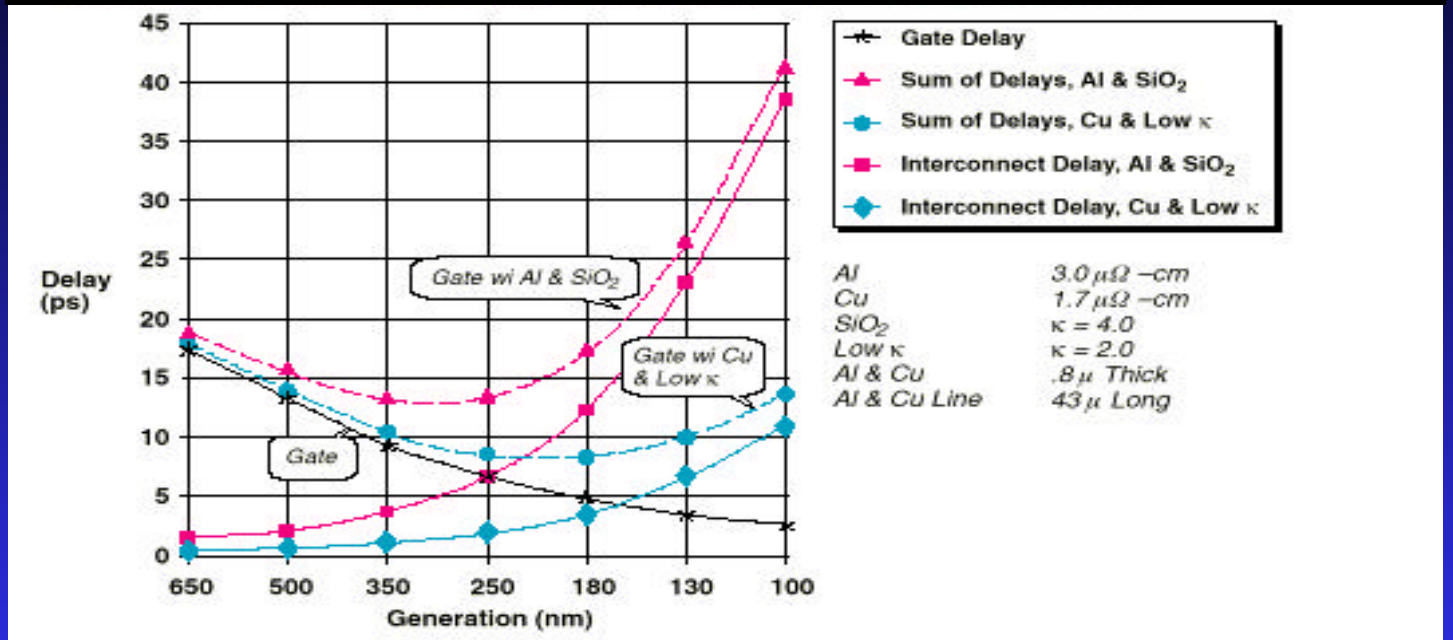
- Background and overview
- LR-based STIS optimization
  - LR -- local refinement
  - STIS -- simultaneous transistor and interconnect sizing
- Conclusions and future works

Upcoming Design Challenges

- Microprocessors for server computers
  - 1998 -- 0.25um, 7.5M FETs, 450MHz
  - 2001 -- 0.18um, ~100M FETs, >1GHz
    - close to tape-out
  - 2005 -- 0.10um, ~200M FETS, ~3.5GHz
    - launch design in 2003
    - begin developing design tools in 2001
    - start research right now

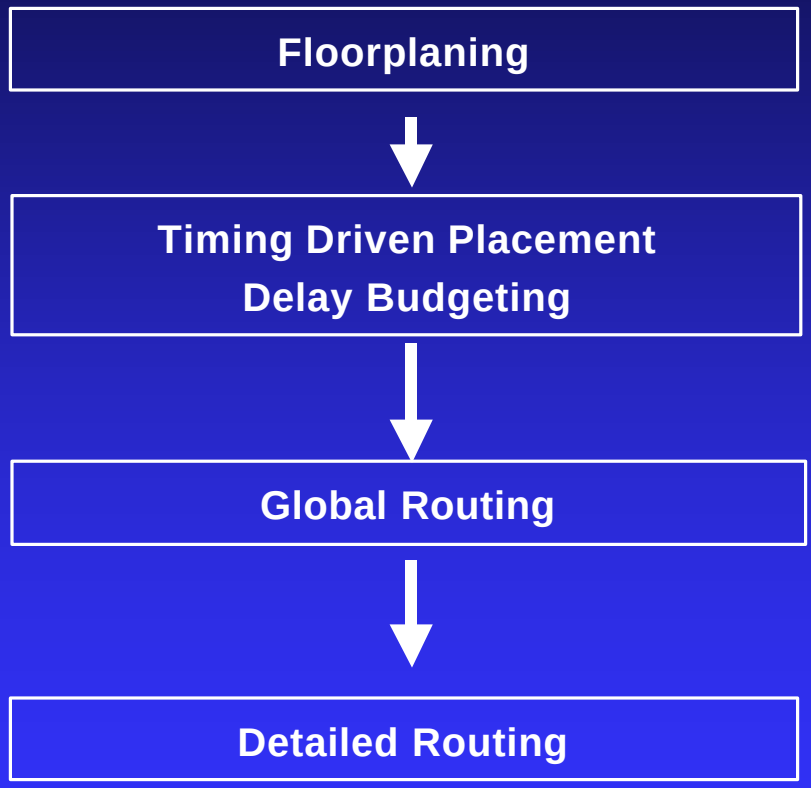
We are moving faster than Moore's Law

Critical Issue: Interconnect Delay

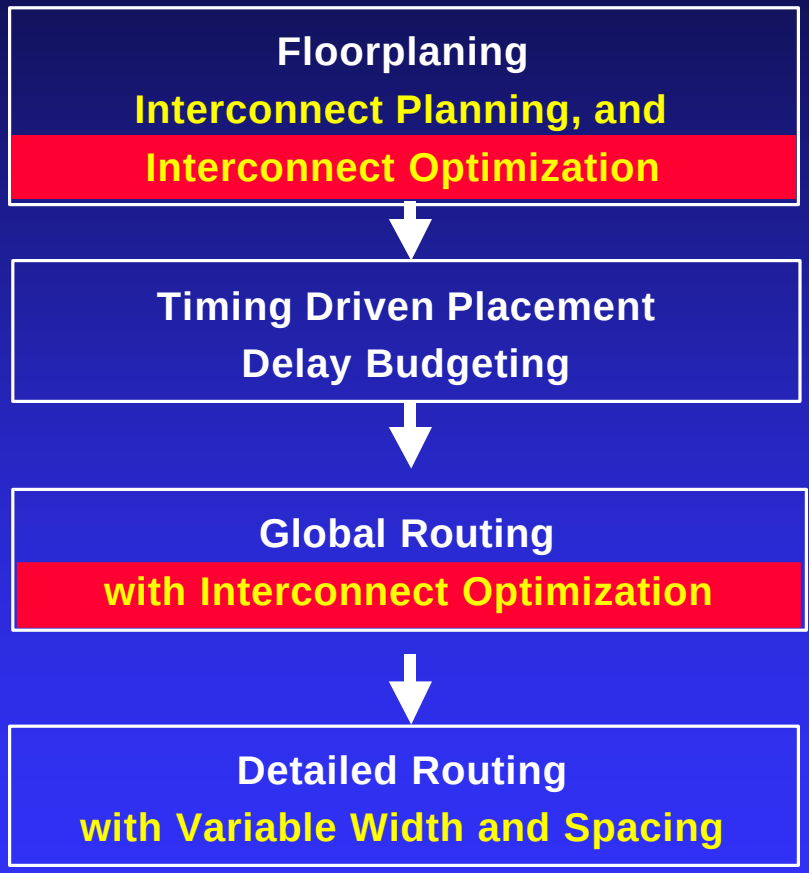


- Starting from 0.25um generation, circuit delay is dominated by interconnect delay
- Efforts to control interconnect delay
  - Processing technology: Cu and low K dielectric
  - Design technology: interconnect-centric design

Layout Design:
Device-Centric versus Interconnect-Centric



Device-Centric



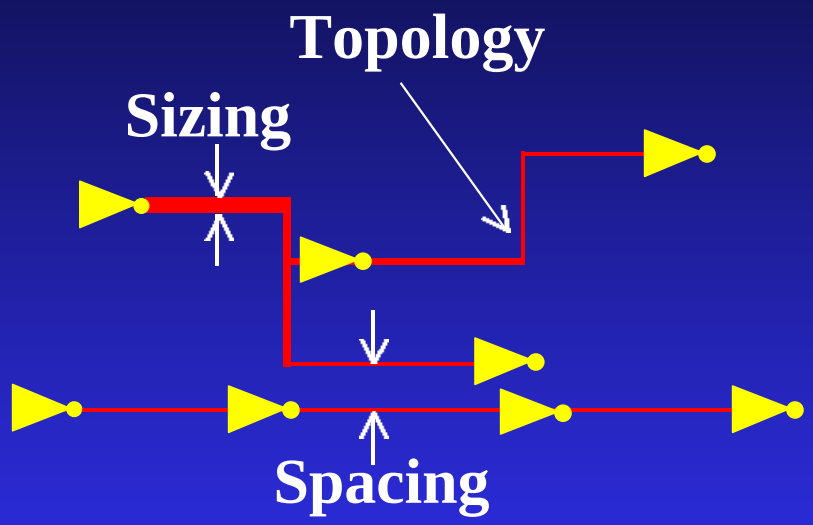
Interconnect-Centric

Interconnect Optimization

Device locations and constraints:

- **Delay**
- **Power**
- **Signal integrity**
- **Skew**

...



- **Other critical optimizations: buffer insertion, simultaneous device and interconnect sizing …**
- **Automatic solutions guided by accurate interconnect and device models**

UCLA TRIO Package

- Integrated system for interconnect design
  - http://cadlab.cs.ucla.edu/~trio
- Efficient polynomial-time optimal/near-optimal algorithms
  - Interconnect topology optimization
  - Optimal buffer insertion
  - Optimal wire sizing
  - Wire sizing and spacing considering Cx
  - Simultaneous device and interconnect sizing
  - Simultaneous topology generation with buffer insertion and wiresizing
- Accurate interconnect models
  - 2 -1/2 D capacitance model
  - 2 -1/2 D inductance model
  - Elmore delay and higher-order delay models
- Improve interconnect performance by up to 7x !
  - Used in industry, e.g., Intel

Contributions to UCLA TRIO Package

- **Integrated system for interconnect design**
  - **http://cadlab.cs.ucla.edu/~trio**
- **Efficient polynomial-time optimal/near-optimal algorithms**
  - **Interconnect topology optimization**
  - **Optimal buffer insertion**
  - **Optimal wire sizing [ICCAD'95, TODAES'96]**
  - **Wire sizing and spacing considering Cx [ICCAD'97, TCAD'99]**
  - **Simultaneous device and interconnect sizing [ICCAD'96, ISPD'98,TCAD'99]**
  - **Simultaneous topology generation with buffer insertion and wiresizing**
- **Accurate interconnect models**
  - **2 -1/2 D capacitance model [DAC'97] (with Cadence)**
  - **2 -1/2 D inductance model [CICC'99] (with HP Labs)**
  - **Elmore delay and higher-order delay models**
- **Improve interconnect performance by up to 7x !**
  - **Used in industry, e.g., Intel**

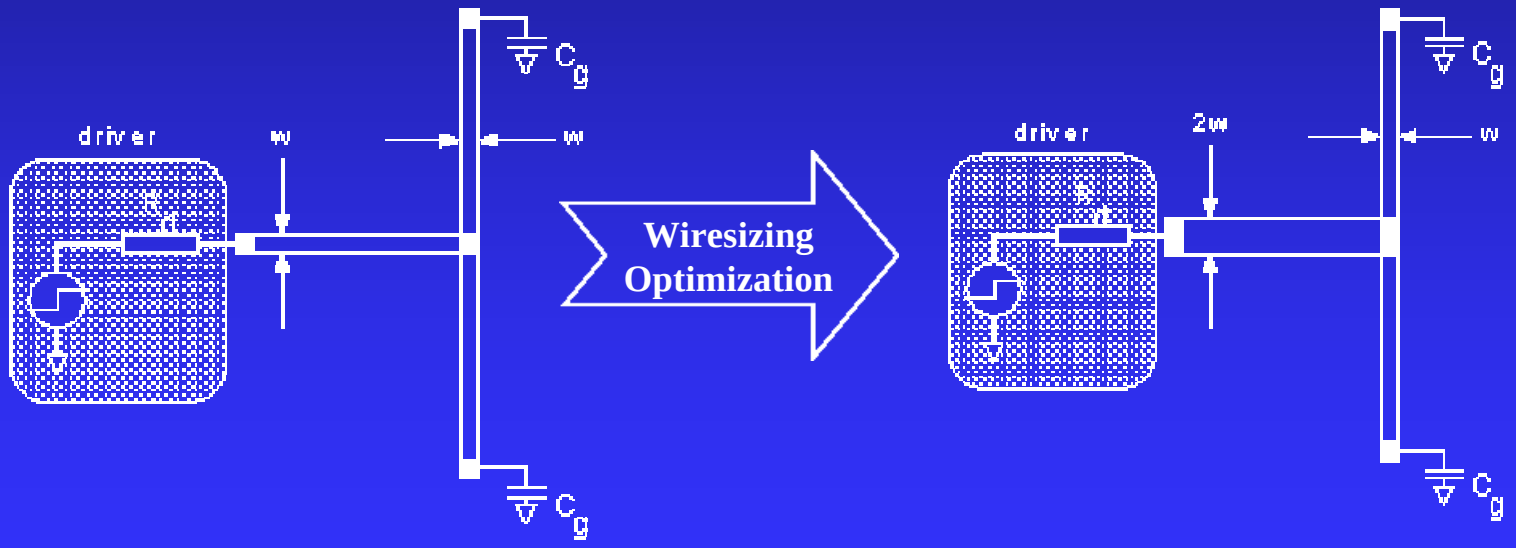
Outline

- Background and overview
- LR-based STIS optimization
  - Motivation for LR-based optimization
- Conclusions and future works

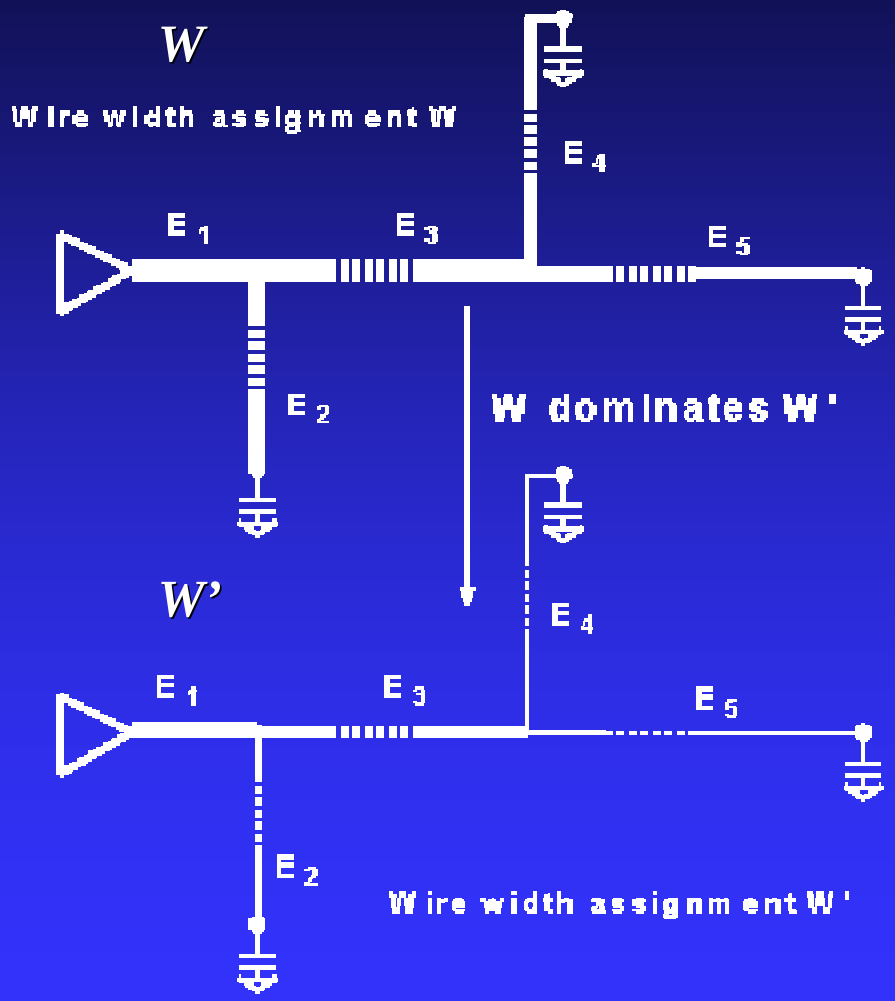
Discrete Wiresizing Optimization

[Cong-Leung, ICCAD'93]

- **Given:** A set of possible wire widths { W_1, W_2, …, W_r }
- **Find:** An optimal wire width assignment to minimize weighted sum of sink delays



Dominance Relation and Local Refinement



- **Dominance Relation**
  - For all E_j, $w(E_j)^3\ w'(E_j)$

    ß

    W dominates *W'* *(i.e.,* $W^3\ W'$*)*
- **Local refinement (LR)**
  - LR for E_1 to find an optimal width for E_1, assuming widths for other wires are fixed with respect to current width assignment
  - Single-variable optimization can be solved efficiently

Dominance Property for Discrete Wiresizing

[Cong-Leung, ICCAD'93]

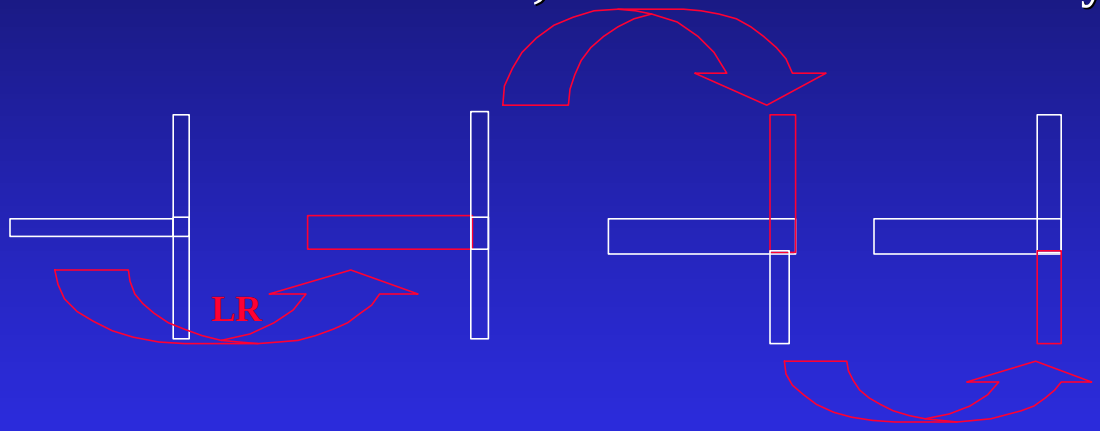
- **If solution *W* dominates optimal solution *W**
  W' = local refinement of *W*
  Then, W' dominates *W**

- **If solution *W* is dominated by optimal solution *W**
  W' = local refinement of *W*
  Then, W' is dominated by *W**

A highly efficient algorithm to compute tight lower and upper bounds of optimal solution

Bound Computation based on Dominance Property

- Lower bound computed starting with minimum widths
  - LR operations on all wires constitute a pass of bound computation
  - LR operations can be in an arbitrary order
  - New solution is wider, but is still dominated by the optimal solution



- Upper bound is computed similarly, but beginning with maximum widths
- We alternate lower and upper bound computations
  - Total number of passes is linearly bounded by size of solution space
- Optimal solution is often achieved in experiments

Other Problems Solved by LR operation

Why does LR operation work?

- Multi-source discrete wiresizing [Cong-He, ICCAD'95]
  - Bundled-LR is proposed to speed up LR by a factor of 100x
- Continuous wiresizing [Chen-Wong, ISCAS'96]
  - Linear convergence is proved [Chu-Wong, TCAD'99]
- Simultaneous buffer and wire sizing [Chen-Chang-Wong, DAC'96]
  - Lagrangian relaxation is proposed to minimize max delay
    - via a sequence of weighted delay minimizations
  - Extended to general gates and multiple nets [Chu-Chen-Wong, ICCAD'98]

Outline

- Background and overview
- LR-based STIS optimization
  - Motivation of LR-based optimization
  - Simple CH-program and application to STIS problem
- Conclusions and future works

Simple CH-function [Cong-He, ICCAD'96, TCAD'99]

- $f(X) = \sum_{p=0}^{m} \sum_{q=0}^{m} \sum_{i=1}^{n} \sum_{j=1,\, j \neq i}^{n} \left(\frac{a_{pi}}{x_i^p}\right) \cdot \left(b_{qj} \cdot x_j^q\right)$

  is a simple CH-function

  - **Variables x_i and x_j are positive, either continuous or discrete**
  - **Coefficients a_{pi} and b_{qj} are positive constants**

- **Examples:**
  - $f = ax^2 + b/x + cy^3, g = \sum (a_{ij} \cdot x_i/x_j)$
- **It includes the objective functions for a number of works**
  - **Discrete or continuous wire sizing [Cong-Leung, ICCAD'93][Cong-He, ICCAD'95][Chen-Wong,ISCAS'96]**
  - **Simultaneous device and wire sizing [Cong-Koh, ICCAD'94][Chen-Chang-Wong, DAC'96][Cong-Koh-Leung, ILPED'96][Chu-Chen-Wong, ICCAD'98]**

Dominance Property for Simple CH-Program

- Optimization problem to minimize a simple CH-function is a simple CH-program.
- The dominance property holds for simple CH-program w.r.t. the LR operation.
  - If *X* dominates optimal solution *X**
    X' = local refinement of *X*
    Then, *X'* dominates *X**
  - If *X* is dominated by optimal solution *X**
    X'=local refinement of *X*
    Then, *X'* is dominated by *X**
- LR operation can be used for all simple CH-programs

Simple CH-function [Cong-He, ICCAD'96, TCAD'99]

- $f(X) = \sum_{p=0}^{m} \sum_{q=0}^{m} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left(\frac{a_{pi}}{x_i^p}\right) \cdot \left(b_{qj} \cdot x_j^q\right)$

  is a simple CH-function

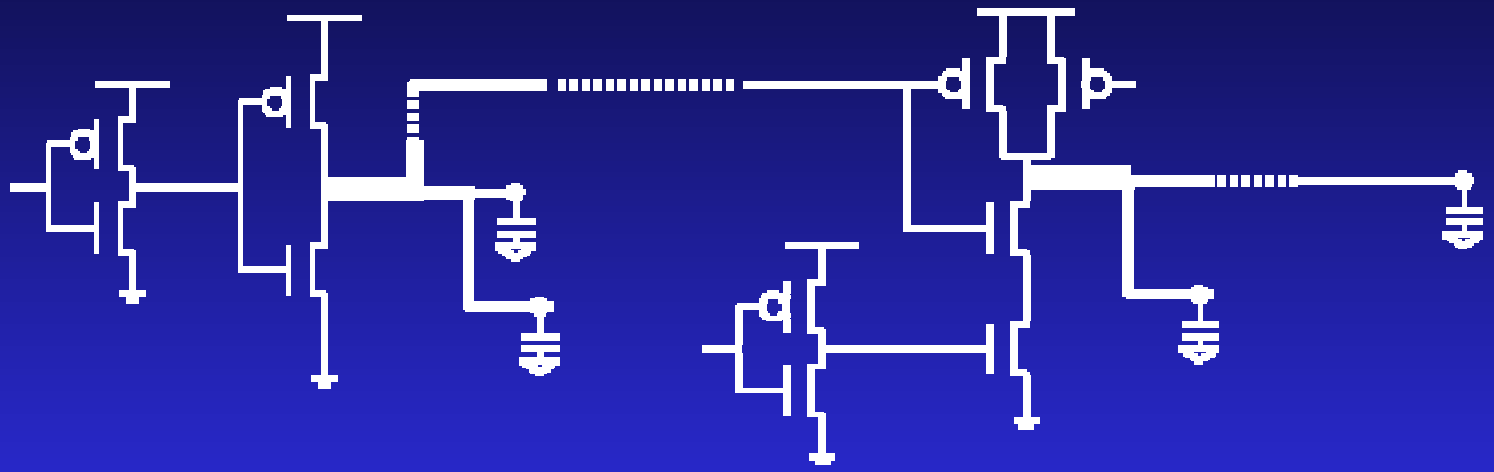
  - **Variables x_i and x_j are positive**
  - **Coefficients a_{pi} and b_{qj} are positive constants**

- **Examples:**
  - $f = ax^2 + b/x + cy^3, g = \sum (a_{ij} \cdot x_i/x_j)$

Unified and efficient solution

- **It includes the objective functions for a number of works**
  - **Discrete or continuous wire sizing [Cong-Leung, ICCAD'93][Cong-He, ICCAD'95][Chen-Wong,ISCAS'96]**
  - **Simultaneous device and wire sizing [Cong-Koh, ICCAD'94][Chen-Chang-Wong, DAC'96][Cong-Koh-Leung, ILPED'96][Chu-Chen-Wong, ICCAD'98]**

General Formulation: STIS
Simultaneous Transistor and Interconnect Sizing



- **Given:** Circuit netlist and initial layout design
- **Determine:** Discrete sizes for devices/wires
- **Minimize:** a Delay + b Power + g Area

⇨ It is the first publication to consider simultaneous device and wire sizing **for complex gates and multiple paths**

STIS Objective for Delay Minimization

- $$t(X) = \sum_{i,j} F(i,j) \bullet \frac{R_0(i)}{x_i} \bullet C_0(j) \bullet x_j + \sum_{i,j} F(i,j) \bullet \frac{R_0(i)}{x_i} \bullet C_1(j)$$
  $$+ \sum_{i} G(i) \bullet \frac{R_0(i)}{x_i} + \sum_{i} H(i) \bullet \frac{R_0(i)}{x_i} \bullet C_1(i)$$
  - R_0 : **unit-width resistance**
  - C_0 : **unit-width area capacitance**
  - C_1 : **effective-fringing capacitance**
  - $X = \{x_1, x_2, ..., x_n\}$: **discrete widths and variables for optimization**

Res = R_0 /x
Cap = C_0 * x + (C_f + C_x)
= C_0 * x + C_1

- **It is a simple CH-function under simple model assuming *R_0, C_0* and *C_1* are constants**
- **STIS can be solved by computing lower and upper bounds via LR operations**
  - **Identical lower and upper bounds often achieved**

SPICE-Delay reduction of LR-Based STIS

- **STIS optimization versus manual optimization for clock net [Chien-et al.,ISCC'94]:**
  - **1.2um process, 41518.2 um wire, 154 inverters**
- **Two formulations for LR-based optimization**
  - **sgws simultaneous gate and wire sizing**
  - **stis simultaneous transistor and interconnect sizing**

	manual	*sgws*	*stis*
max delay (ns)	4.6324	4.34 (-6.2%)	3.96 (-14.4)
power(mW)	60.85	46.1 (-24.3%)	46.3 (-24.2%)
clock skew (ps)	470	130 (-3.6x)	40 (-11.7x)

- **Runtime (wire segmenting: 10um)**
  - **LR-based** ***sgws*** **1.18s,** ***stis*** **0.88s**
  - **HSPICE simulation** **~2100s in total**

STIS Objective for Delay Minimization

- $$t(X) = \sum_{i,j} F(i,j) \bullet \frac{R_0(i)}{x_i} \bullet C_0(j) \bullet x_j + \sum_{i,j} F(i,j) \bullet \frac{R_0(i)}{x_i} \bullet C_1(j)$$
  $$+ \sum_{i} G(i) \bullet \frac{R_0(i)}{x_i} + \sum_{i} H(i) \bullet \frac{R_0(i)}{x_i} \bullet C_1(i)$$
  - R_0 : **unit-width resistance**
  - C_0 : **unit-width area capacitance**
  - C_1 : **fringing capacitance**
  - $X = \{x_1, x_2, ..., x_n\}$: **discrete widths and variables for optimization**

Over-simplified for DSM (Deep Submicron) designs

- **It is a simple CH-function under simple model assuming R_0 ,C_0 and C_1 are constants**

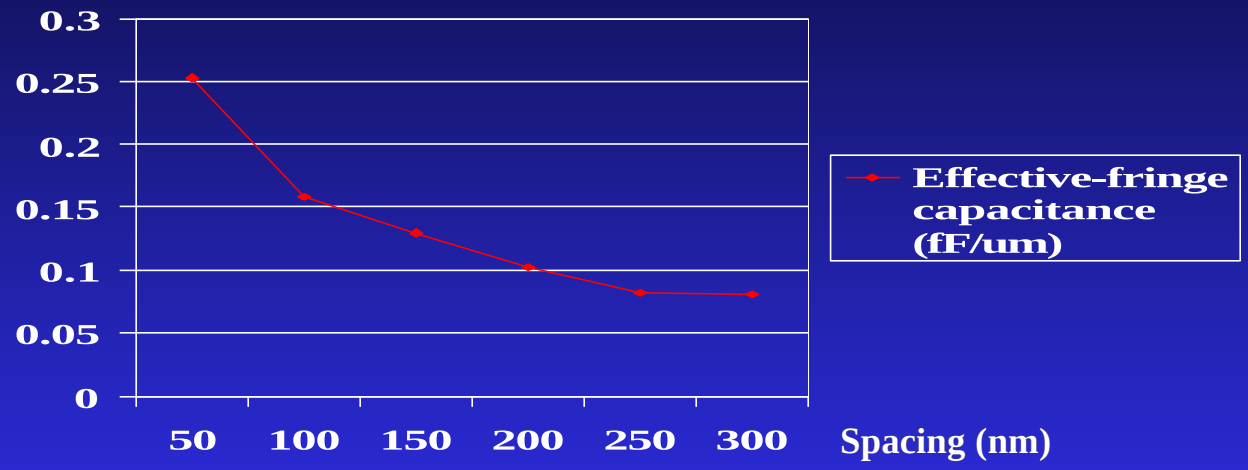
R_0 is far away from a Constant!

effective-resistance R_0 for unit-width n-transistor							
size = 100x				size = 400x			
$c_l \setminus t_t$	0.05ns	0.10ns	0.20ns	$c_l \setminus t_t$	0.05ns	0.10ns	0.20ns
0.225pf	12200	12270	19180	0.501pf	12200	15550	19150
0.425pf	8135	9719	12500	0.901pf	11560	13360	17440
0.825pf	8124	8665	10250	1.701pf	8463	9688	12470

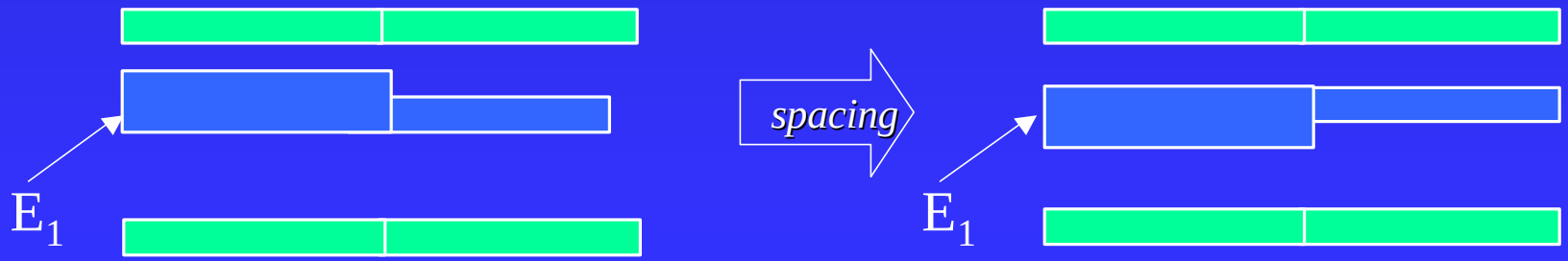
- R_0 depends on size, input slope t_t and output load c_l
  - May differ by a factor of 2
- Using more accurate model like the table-based device model has the potential of further delay reduction.
  - But easy to be trapped at local optimum, and to be even worse than using simple model [Fishburn-Dunlop, ICCAD'85]

Neither C_0 nor C_1 is a Constant

- Both depend on wire width and spacing
  - Especially $C_1 = C_f + C_x$ is highly sensitive to spacing



- C_x accounts for >50% capacitance in DSM
  - Proper spacing may lead to extra delay reduction
  - But no existing algorithm for optimal spacing



STIS-DSM Problem to Consider DSM Effects

Non-constant R_0, C_0 and C_1

- **STIS-DSM problem**
  - **Find:** device sizing, and wire sizing and spacing solution optimal w.r.t. accurate device model and multiple nets
- **Easier but less appealing formulation: single-net STIS-DSM**
  - **Find:** device sizing, and wire sizing and spacing solution optimal w.r.t. accurate device model and a single-net
  - **Assume:** its neighboring wires are fixed

Outline

- Background and overview
- LR-based STIS optimization
  - Motivation: LR-based wire sizing
  - Simple CH-program and application to STIS problem
  - Bounded CH-program and application to STIS-DSM problem
- Conclusions and future works

Go beyond Simple CH-function

$$f(X) = \sum_{p=0}^{m} \sum_{q=0}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{a_{pi}(X)}{x_i^p} \right) \cdot \left(b_{qj}(X) \cdot x_j^q \right)$$

- **It is a simple CH-function if**
  - a_{pi} **and** b_{qj} **are positive constants**
- **It is a bounded CH-function if**
  - a_{pi} **and** b_{qj} **are arbitrary functions of** X
  - a_{pi} **and** b_{qj} **are positive and bounded**
    - $a_{pi}^L \le a_{pi}(X) \le a_{pi}^U$ **and** $b_{qj}^L \le b_{qj}(X) \le b_{qj}^U$
- **Examples:**
  - $f(x_1, x_2) = \frac{1}{\ln x_1} \cdot x_1 + \frac{x_2}{x_1}, x_1 > e$
  - **Objective function for STIS-DSM problem**

Extended-LR Operation

- **Extended-LR (ELR) operation is a relaxed LR operation**
  - **Replace a_{pi} and b_{qj} by its lower or upper bound during LR operation.**
  - **Make sure that the resulting lower or upper bound is always correct**
    - **but might be conservative.**

- **Example:** $f(x) = a(x) \cdot x + b(x)/x$
  - **ELR for a lower bound is** $x_{ELR}^{L} = \sqrt{b^{L}/a^{U}}$
  - **ELR for an upper bound is** $x_{ELR}^{U} = \sqrt{b^{U}/a^{L}}$

?

$$x_{ELR}^{L} \le x^{*} \le x_{ELR}^{U}$$

General Dominance Property

- **Theorem ([Cong-He, ISPD'98, TCAD'99])**
  - **Dominance property holds for bounded CH-program with respect to ELR operation**

General Dominance Property

- **Theorem ([Cong-He, ISPD'98, TCAD'99])**
  - **Dominance property holds for bounded CH-program with respect to ELR operation**

- **To minimize** $f(X) = \sum_{p=0}^{m} \sum_{q=0}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{a_{pi}(X)}{x_i^p} \right) \cdot \left(b_{qj}(X) \cdot x_j^q \right)$
  - **If *X* dominates optimal solution *X**
    X' = Extended-LR of *X*
    Then, *X'* dominates *X**
  - **If *X* is dominated by optimal solution *X***
    X'= Extended-LR of *X*
    Then, *X'* is dominated by *X**

Solution to STIS-DSM Problem

- STIS-DSM can be solved as a bounded CH-program
  - Lower bound computed by ELR starting with minimum sizes
  - Upper bound computed by ELR starting with maximum sizes
  - Lower- and upper-bound computations are alternated to shrink solution space
  - Up-to-date lower and upper bounds of R_0, C_0 and C_1 are used
    - Uncertainty of R_0, C_0 and C_1 is reduced when the solution space is shrunk
- There exists an optimal solution to the STIS-DSM problem between final lower and upper bounds
  - How large is the gap?

Gaps between Lower and Upper Bounds

- Two nets under 0.18um technology: *DCLK* and *2cm line*
  - STIS-DSM uses table-based device model and ELR operation
  - STIS uses simple device model and LR operation
- We compare average lower-bound width / average gap

	Transistors		Wires	
DCLK	STIS	STIS-DSM	STIS	STIS-DSM
sgws	5.39/0.07	13.0/1.91	2.50/0.003	2.78/0.025
stis	17.2/1.53	21.6/2.36	2.69/0.017	2.82/0.030
2cm line	STIS	STIS-DSM	STIS	STIS-DSM
sgws	108/0.108	112/0.0	4.98/0.004	4.99/0.106
stis	126/0.97	125/1.98	5.05/0.032	5.11/0.091

- Gap is almost negligible
  - About 1% of lower-bound width in most cases

Delay Reduction by Accurate Device Model

- STIS-DSM versus STIS
  - STIS-DSM uses table-based device model and ELR operation
  - STIS uses simple device model and LR operation

DCLK	STIS	STIS-DSM
sgws	1.16 (0.0%)	1.08 (-6.8%)
stis	1.13 (0.0%)	0.96 (-15.1%)

2cm line	STIS	STIS-DSM
sgws	0.82 (0.0%)	0.81 (-0.4%)
stis	0.75 (0.0%)	0.69 (-7.6%)

- STIS-DSM achieves up to 15% extra reduction
- Runtime is still impressive
  - Total optimization time ~10 seconds

Delay Reduction by Wire Spacing

- Multi-net STIS-DSM versus single-net STIS-DSM
  - Test case:
    - 16-bit bus
    - each bit is 10mm-long with 500um per segment

pitch-spacing	delay		runtime
	single-net	multi-net	multi-net
1.10um	1.31	0.79 (-39%)	2.0s
1.65um	0.72	0.52 (-27%)	2.4s
2.20um	0.46	0.42 (-8.7%)	2.3s
2.75um	0.38	0.36 (-5.3%)	4.9s
3.30um	0.35	0.32 (-8.6%)	7.7s

- Multi-net STIS-DSM achieves up to 39% delay reduction
  - Single-net STIS-DSM has a significant delay reduction if we compare it with previous wire sizing formulations

Conclusions

- Interconnect-centric design is the key to DSM and GHz IC designs
- Interconnect optimization is able to effectively control interconnect delay
  - Problem formulations should consider DSM effects
    - e.g., LR-based optimization for STIS-DSM problem

Valid for general problems

- More is needed to close the loop of interconnect-centric design
  - Interconnect planning
  - Interconnect optimization for inductance and noise
  - Interconnect verification, especially for pattern-dependent noise and delay
  - ... ...