Subthreshold Logical Effort: A Systematic Framework for Optimal Subthreshold Device Sizing

John Keane, Hanyong Eom, Tae-Hyoung Kim, Sachin Sapatnekar, Chris Kim

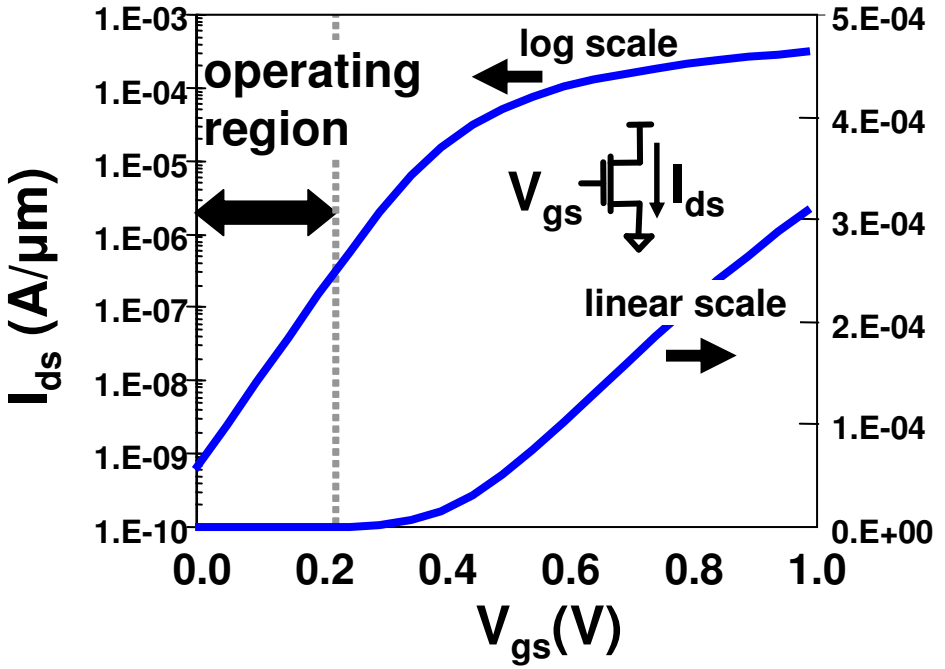
University of Minnesota
Department of Electrical and Computer Engineering

jkeane@ece.umn.edu

Presentation Agenda

- Subthreshold design
- Conventional logical effort
- Proposed subthreshold logical effort formulation
- Simulation results
- Conclusion

Subthreshold Operation



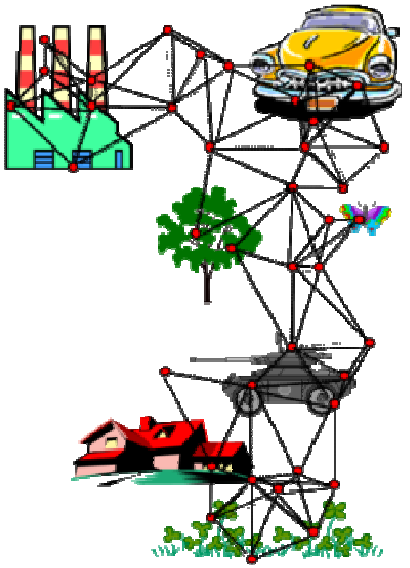
- **Main Benefit**
  - Super-linear power savings

$$P_{total} = \alpha \cdot f \cdot C \cdot V_{dd}^{2} + I_{leak} \cdot V_{dd}$$

  - Minimum energy solution for low-performance designs
- **Limitations**
  - PVT variation
  - Interconnect delay
  - Lack of a systematic design methodology

Subthreshold Operation

Hearing Aids



Distributed Sensor Networks

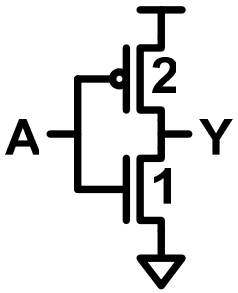
Watches

- **Main Benefit**
  - Super-linear power savings

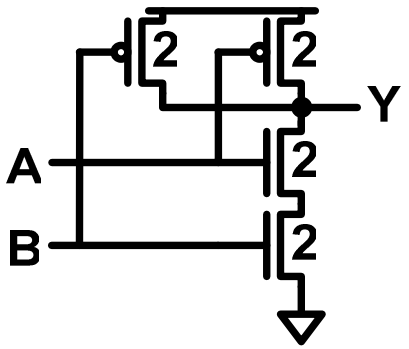
$$P_{total} = \alpha \cdot f \cdot C \cdot V_{dd}^{2} + I_{leak} \cdot V_{dd}$$

  - Minimum energy solution for low-performance designs
- **Limitations**
  - PVT variation
  - Interconnect delay
  - Lack of a systematic design methodology

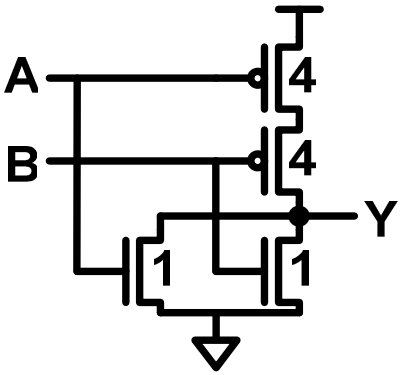
Conventional Logical Effort



$p = 1$
$g = 1$



$p = 6/3 = 2$
$g = 4/3$

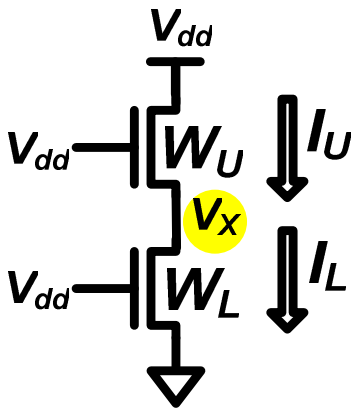


$p = 6/3 = 2$
$g = 5/3$

p: parasitic delay, *g*: logical effort

- **Key Assumptions**
  1) Devices in a stack are equally sized
  2) Resistance of an *n*-stack is *n* times the resistance of a single device
  3) $W_P:W_N \sim 2:1$
- **These assumptions must be revisited for optimal sizing of subthreshold circuits**

Assumption 1: Sizing within the Stack



$$I_U = W_U e^{\frac{(V_{dd}-V_X)-(V_{t0}+\gamma V_X+\lambda_d(V_{dd}-V_X))}{mV_T}}\left(1-e^{\frac{-(V_{dd}-V_X)}{V_T}}\right)$$

$$I_L = W_L e^{\frac{V_{dd}-(V_{t0}+\lambda_d V_X)}{mV_T}}\left(1-e^{\frac{-V_X}{V_T}}\right)$$

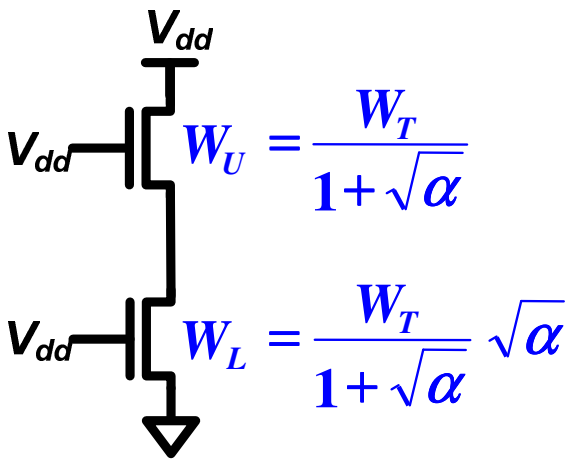
- **We introduce the following constant:**
  - $\alpha = e^{\frac{-\lambda_d V_{dd}}{mV_T}}$,

  and use straight forward steps to solve for V_x:
  - $V_X = \frac{kT}{q}\ln\left(1+\frac{\alpha W_U}{W_L}\right)$

Assumption 1: Continued

- **Equate the currents and define $W_T = W_U + W_L$ to eliminate W_L :**
  - $$I_U = I_L = \frac{\alpha W_U (W_T - W_U)}{\alpha W_U + W_T - W_U} e^{\frac{V_{dd} - V_{t0}}{m V_T}}$$

- **Setting $\partial I_U / \partial W_U = 0$ gives us the optimal sizing factors:**



$$W_U = \frac{W_T}{1+\sqrt{\alpha}}$$

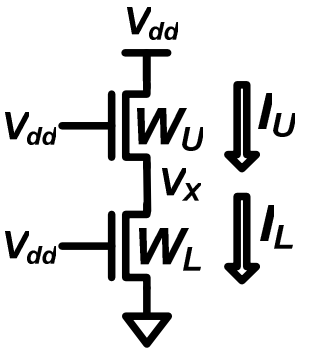
$$W_L = \frac{W_T}{1+\sqrt{\alpha}} \sqrt{\alpha}$$

$$\left(\alpha = e^{\frac{-\lambda_d V_{dd}}{m V_T}} \right)$$

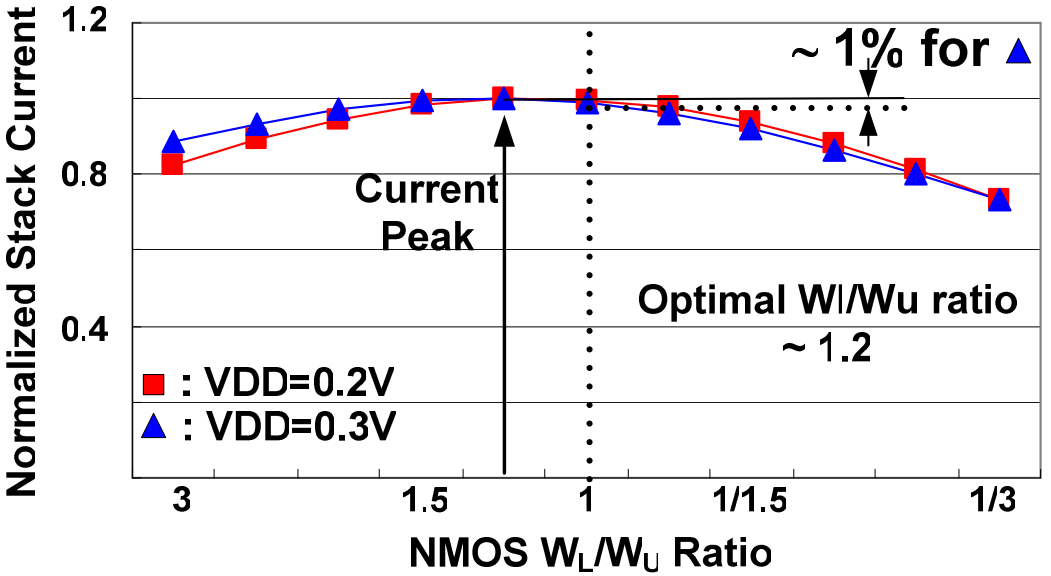
Device nearest to the supply rail is sized up by a factor of $\sqrt{\alpha}$

Sizing Ratio within the Stack

	Device	V_{dd} = 0.3V	V_{dd} = 0.2V
NMOS	W_U	1μm	1μm
	$W_L = \sqrt{\alpha} W_U$	1.189 μm	1.122 μm



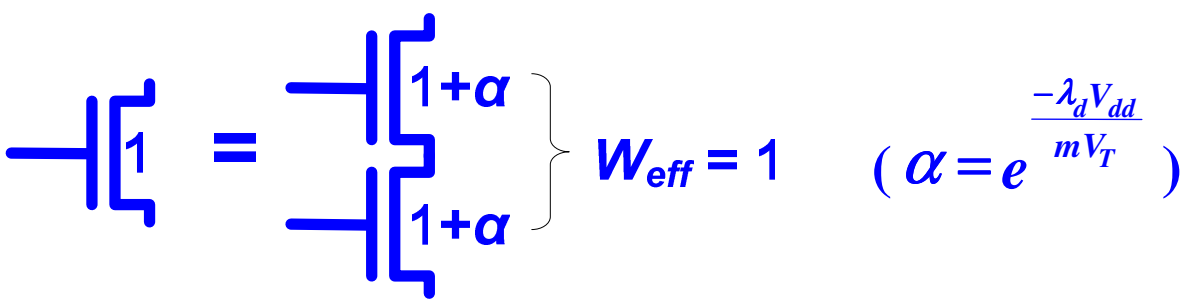
0.13μm, 1.2V, RT



- Benefit of optimal sizing is negligible (~1%)
- Use 1:1 stack ratio for simplicity

Assumption 2: Effective Width of Stacks

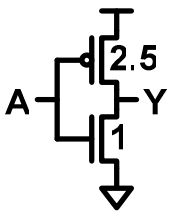
- **Defining $W = W_U = W_L$, we find:**
  - $$I_U = I_L = \frac{\alpha W^2}{\alpha W + W} e^{\frac{V_{dd} - V_{t0}}{mV_T}} = \frac{\alpha}{1+\alpha} W e^{\frac{V_{dd} - V_{t0}}{mV_T}}$$
- **For a single transistor, the current equation is:**
  - $$I = W_{eff} e^{\frac{V_{dd} - (V_{t0} + \lambda_d V_{dd})}{mV_T}} \approx \alpha W_{eff} e^{\frac{V_{dd} - V_{t0}}{mV_T}}$$
- **These two equations tell us:** $$\alpha W_{eff} = \frac{\alpha}{1+\alpha} W \rightarrow W_{eff} = \frac{1}{1+\alpha} W$$
- **Two stacked transistors should be sized up by a factor of 1+α**



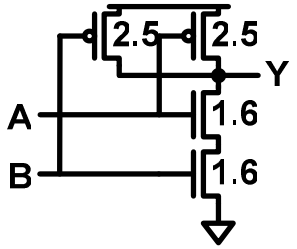
New Subthreshold Logical Effort

Simulation results vs. our theoretical results (UMC 0.13µm)

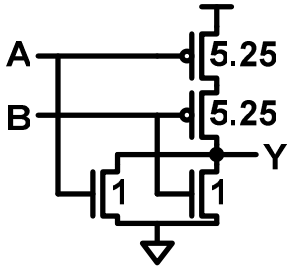
	V_{dd} = 0.3V		V_{dd} = 0.2V	
	Measured	Theoretical 1+α	Measured	Theoretical 1+α
PMOS	2.64	2.707	2.4	2.428
NMOS	2.44	2.413	2.25	2.259



p = 1
g = 1



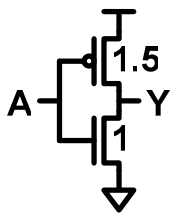
p = 6.6/3.5
g = 4.1/3.5



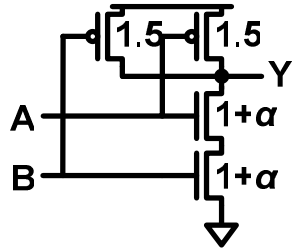
p = 7.25/3.5
g = 6.25/3.5

Conventional logical effort

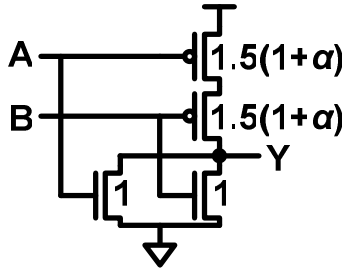
1) 1:1 inter-stack sizing
2) Size up by factor of 2
3) 2.5 : 1 beta ratio



p = 1
g = 1



p = (4+α)/2.5
g = (2.5+α)/2.5



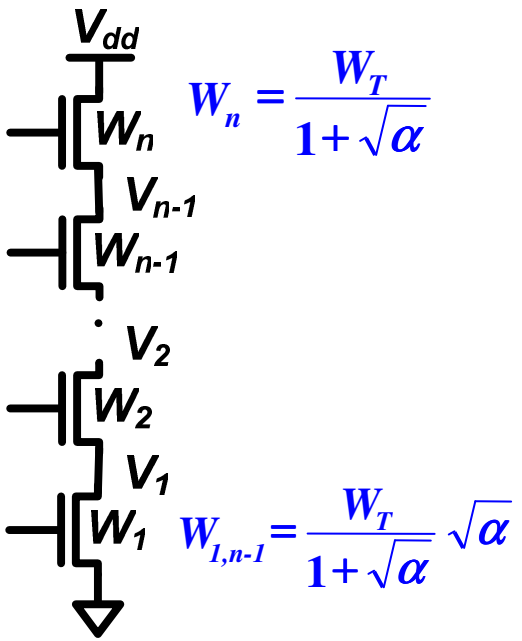
p = (3.5+1.5α)/2.5
g = (2.5+1.5α)/2.5

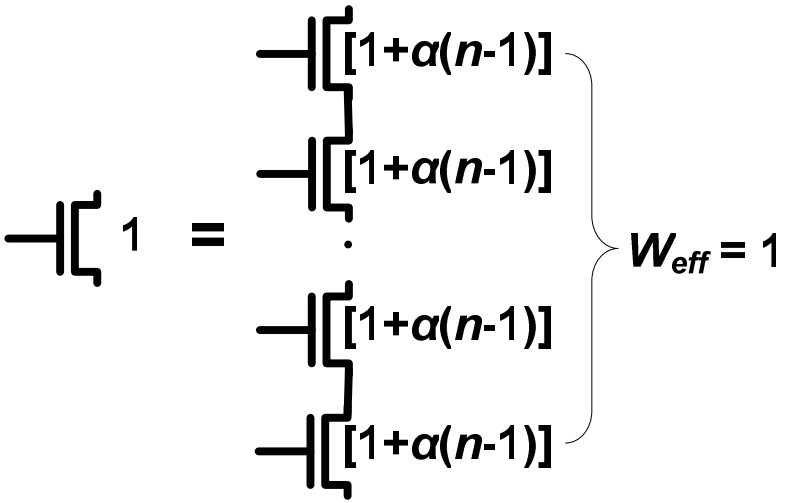
Subthreshold logical effort

1) 1:1 inter-stack sizing
2) Size up by factor of 1 + α
3) 1.5 : 1 beta ratio

Optimal Sizing for *n*-Stacks

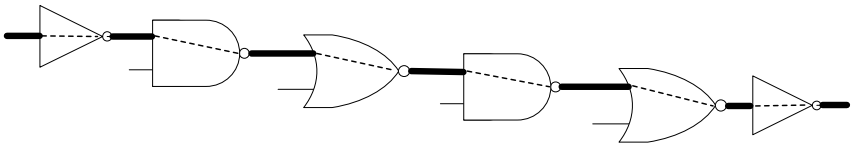
- Use same procedure introduced for the 2-stack
- Choose equivalent size for all devices for negligible performance hit
- Each device in an *n*-stack should be scaled up by a factor of [1+α(*n*-1)]





[1+α(*n*-1)] : stack sizing factor for *n*-stack

Test Cases: Optimizing a Critical Path



- Modified the widths in order to achieve equal delays through the two paths using the worst-case input patterns for each ("stack" and "fast" paths)
- Test 3 cases:
  1) Only the new stack sizing factors are used
  2) Use 1.5:1 PMOS:NMOS width ratio with strong-inversion stack sizing
  3) Combine the two improvements

	W_P:W_N	New stack sizing?	V_{dd} = 0.3V		V_{dd} = 0.2V	
			Fast Path	Stack Path	Fast Path	Stack Path
Conventional	2.5:1	No	18.08ns	20.34ns	140.0ns	150.1ns
Case 1	2.5:1	Yes	-9.79%	13.17%	-4.54%	13.38%
Case 2	1.5:1	No	22.08%	11.22%	20.78%	11.32%
Case 3 (proposed)	1.5:1	Yes	7.21%	22.17%	11.40%	20.99%

ISCAS Benchmark Results

Circuit	V_{dd} = 0.3V			V_{dd} = 0.2V		
	Conv.	Proposed	Speedup	Conv.	Proposed	Speedup
C432	12.93 ns	11.55 ns	10.67%	99.44 ns	89.38 ns	10.11%
C6288	24.71 ns	21.59 ns	12.63%	186.0 ns	170.6 ns	8.31%
C3540	35.06 ns	33.53 ns	4.38%	270.6 ns	253.6 ns	6.29%
C1355	12.40 ns	10.73 ns	13.46%	103.1 ns	90.41 ns	12.32%
74283	43.74 ns	41.45 ns	5.25%	340.7 ns	323.4 ns	5.08%
74181	47.70 ns	44.74 ns	6.20%	378.8 ns	353.1 ns	6.78%
74L85	22.88 ns	21.37 ns	6.59%	185.2 ns	170.7 ns	7.80%
74182	29.18 ns	19.52 ns	33.1%	215.3 ns	146.2 ns	32.1%

- **Three cell libraries were created**
  1) **Conventional strong-inversion sizing**
  2) **Optimized for 0.3V supply** ($\alpha = e^{\frac{-\lambda_d \times 0.3}{mV_T}}$)
  3) **Optimized for 0.2V supply** ($\alpha = e^{\frac{-\lambda_d \times 0.2}{mV_T}}$)

Conclusion

- **Subthreshold design holds promise for emerging ultra-low power applications**
- **Design methodologies must be reformulated**
- **New derivation of the optimal stack sizing for subthreshold**
- **Proposed subthreshold logical effort methodology led to performance benefits ranging from 4.4% to 33.1%**