EECS 427
Lecture 22: Low and Multiple-Vdd Design

Reading: 11.7.1

Last Time

- Low power ALUs
  - Glitch power
  - Clock gating
  - Bus recoding
- The low power design space
  - Dynamic vs static

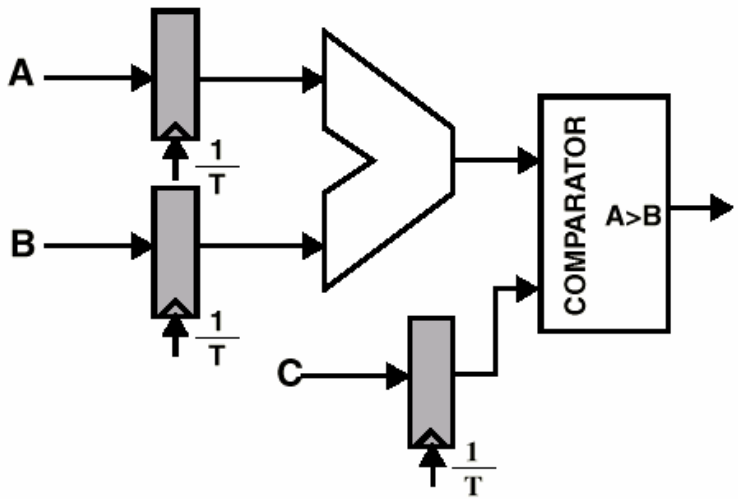
Lecture Overview

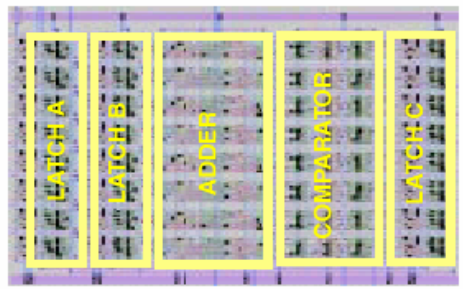
- Low Vdd design
  - Pipelining
  - Parallel
- Multiple Vdd design
  - Concept
  - Level converter topologies
  - Dual-Vdd buffer design for global wires

Power and Energy Design Space

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	Design Time	Non-active Modules	Run Time
Active	Logic Design Reduced V_{dd} Sizing Multi-V_{dd}	Clock Gating	DFS, DVS (Dynamic Freq, Voltage Scaling)
Leakage	+ Multi-V_T	Sleep Transistors Variable V_T	+ Variable V_T

Architecture Tradeoff for Fixed-rate Processing Reference Datapath



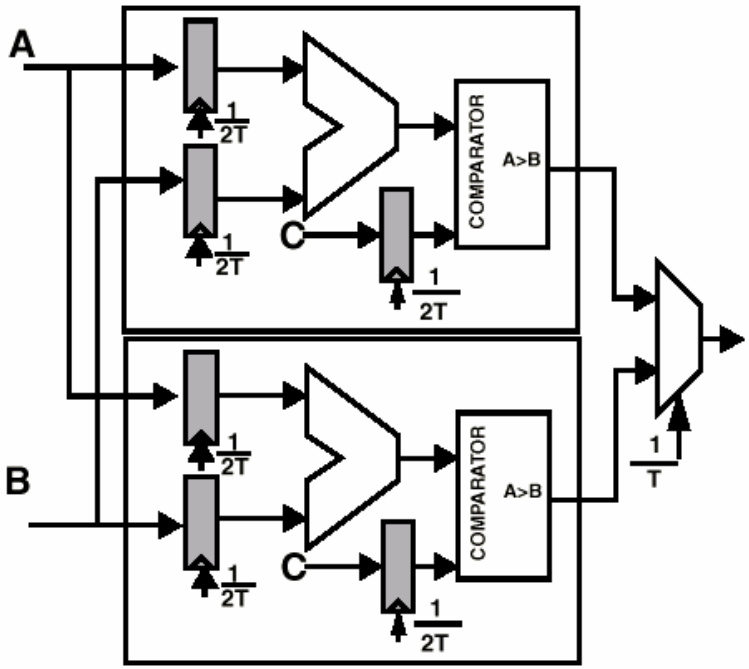


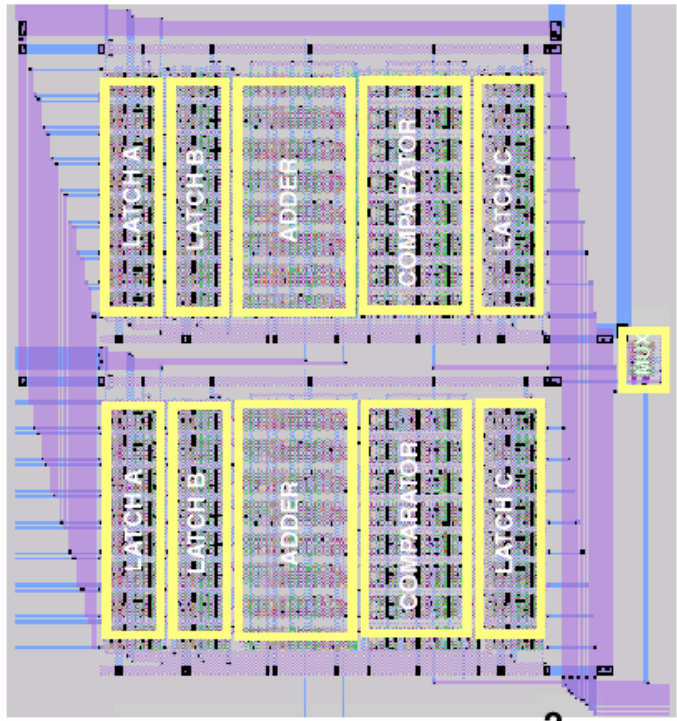
Area = 636 x 833 μ^2

- Critical path delay $\Rightarrow T_{adder} + T_{comparator}$ (= 25ns)
  $\Rightarrow f_{ref} = 40Mhz$
- Total capacitance being switched = C_{ref}
- $V_{dd} = V_{ref} = 5V$
- Power for reference datapath = $P_{ref} = C_{ref} V_{ref}^2 f_{ref}$

from [Chandrakasan92] (*IEEE JSSC*)

Parallel Datapath

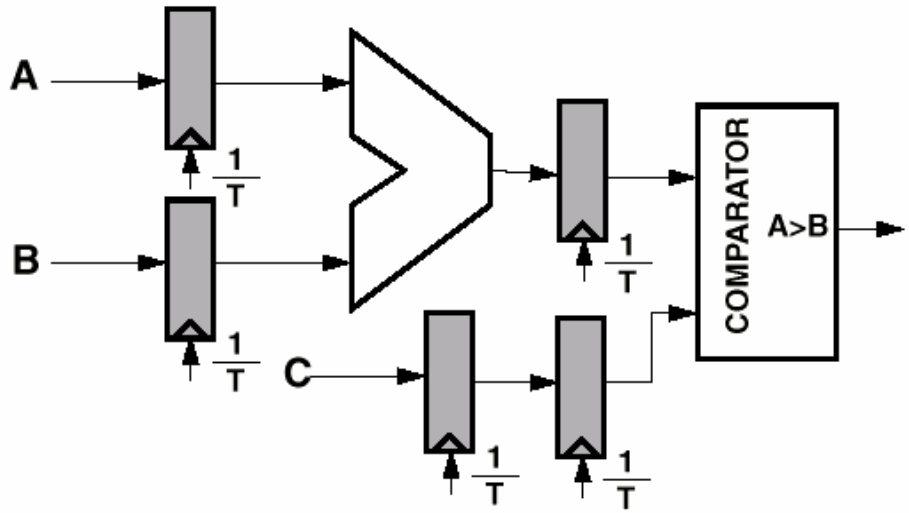


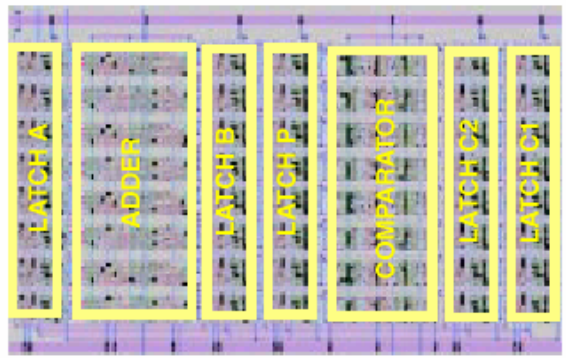


Area = 1476 x 1219 μ^2

- The clock rate can be reduced by half with the same throughput $\Rightarrow f_{par} = f_{ref} / 2$
- $V_{par} = V_{ref} / 1.7$, $C_{par} = 2.15 C_{ref}$
- $P_{par} = (2.15 C_{ref}) (V_{ref}/1.7)^2 (f_{ref}/2) \approx 0.36 P_{ref}$

Pipelined Datapath





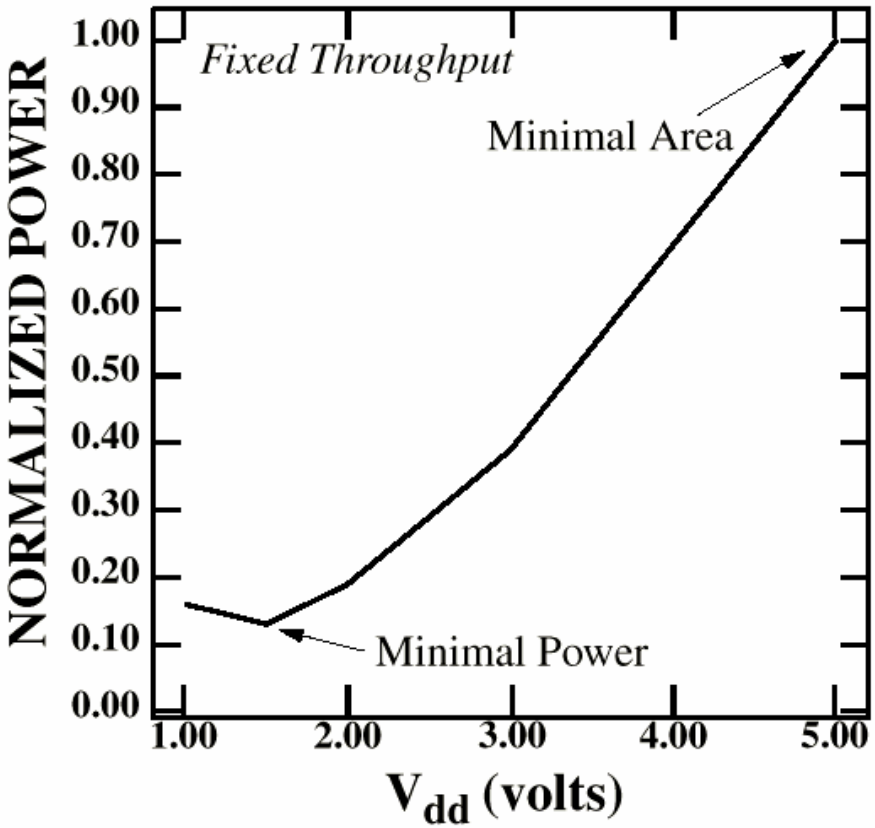
Area = 640 x 1081 μ^2

- Critical path delay is less $\Rightarrow$ max [T_{adder} , $T_{comparator}$]
- Keeping clock rate constant: $f_{pipe} = f_{ref}$
  Voltage can be dropped $\Rightarrow V_{pipe} = V_{ref} / 1.7$
- Capacitance slightly higher: $C_{pipe} = 1.15C_{ref}$
- $P_{pipe} = (1.15C_{ref})\ (V_{ref}/1.7)^2\ f_{ref} \approx 0.39\ P_{ref}$

A Simple Datapath: Summary

Architecture type	Voltage	Area	Power
Simple datapath (no pipelining or parallelism)	5V	1	1
Pipelined datapath	2.9V	1.3	0.39
Parallel datapath	2.9V	3.4	0.36
Pipeline-Parallel	2.0V	3.7	0.2

How Low a Voltage can be Used?



- Capacitance overhead starts to dominate at "high" levels of parallelism and results in an optimum voltage

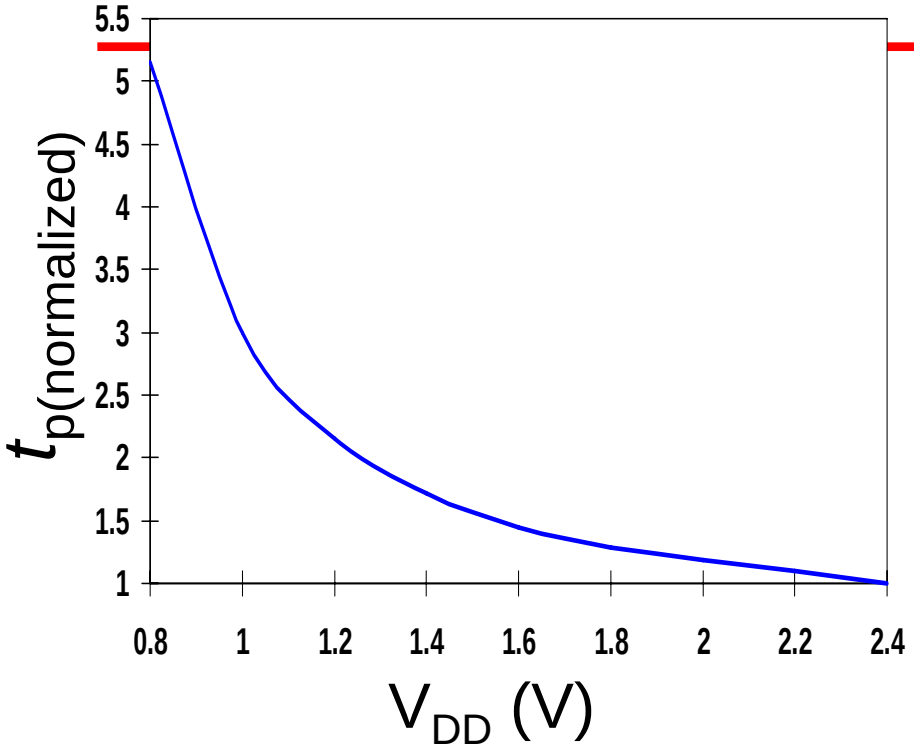
Power and Energy Design Space Revisited

	Constant Throughput/Latency		Variable Throughput/Latency
Energy	Design Time	Non-active Modules	Run Time
Active	Logic Design Reduced V_{dd} Sizing Multi-V_{dd}	Clock Gating	DFS, DVS (Dynamic Freq, Voltage Scaling)
Leakage	+ Multi-V_T	Sleep Transistors Multi-V_{dd} Variable V_T	+ Variable V_T

Supply Voltage Scaling

- How to maintain throughput under reduced supply?
- Introducing more parallelism/pipelining
  - Area increase – cost increases
  - Cost/power tradeoff
- **Multiple voltage domains**
  - Separate supply voltages for different blocks
  - Lower VDD for slower blocks
  - Cost of DC-DC converters or additional off-chip supplies, distributing multiple power supplies on-chip
- Dynamic voltage scaling – with variable throughput
- Reduce V_{th} to improve speed
  - Exponentially increased leakage eventually dominates

Delay as a Function of V_{DD}



$$T_d = \frac{C_L * V_{dd}}{I}$$

$$I \sim (V_{dd} - V_t)^{1.3}$$

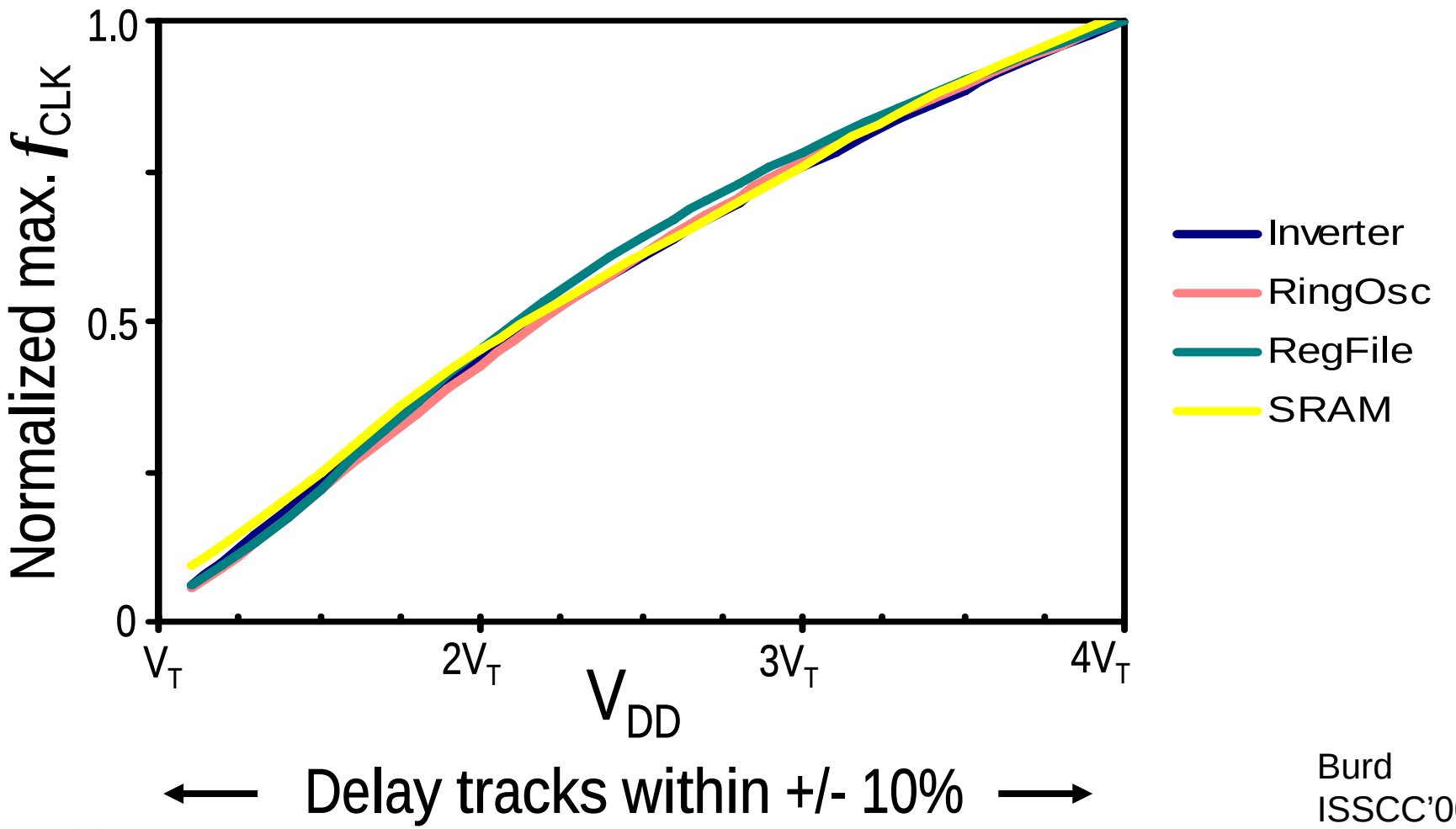
$$\frac{T_{d(Vdd=1.5)}}{T_{d(Vdd=2.5)}} = \frac{(1.5) * (2.5 - 0.4)^{1.3}}{(2.5) * (1.5 - 0.4)^{1.3}}$$

$$\approx 1.4$$

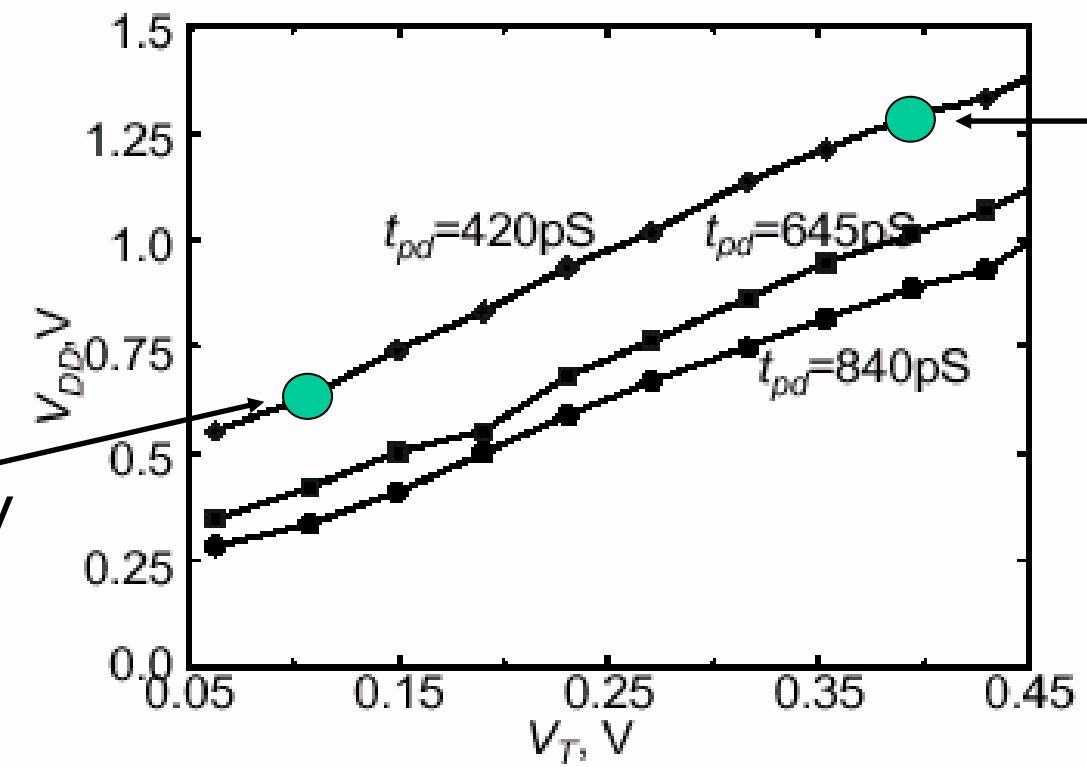
- Decreasing V_{DD} reduces dynamic energy consumption quadratically
- But increases gate delay (decreases performance)

- Determine critical path(s) at **design time** & use high V_{DD} for transistors on those paths for speed. Use lower V_{DD} on other gates

CMOS Circuits Track Over V_{DD}



Changing V_{dd} and V_{th} Together



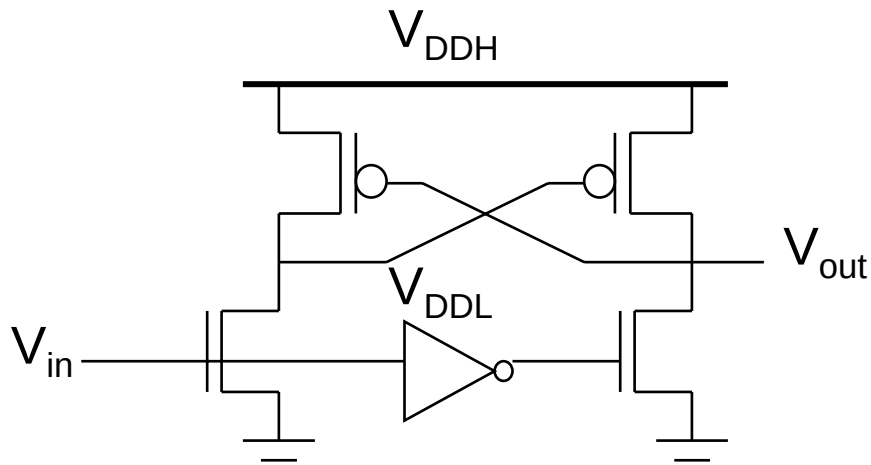
Vdd=1.3V
Vth = 0.39V

Vdd=0.62V
Vth = 0.11V

Contours of constant delay show that reductions in V_{th} must accompany smaller V_{dd}'s to maintain speed

Multiple V_{DD} Considerations

- How many V_{DD}? – 2 is becoming more popular
  - Many chips already have 2 supplies (1 for core and 1 for I/O)
- When combining multiple supplies, **level converters** are required when a module at lower supply drives gate at higher supply (step-up)
  - If a gate supplied with V_{DDL} drives a gate at V_{DDH}, PMOS never turns off
    - Cross-coupled PMOS transistors perform the level conversion
    - NMOS transistors operate at reduced supply
  - Level converters are **not** needed for step-down changes in voltage

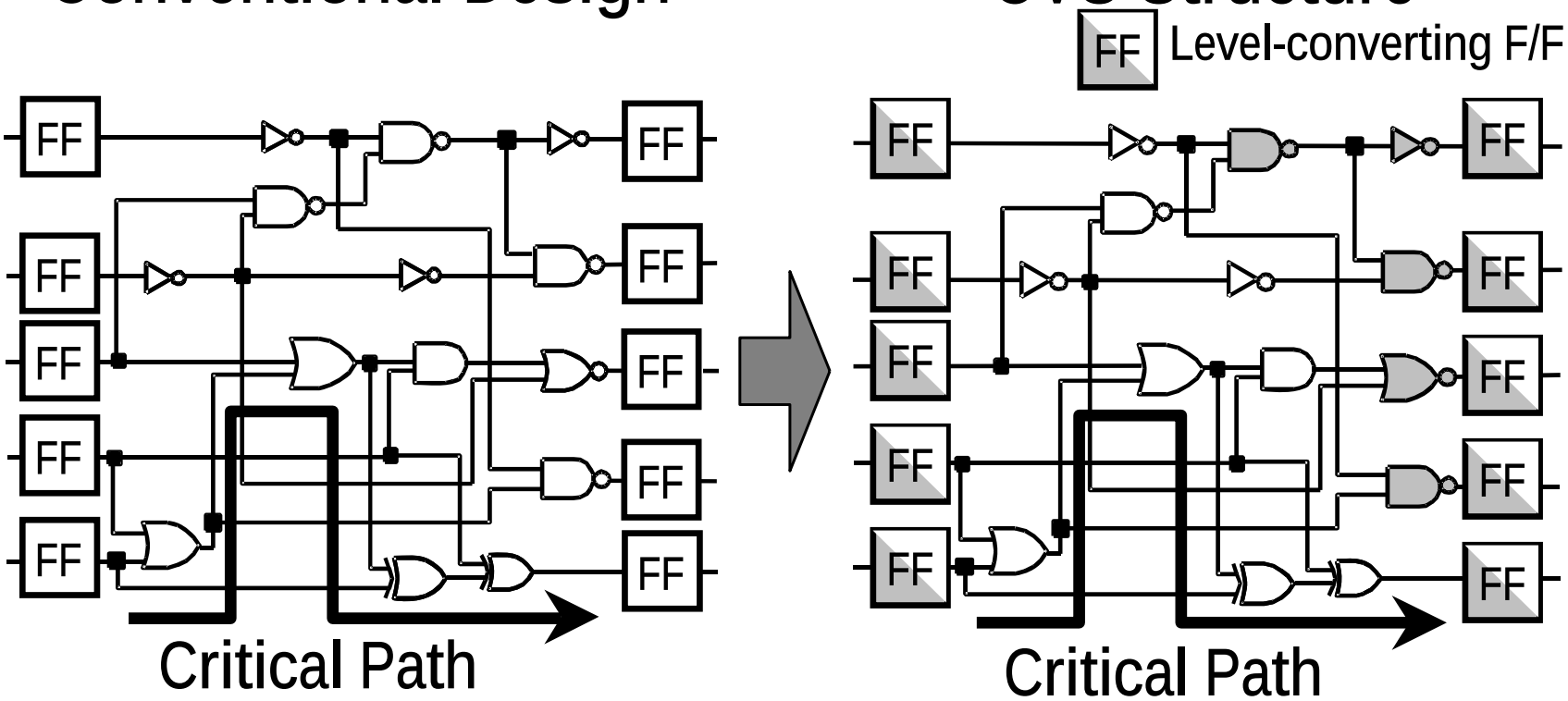


  - Overhead of level converters can be reduced by converting at register boundaries & embedding level conversion inside the flop

Multiple Vdd Design

Conventional Design

CVS Structure



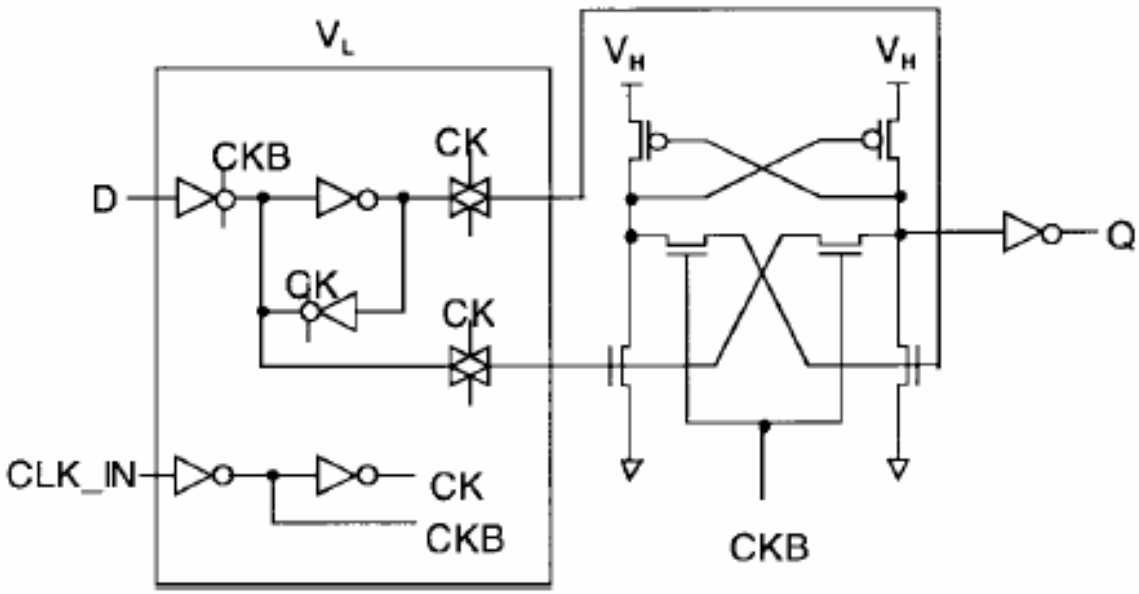
Lower V_{DD} portion is shaded

"Clustered voltage scaling"

M.Takahashi, ISSCC'98.

Level converting flip flops

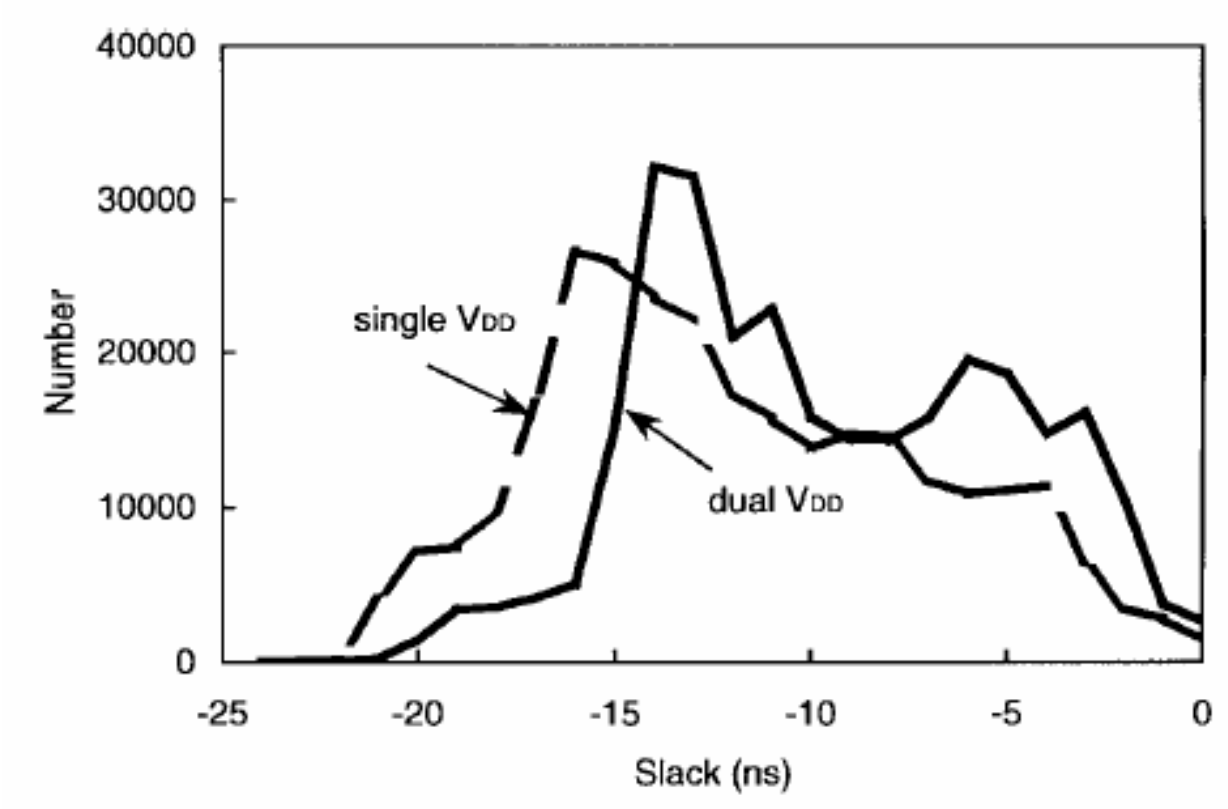
- Needed to restore the input to the next pipeline to V_H



Takahashi et. al *JSSC 1998*

Effect of CVS on path distribution

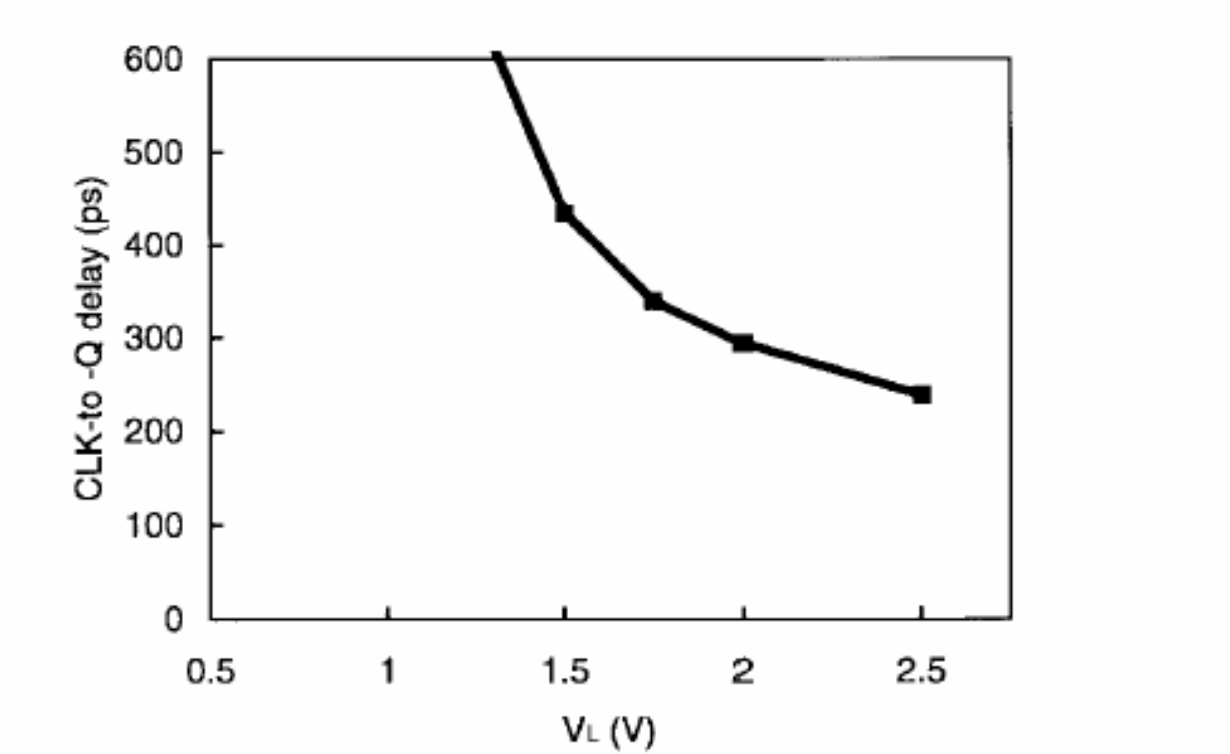
- "Shift" the histogram towards the right



Takahashi et. al *JSSC 1998*

Delay Penalty

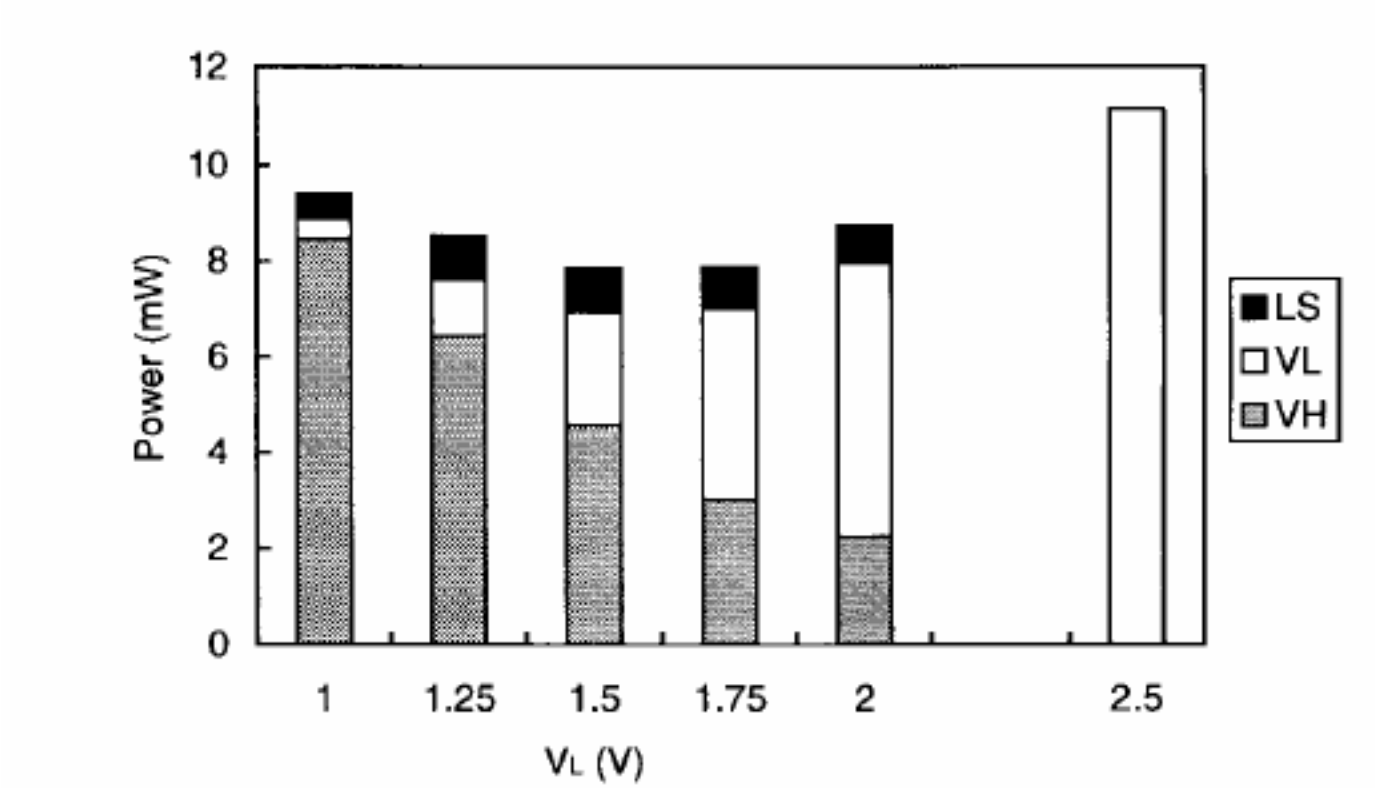
- Significant delay penalty
  - Swing voltage unchanged (Linear effect)
  - Drive voltage shrinks (Quadratic effect)



Takahashi et. al *JSSC 1998*

Power dissipation dependence on V_L

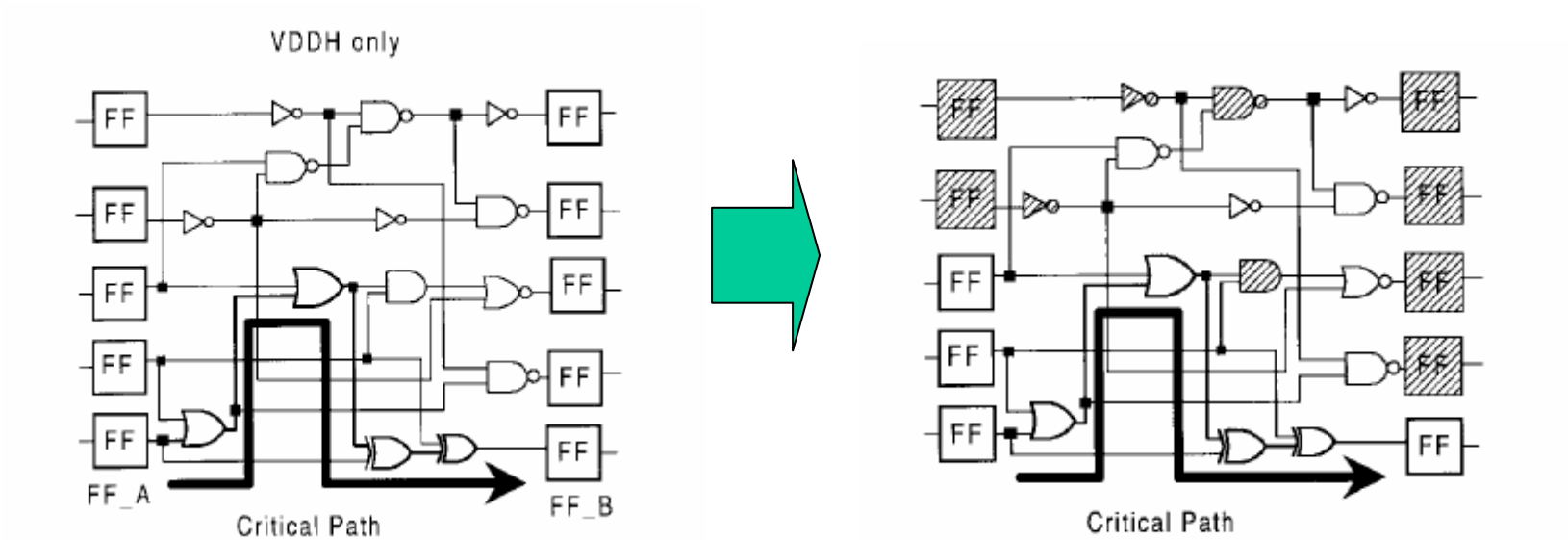
- Setting V_L too low results in less paths with low V_{dd} assignments



Takahashi et. al *JSSC 1998*

ECVS

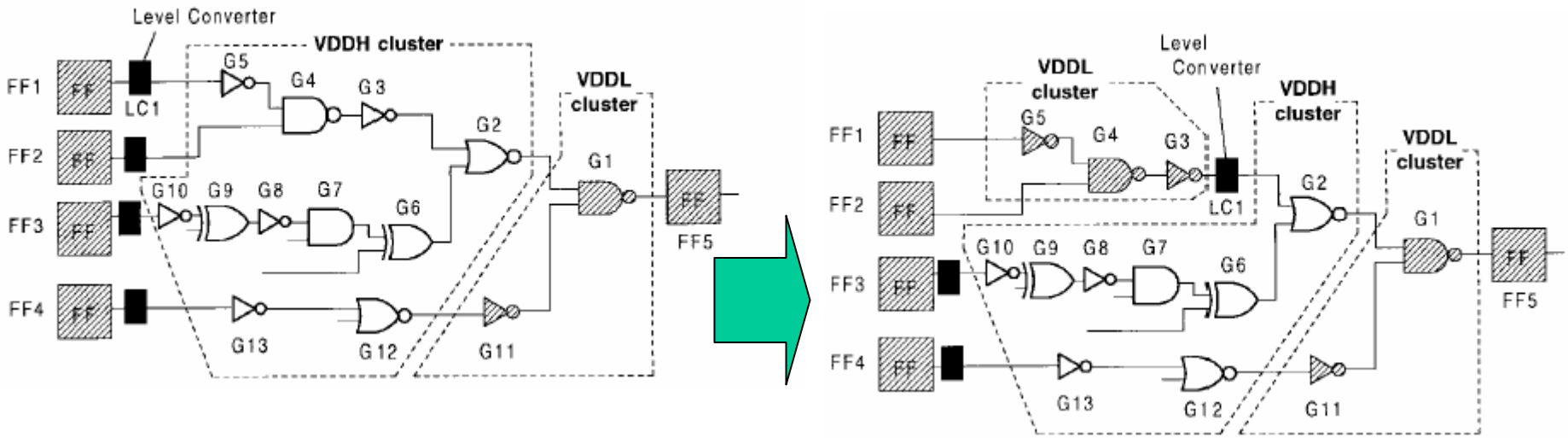
- No longer constrained to a monotonic voltage profile from input to output.
- Requires a level-converter to restore a higher voltage
  - Level converting buffers
  - Level converting gates
  - Level conversion is therefore not restricted to latches



Usami et. al *JSSC 1998*

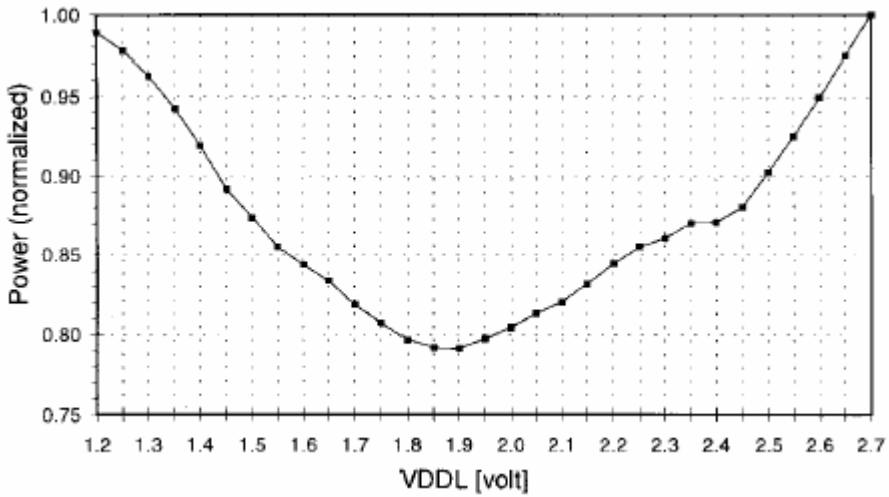
ECVS allows more paths to be assigned to V_L

- Allows delay balancing through voltage assignment
- Must pay delay and power penalty in performing every level conversion (Small clusters may not be worthwhile)
- Algorithms used for concurrent sizing-voltage assignment



Optimal choice for V_L

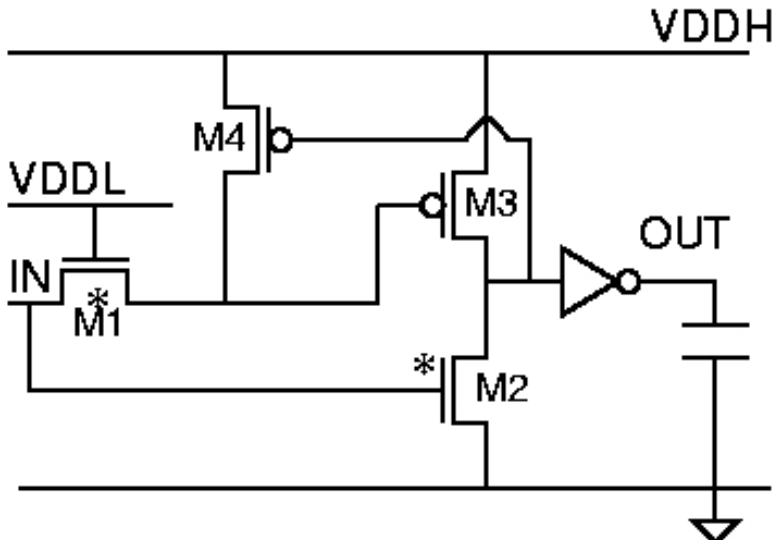
- The choice for V_L depends on the delay histogram with single $V_{DD.}$
- Choosing too large a V_L nullifies the effects of lower power dissipation.
- Choosing too low a V_L results in too few paths being assigned to V_l .



Usami et. al *JSSC 1998*

Existing Level Converters

- DCVS
- Pass gate (PG)

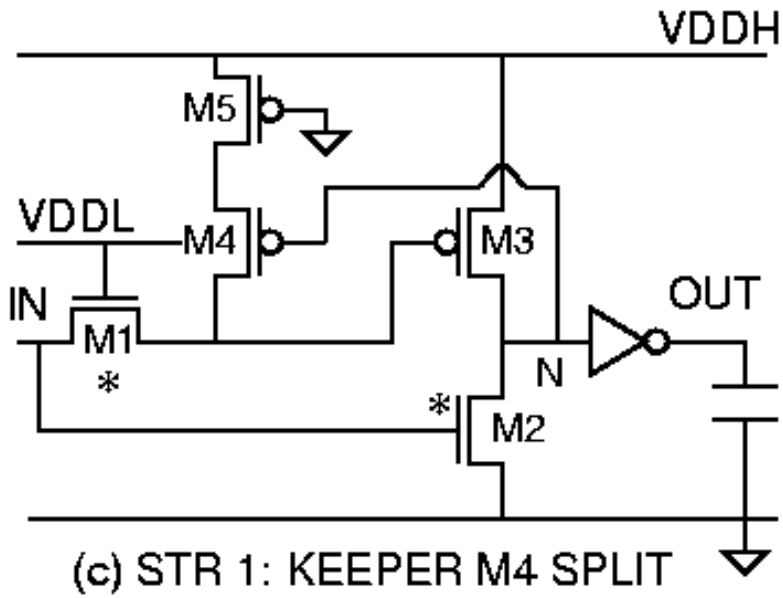


* = low-Vth candidate

- DCVS – Higher power dissipation due to greater contention and higher transistor count
- PG – Simpler design, faster, lower power than DCVS, critical path is falling input (and output)
  - Key: Purpose of M1

Alternate LC 1 : STR1

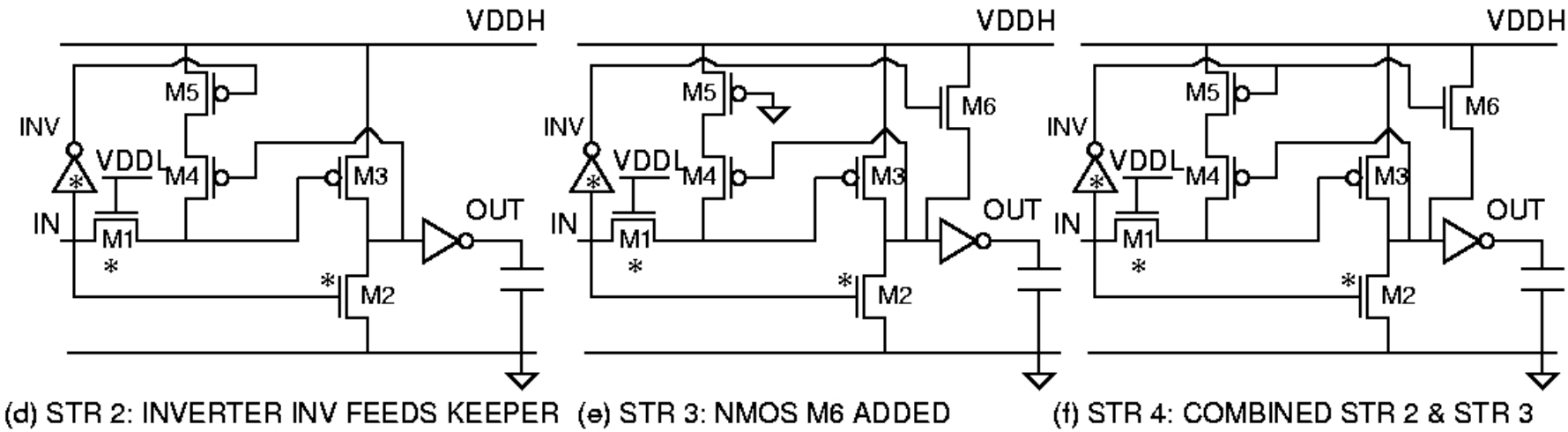
- STR1



- Known high-performance design technique, with much improved results in this application space
- Keeper M4 from PG split into M4 and M5
- Reduced loading on node N and reduced contention

Alternate LCs : STR2, 3 and 4

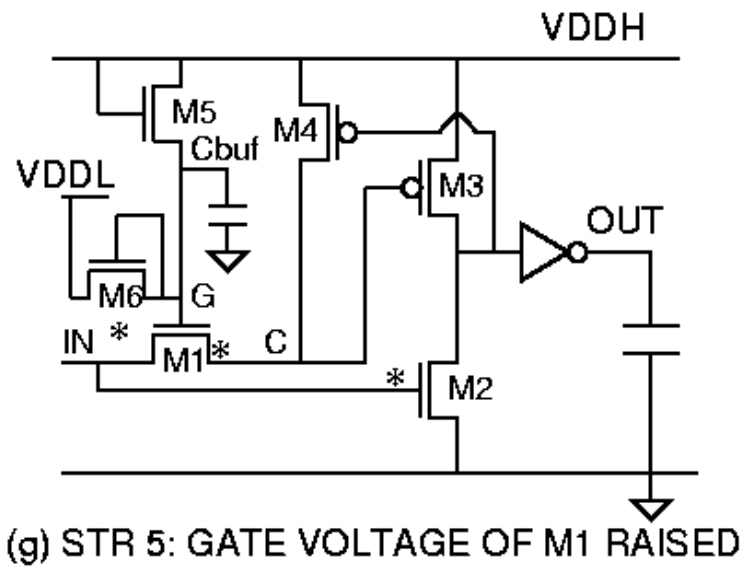
- STR2 STR3 STR4



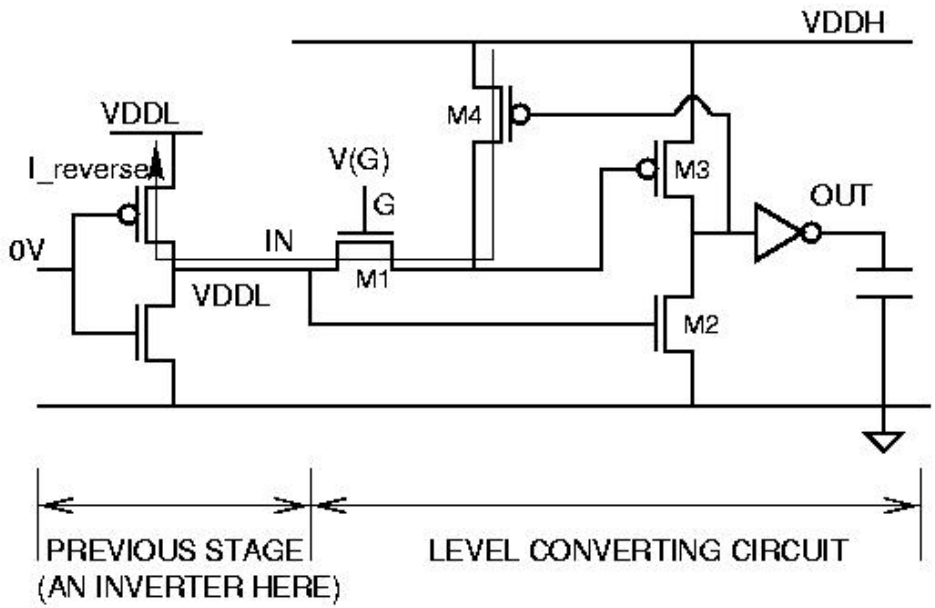
- INV and M6 added to turn off feedback path faster and speed up critical path of the circuit

Alternate LC 5 : STR5

- STR5



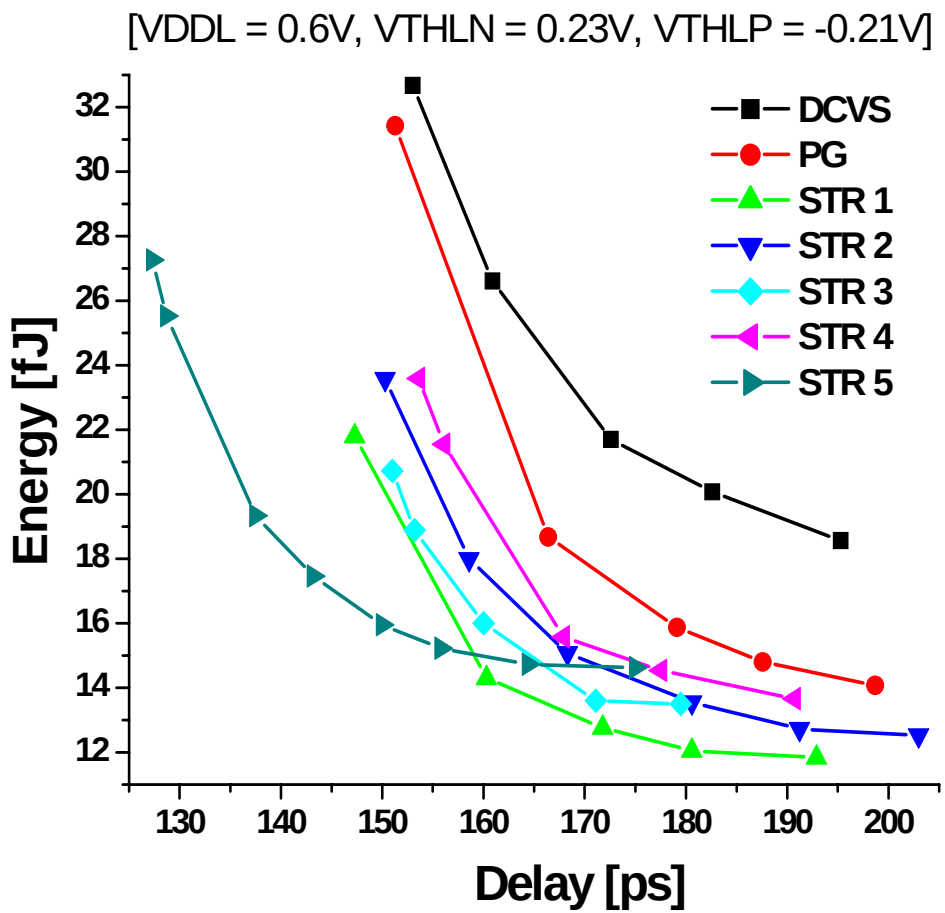
(g) STR 5: GATE VOLTAGE OF M1 RAISED



- Raised gate voltage on pass transistor boosts performance
- Leakage current I_reverse creates tradeoff between power and speed

Simulation Results

- Low VDDL/High VTH
  - STR1,..,4 consume about 40-50% less energy
  - STR1 about 3-4% faster than DCVS and PG
  - STR2, 3 also slightly faster
- Low VDDL/Low VTH
  - STR1 consumes 37% and 15% lower energy than DCVS and PG respectively
- High VDDL
  - STR1 consumes 40% and 15% less energy than DCVS and PG respectively
  - STR1 and 4 faster than DCVS and PG



Summary

- Use of 2 Vdd's on a chip is growing
  - Brings up level conversion, layout, power distribution issues
- Fast, energy efficient level converter topologies are critical to maximize dual-Vdd benefit
- What else can you do with 2 supplies available?