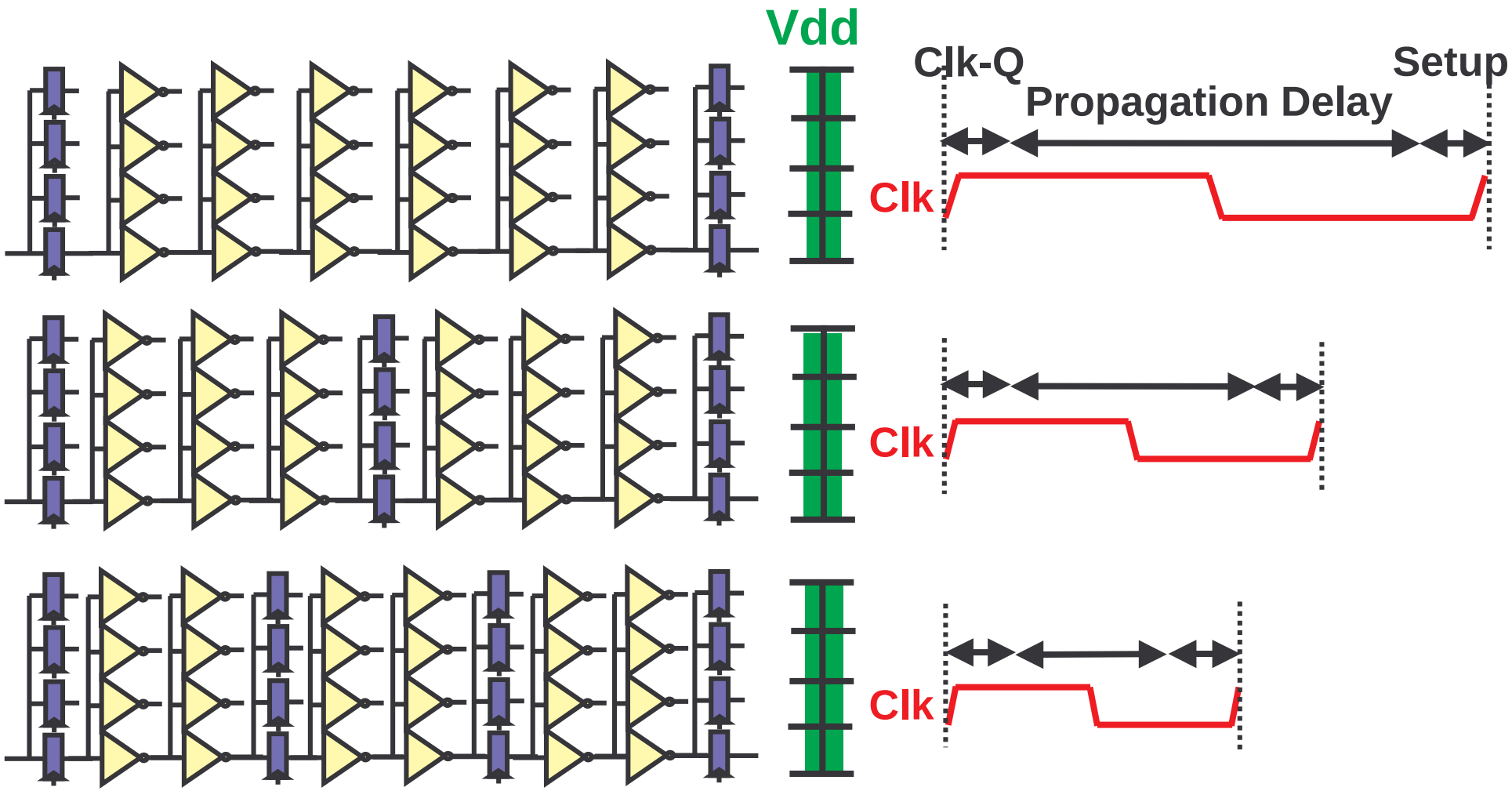
ISLPED 2004
8/10/2004

Power-Optimal Pipelining in Deep Submicron Technology

Seongmoo Heo and Krste Asanović
Computer Architecture Group, MIT CSAIL

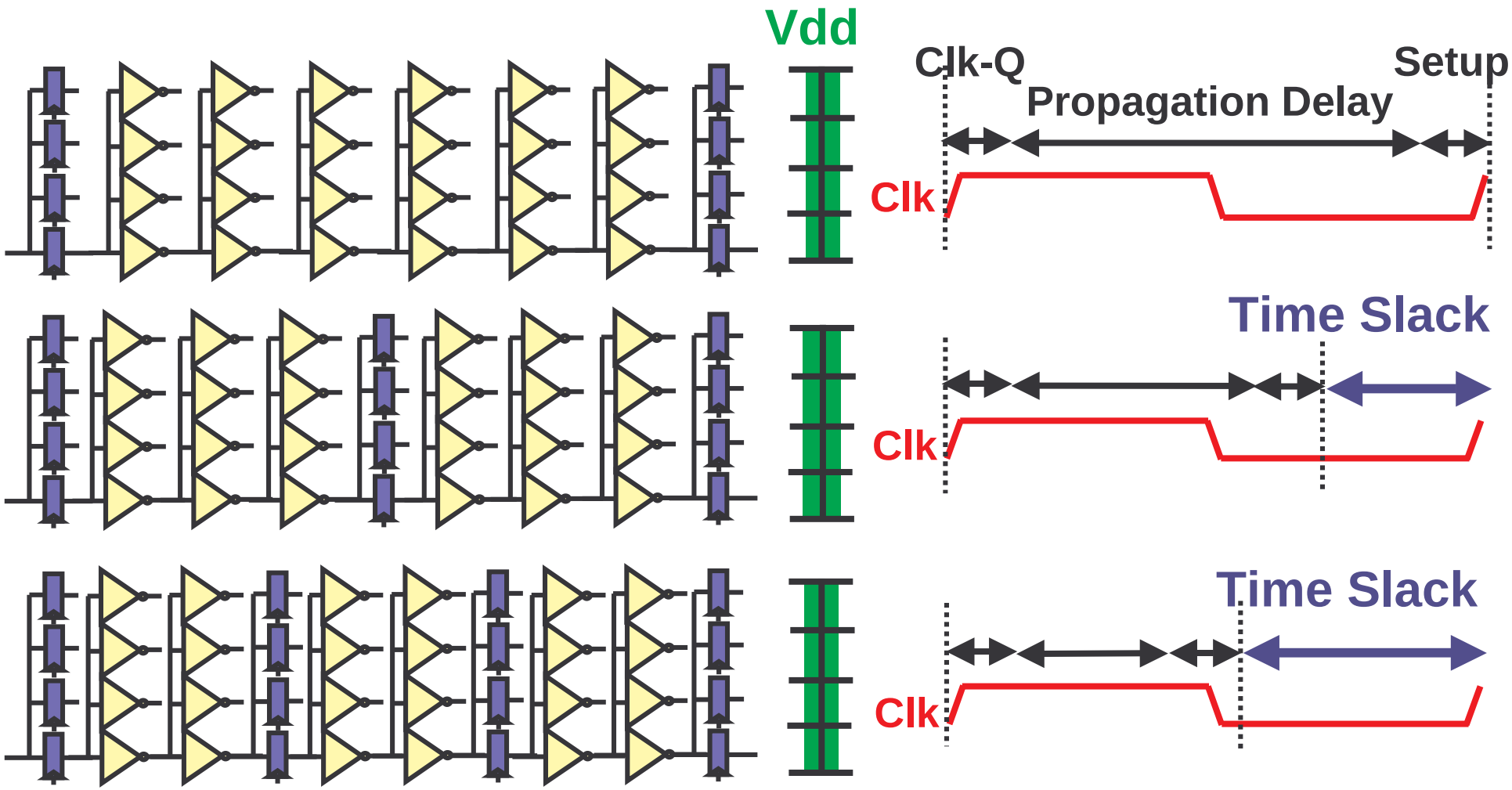
Traditional Pipelining

- **Goal: Maximum performance**



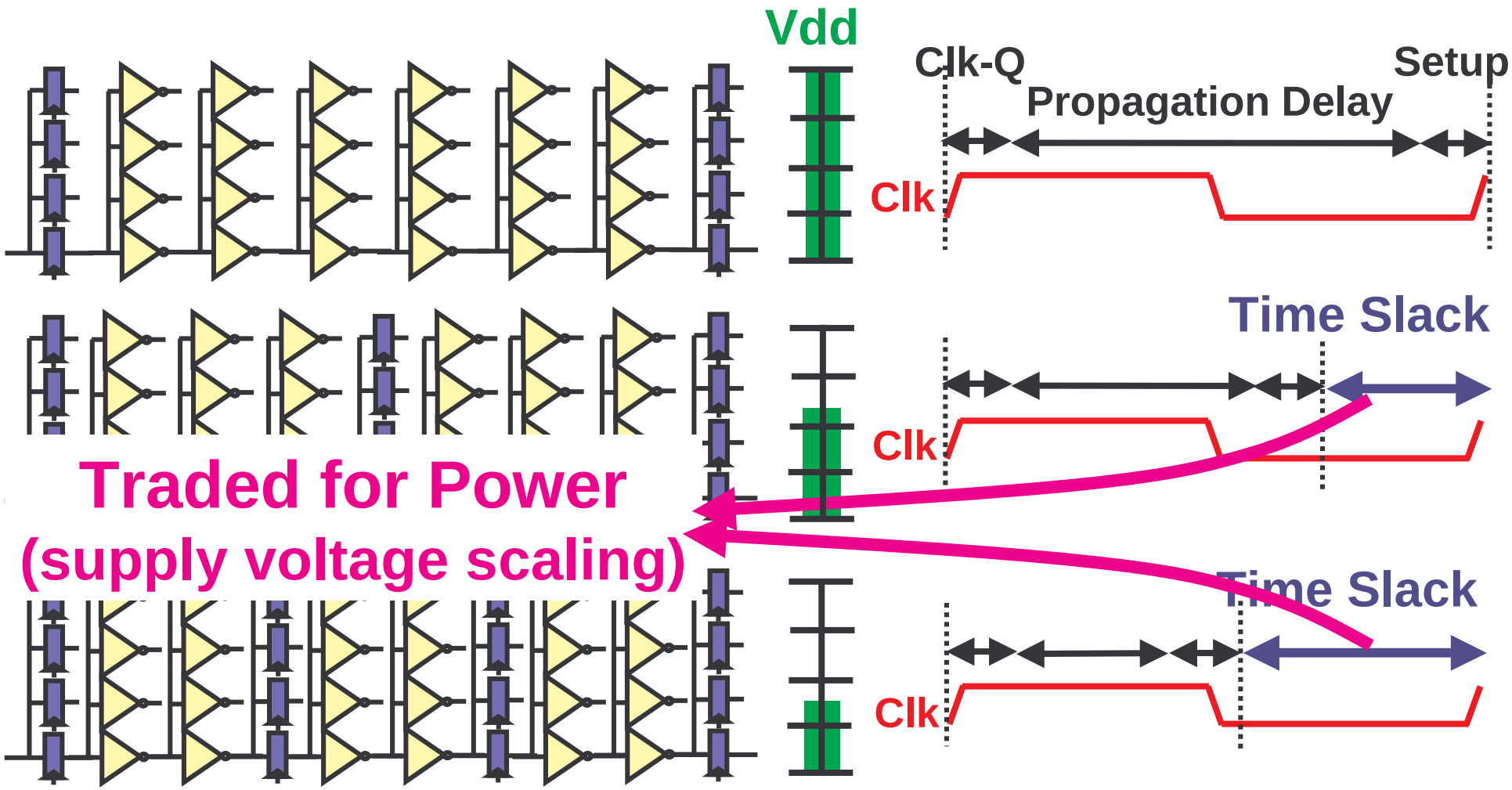
Pipelining as a Low-Power Tool

- Goal: Low-Power, Fixed Throughput

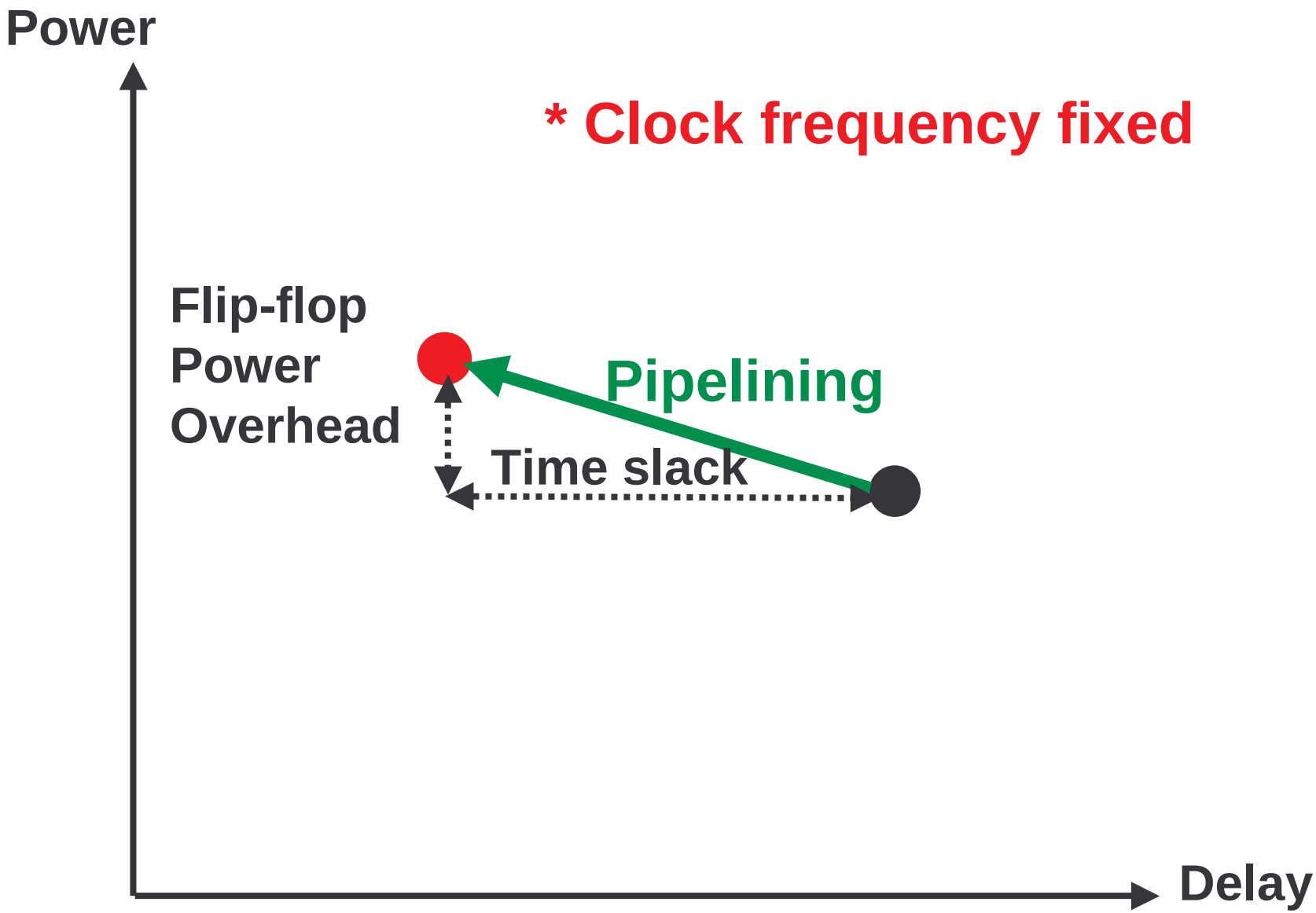


Pipelining as a Low-Power Tool

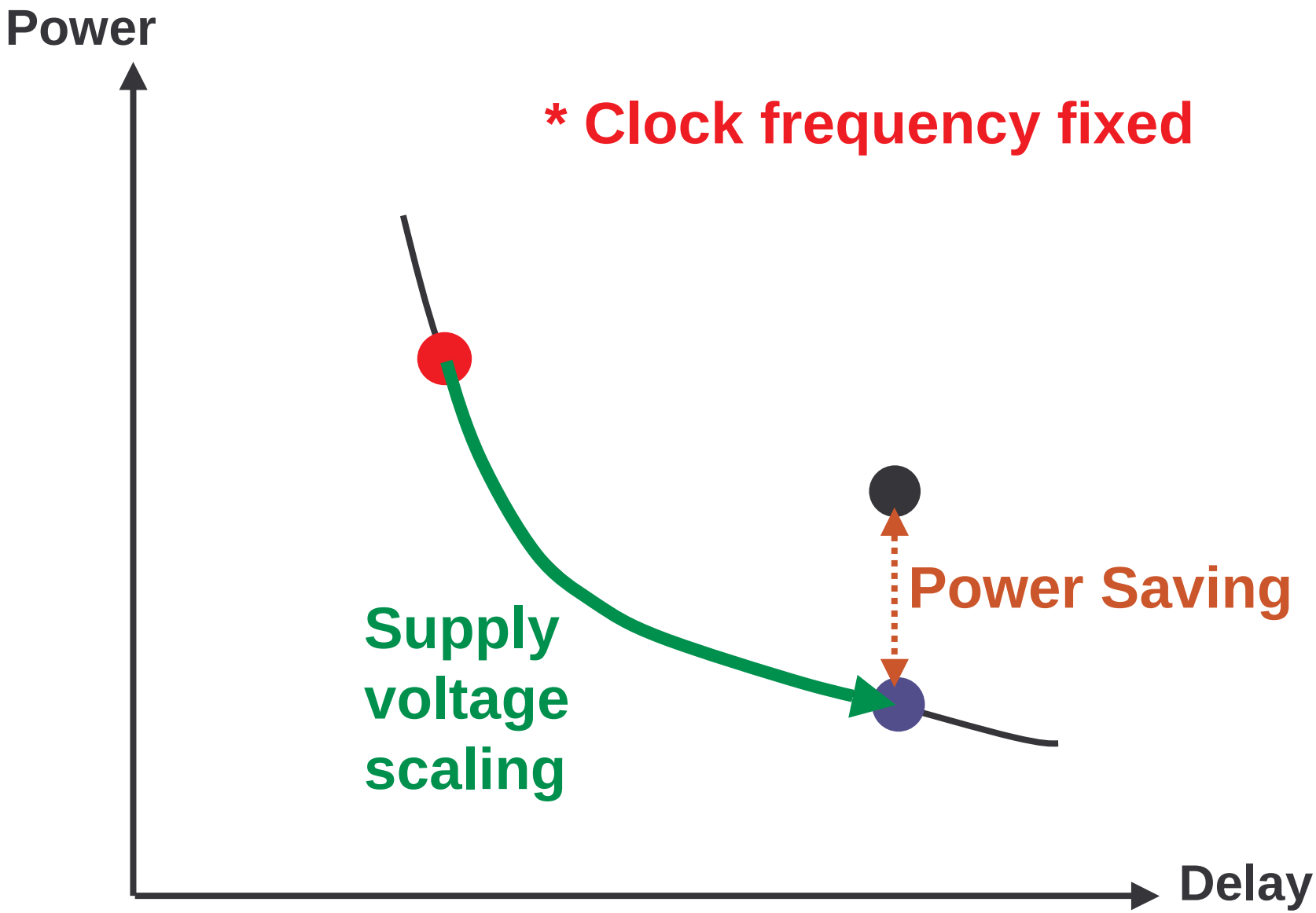
- **Goal: Low-Power, Fixed Throughput**



Pipelining as a Low-Power Tool



Pipelining as a Low-Power Tool

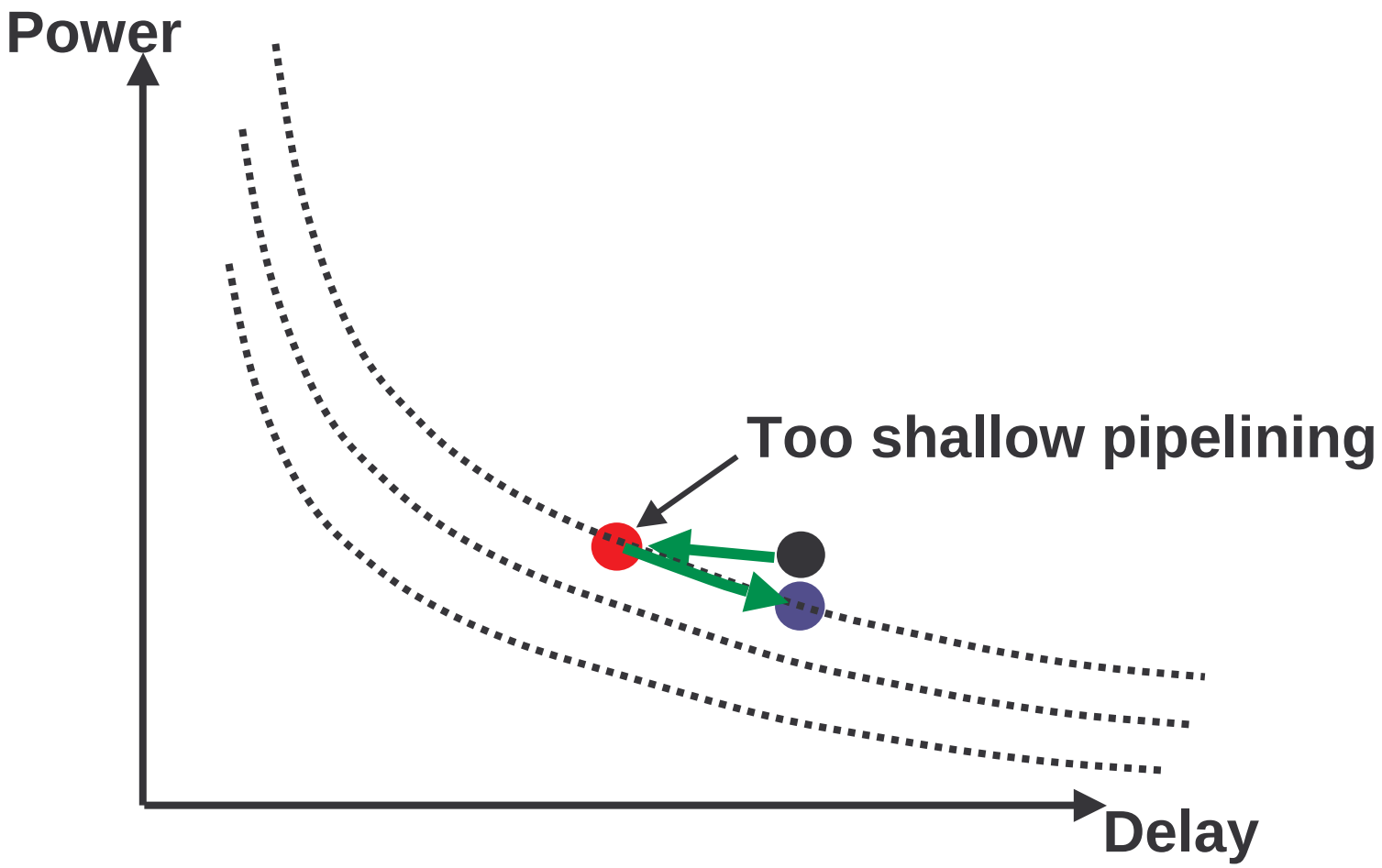


Power-Optimal Pipelining

- **Power reduction from pipelining limited by power overhead of increased number of flip-flops**
  → Power-Optimal Pipelining

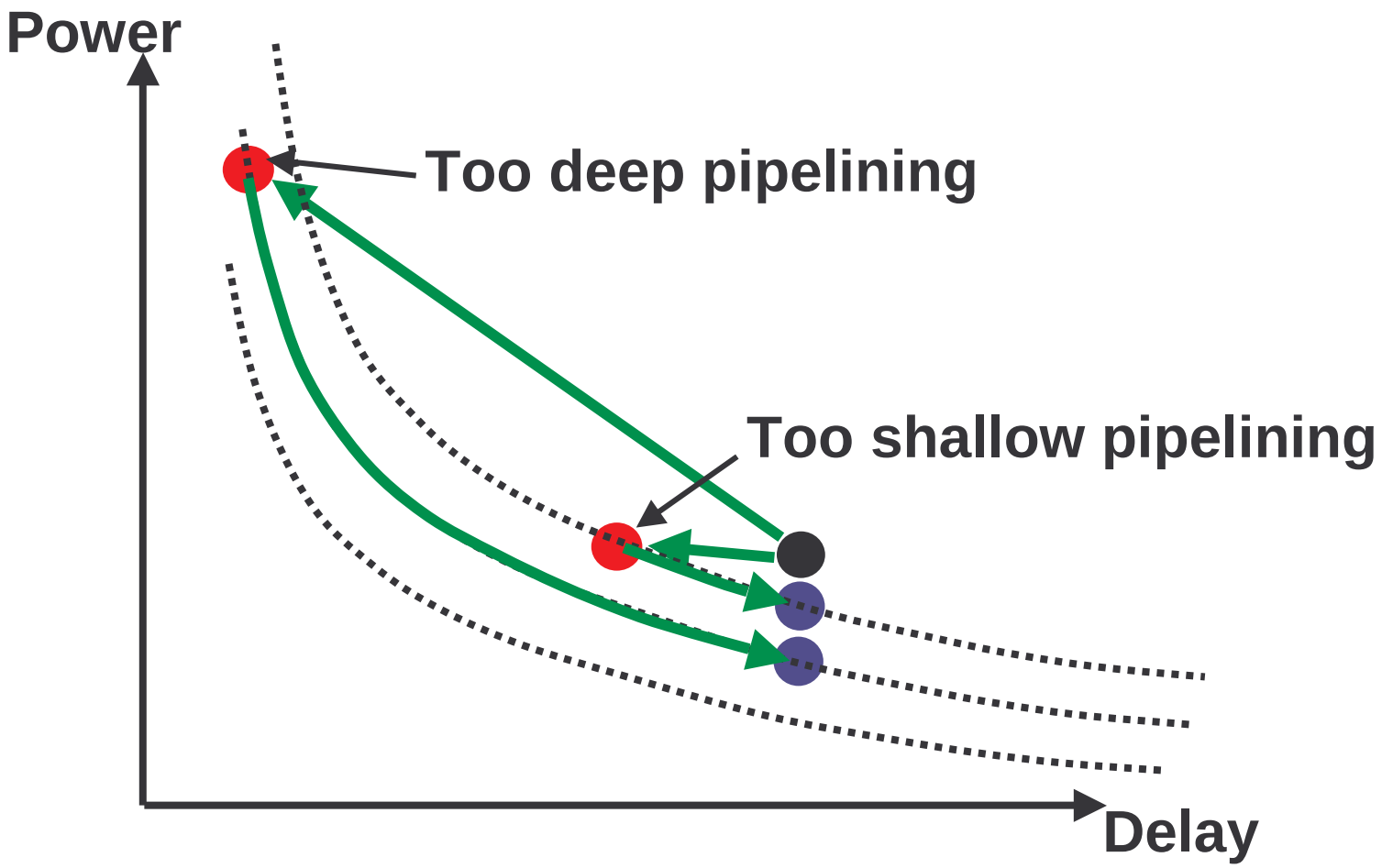
Power-Optimal Pipelining

- **Power reduction from pipelining limited by power overhead of increased number of flip-flops**
  → Power-Optimal Pipelining



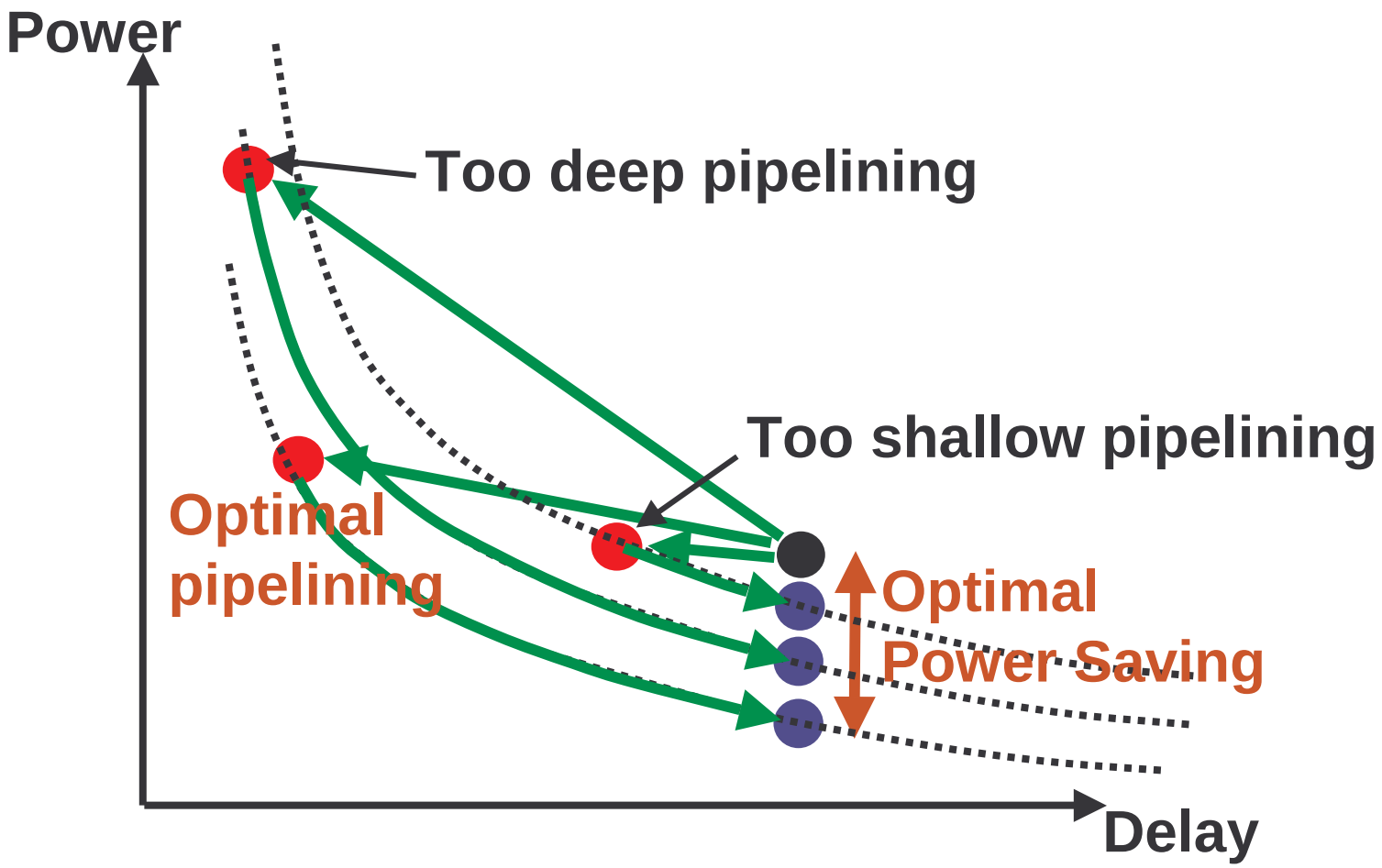
Power-Optimal Pipelining

- **Power reduction from pipelining limited by power overhead of increased number of flip-flops**
  → Power-Optimal Pipelining



Power-Optimal Pipelining

- **Power reduction from pipelining limited by power overhead of increased number of flip-flops**
  → Power-Optimal Pipelining



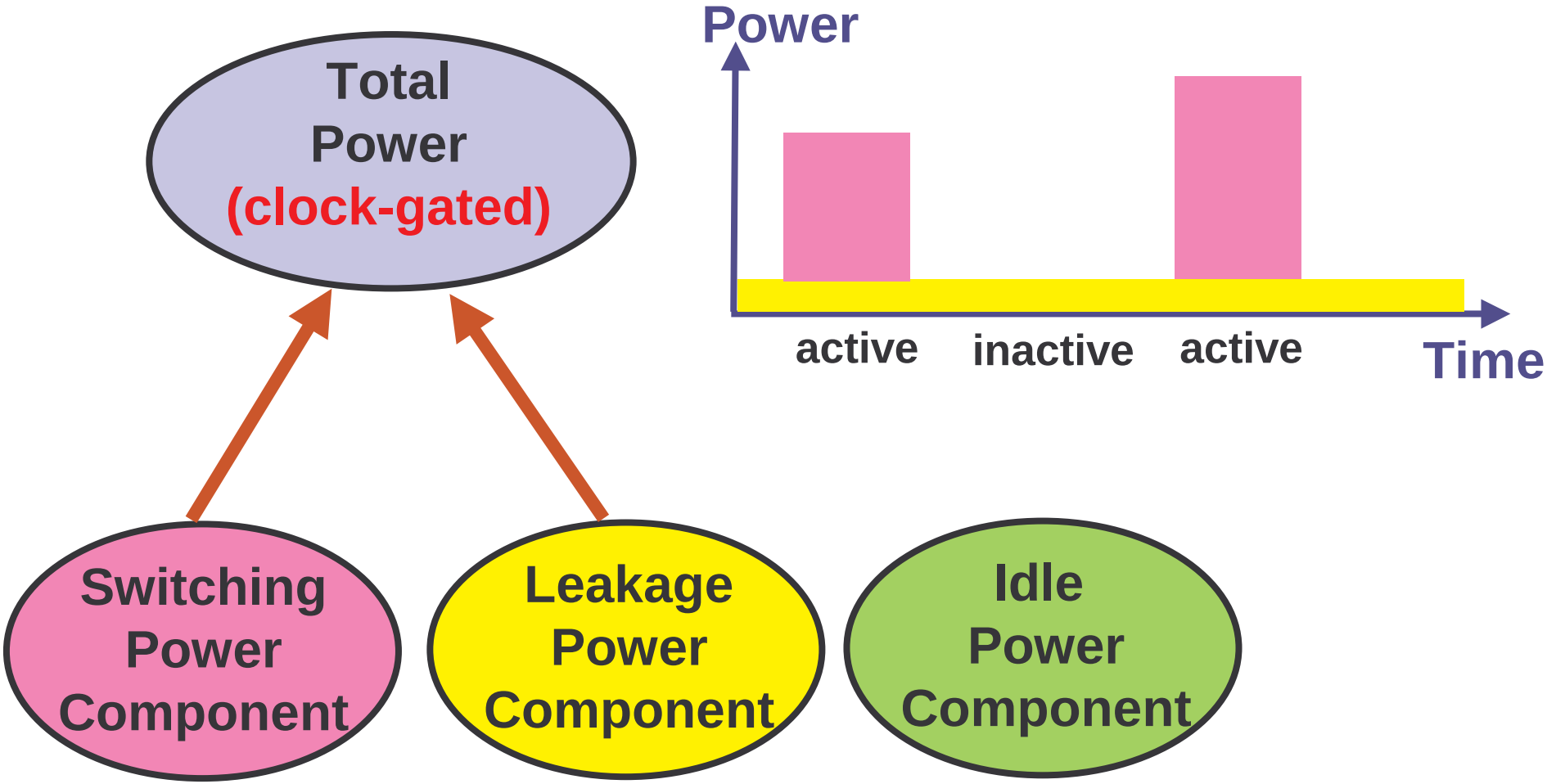
Contribution

- **Pipelining is an old idea.**
- **Research focus has been on performance impact of pipelining.**
- **Idea of using pipelining [Chandrakasan ’92] to lower power has not been fully explored in deep submicron technology.**

- **Analysis and circuit-level simulation of Power-Optimal Pipelining for different regimes of V_{th}, activity factor, clock gating**

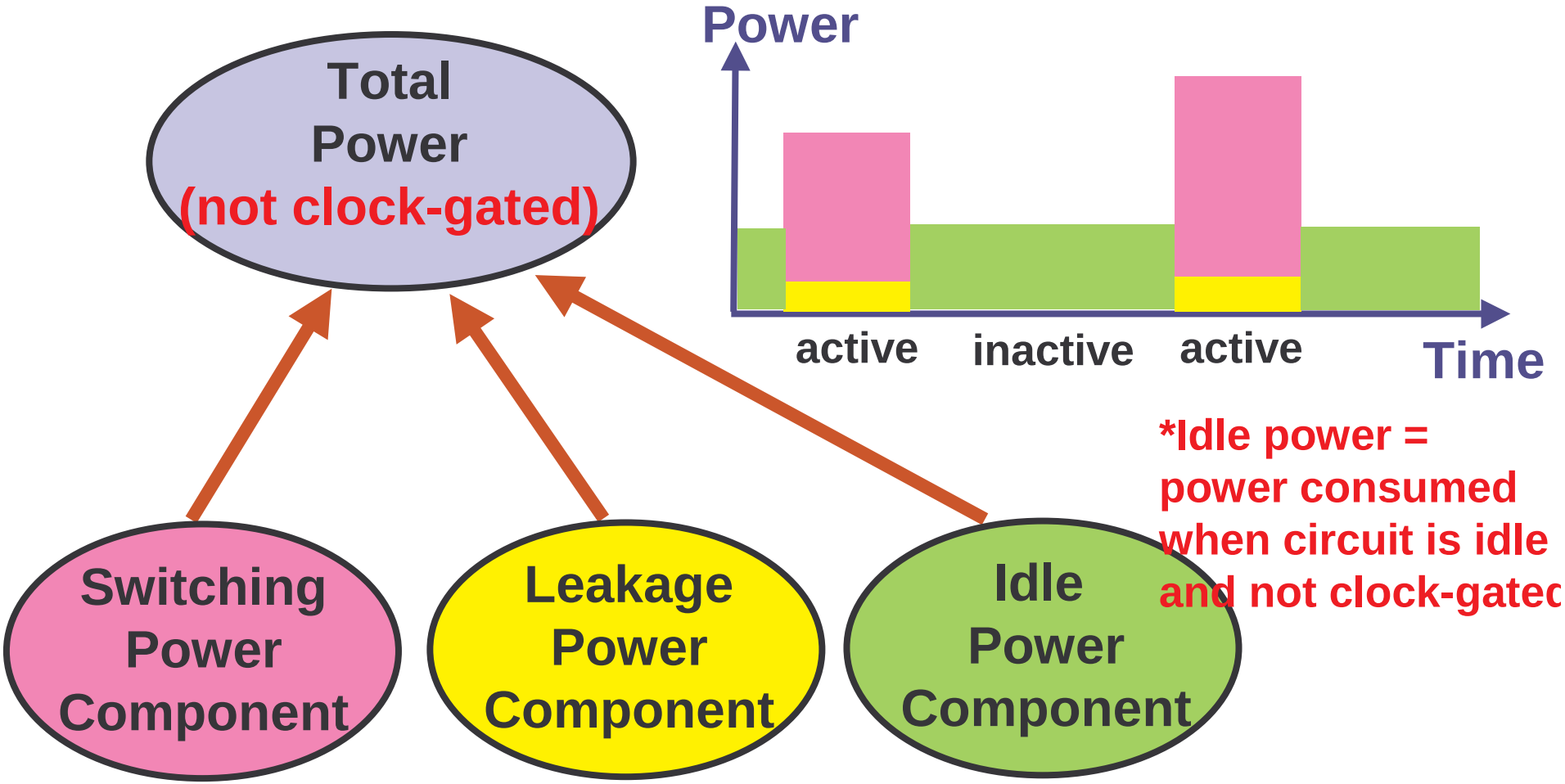
Bottom-to-Top Approach

1. Impact of pipelining on power component
2. Impact of pipelining on total power (with/without clock-gating)



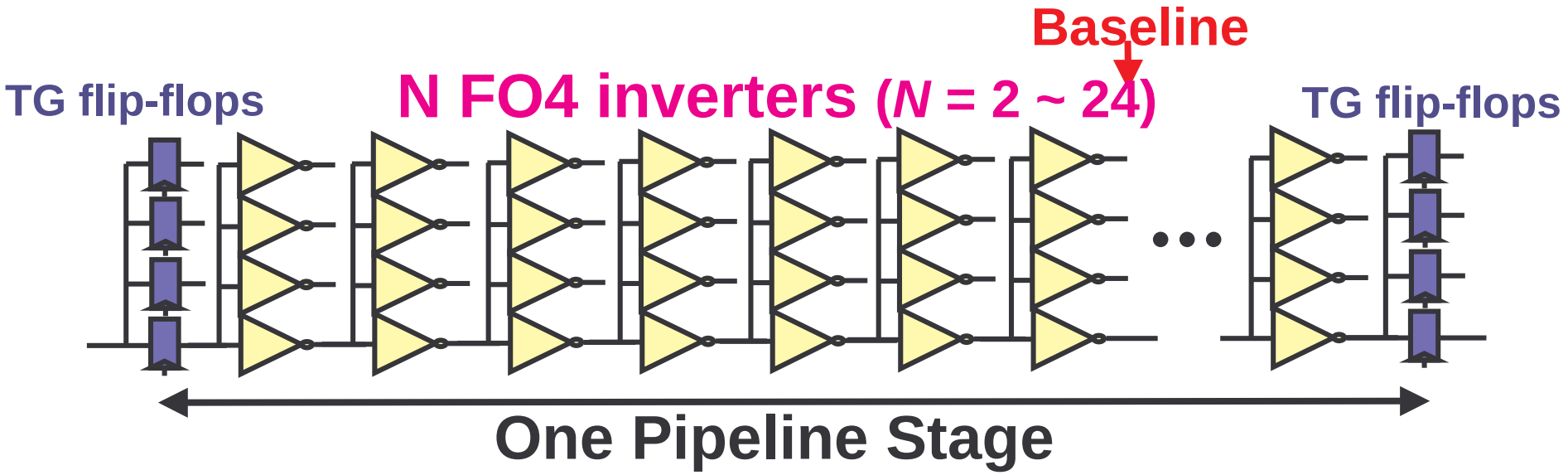
Bottom-to-Top Approach

1. Impact of pipelining on power component
2. Impact of pipelining on total power (with/without clock-gating)

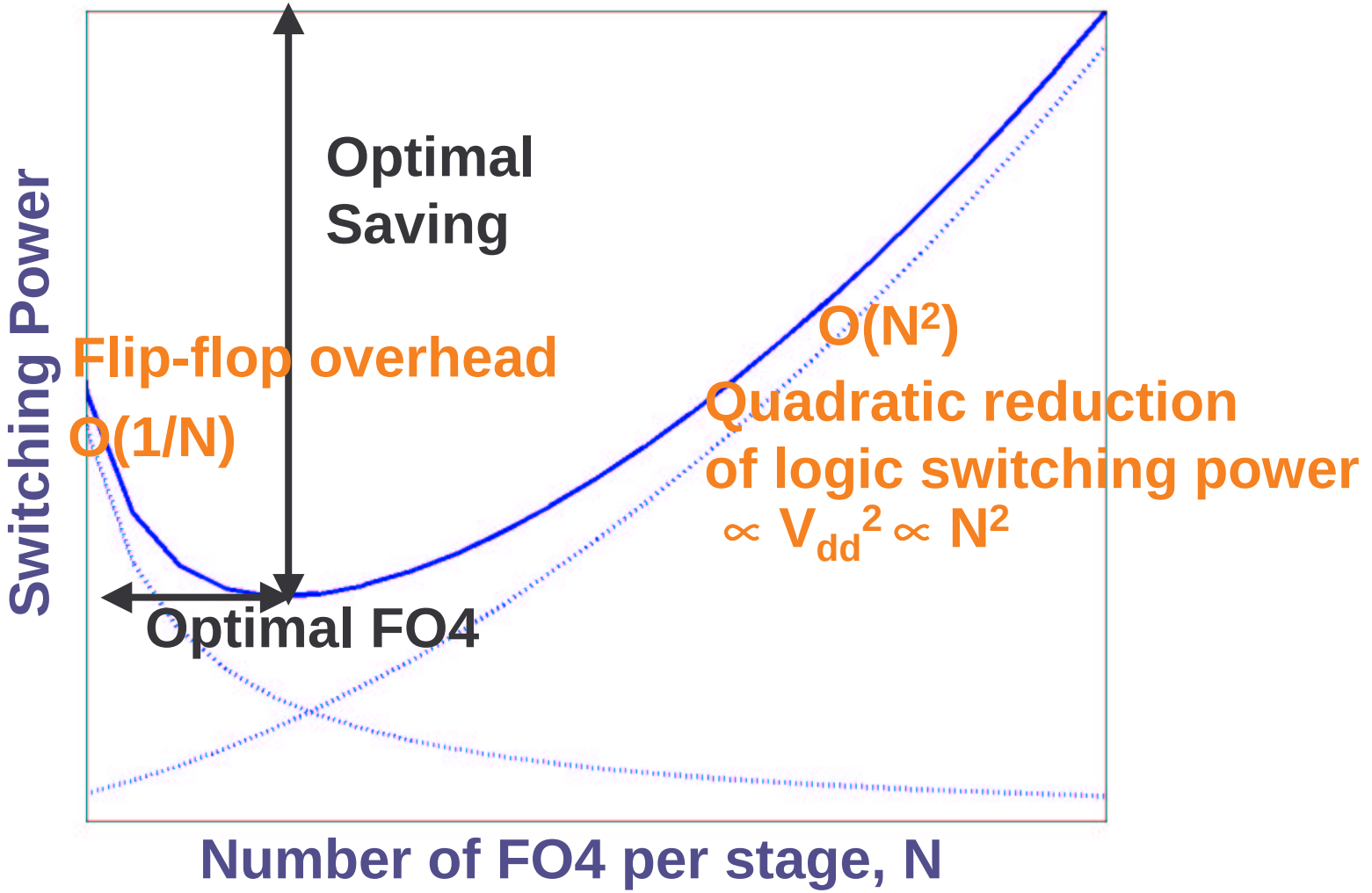


Methodology

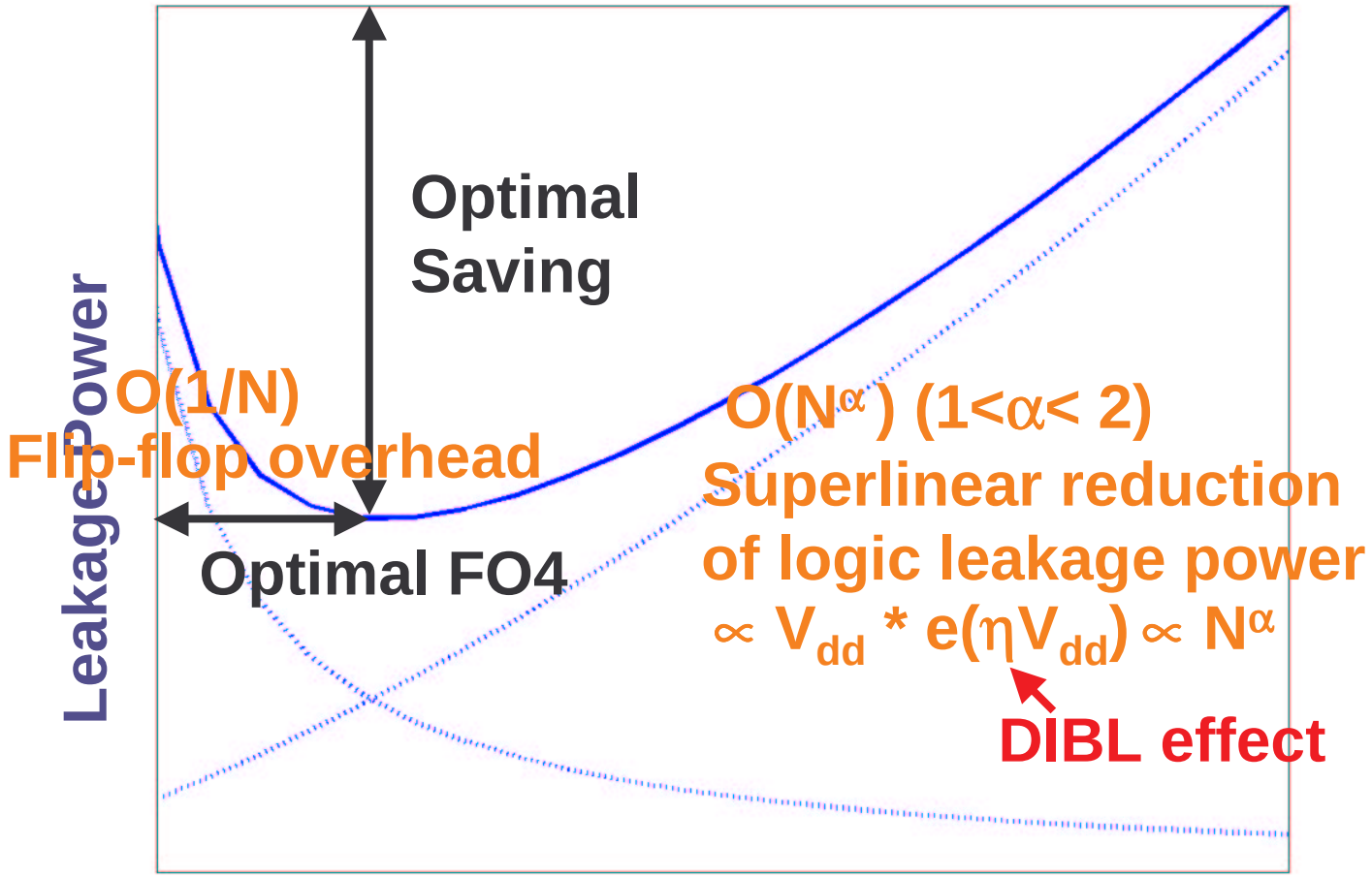
- **Target digital system: Fixed throughput, Highly parallel computation, Logic-dominant**
- **Test bench**
  - **BPTM (Berkeley Predictive Technology Model) 70nm process:**
  - **LVT(0.17/-0.2), MVT(0.19/-0.22), HVT(0.21/-0.24)**
  - **Hspice simulation at 100°C, Clock = 2 GHz**



Pipelining and Switching Power: Analytical Trend



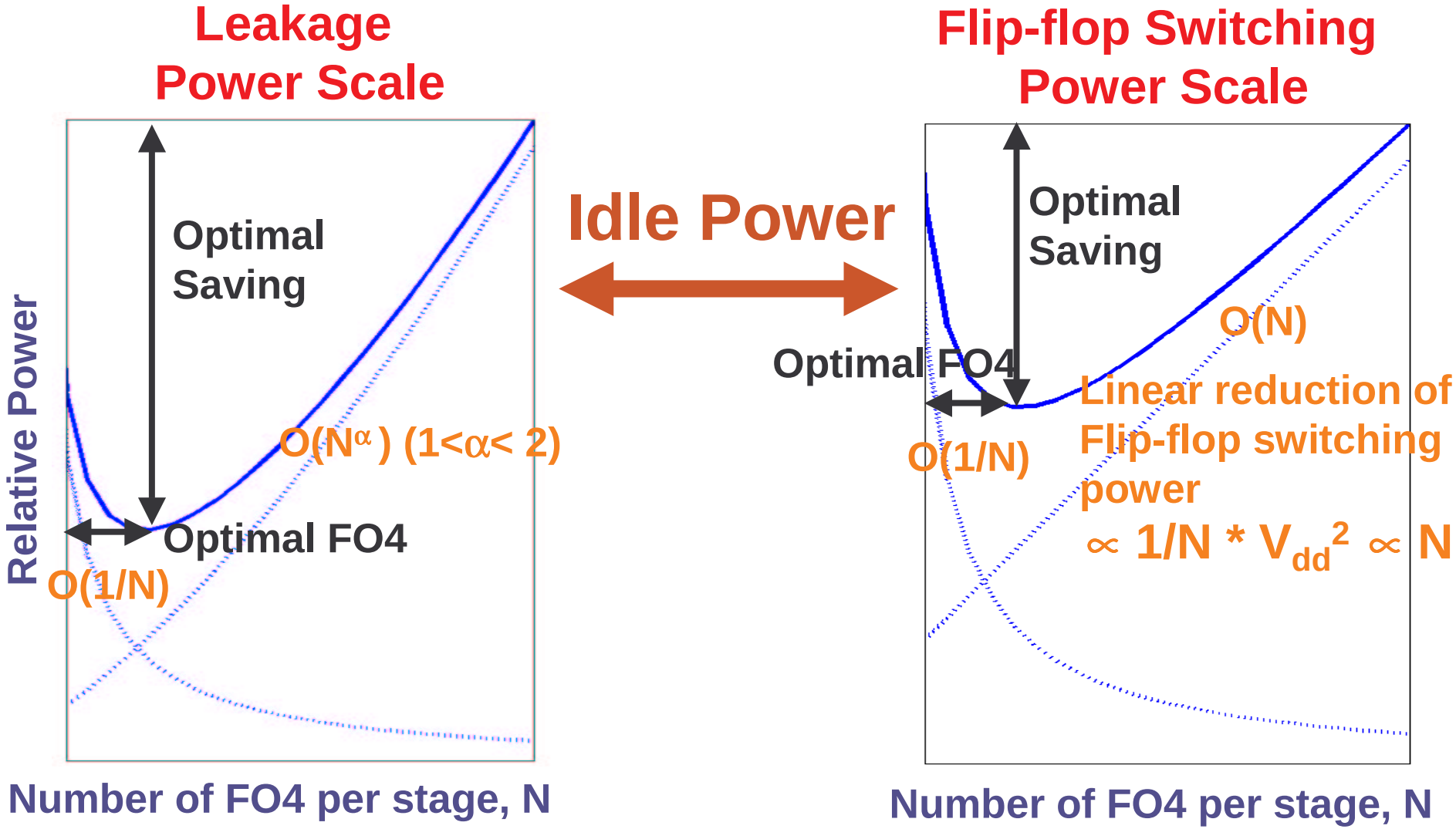
Pipelining and Leakage Power: Analytical Trend



Pipelining and Idle Power: Analytical Trend

- Clock-gating is not always possible
  - Increased control complexity
  - insufficient setup time of clock enable signal
- Leakage Power + Flip-flop Switching Power
  - Between leakage power scaling and flip-flop switching power scaling depending on leakage level

Pipelining and Idle Power: Analytical Trend



Simulation Results: Power Components

Fixed Throughput @ 2 GHz

Power Components	Switching Power	Leakage Power	Idle Power
Right hand side curve	$O(N^2)$	$O(N^{\alpha})$ $(1<\alpha< 2)$	$O(N)$ or $O(N^{\alpha})$ $(1<\alpha< 2)$
Saving*	79(HVT)~ 82(LVT)%	70(LVT)~ 75(HVT)%	55(HVT)~ 70(LVT)%
N*	6	6	8

N = Number of FO4 inverters per stage

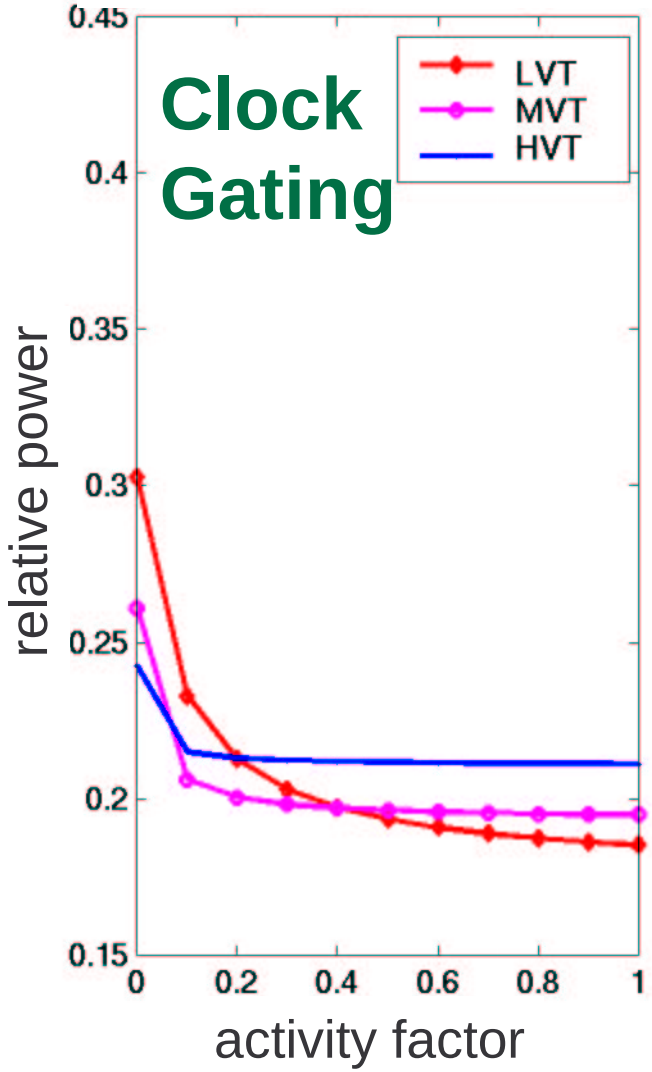
(Not including flip-flop delay)

N* = Optimal N

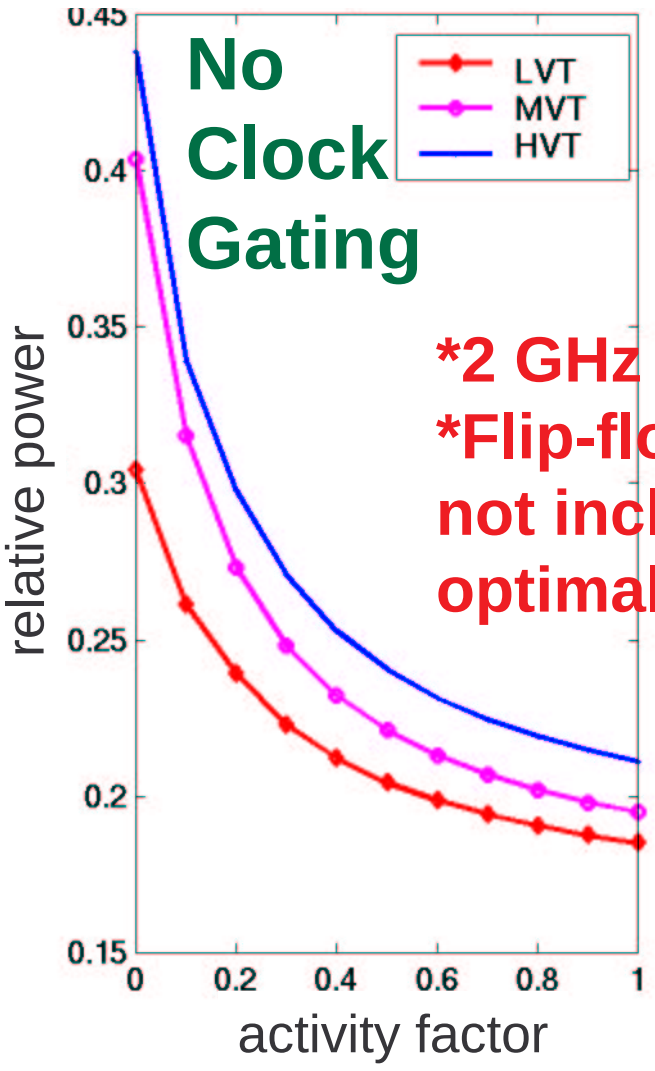
Saving* = Optimal power saving by pipelining

Optimal Power Saving

Optimal FO4 = 6



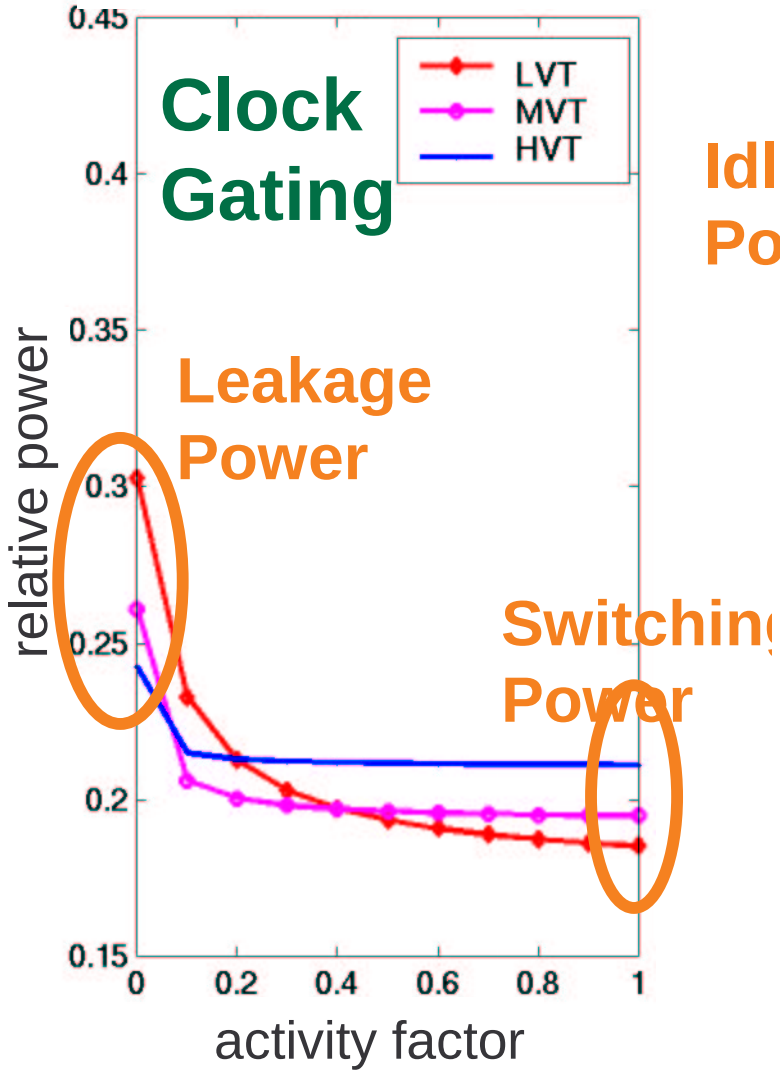
Optimal FO4 = 6~8



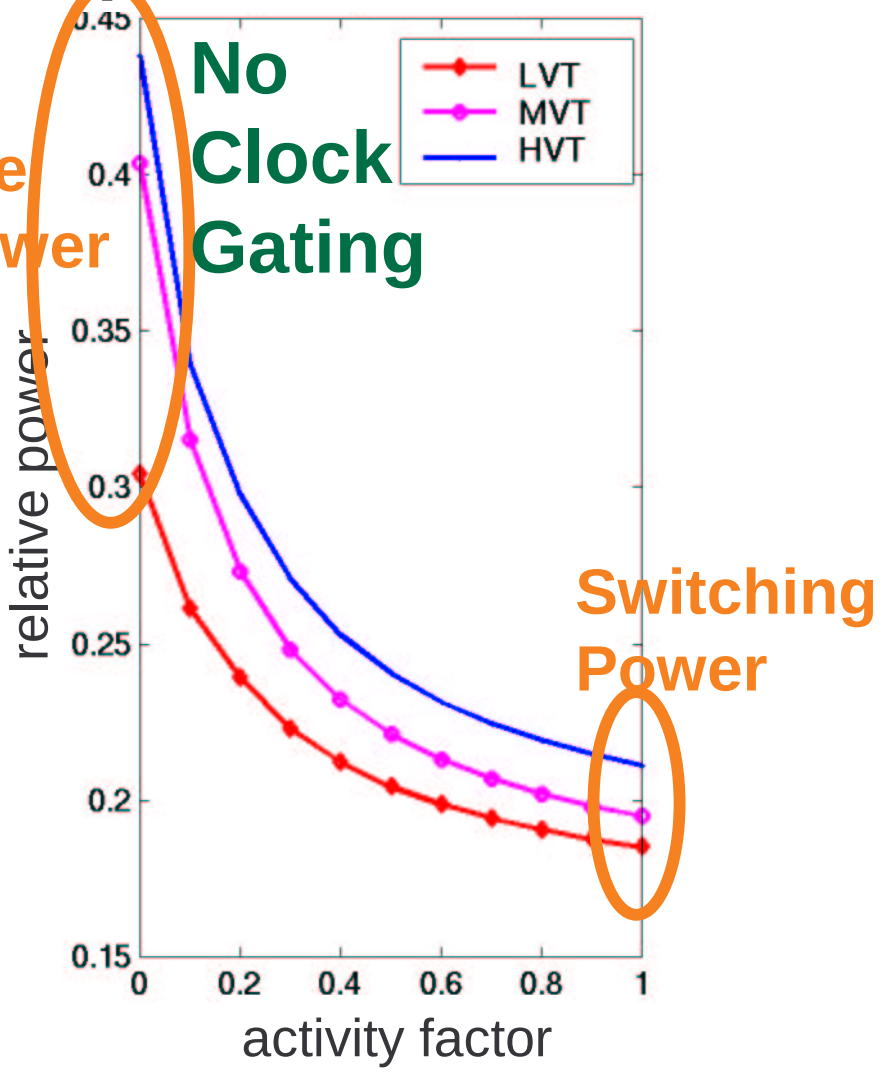
*2 GHz
*Flip-flop delay not included in optimal FO4

Optimal Power Saving

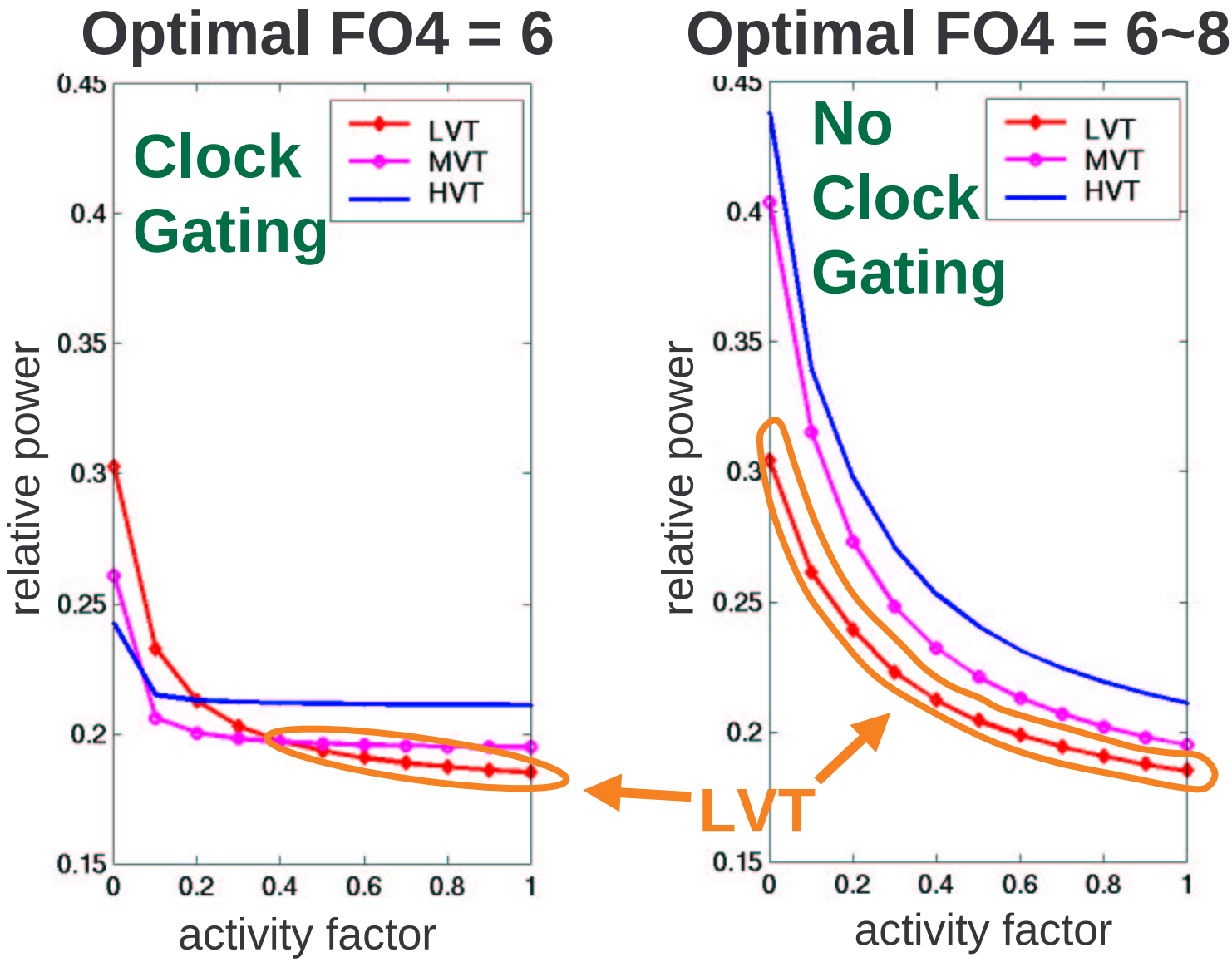
Optimal FO4 = 6



Optimal FO4 = 6~8



Optimal Power Saving



Discussion

- LVT can be fast and power-efficient
  - enables lower V_{dd}
- Flip-flop delay more important than flip-flop power for power-optimal pipelining

Limitation of This Work

	Effect on optimal logic depth	Effect on optimal power saving
Super-linear growth of flip-flops	↑	↓
Additional memory	↑	↓
Reduced glitches	↓	↑
Parasitic wire capacitance	↑	↓

Conclusion

- Pipelining is an effective low-power tool when used to support voltage scaling in digital system implementing highly parallel computation.
- Optimal Logic Depth: 6-8 FO4
  - ~ 8-10 FO4 including flip-flop delay
- Optimal Power Saving: 55 – 80%
  - It depends on V_{th}, AF, Clock-Gating
- Insights:
  - Pipelining is more effective with High AF
    - Pipelining is most effective at saving switching power
  - Pipelining is more effective with lower V_{th}
    - Except for when leakage power is dominant.
  - Pipelining is more effective with clock-gating
    - reduced flip-flop overhead.

Acknowledgments

- **Thanks to SCALE group members and anonymous reviewers**

- **Funded by NSF CAREER award CCR-0093354, NSF ITR award CCR-0219545, and a donation from Intel Corporation.**

BACKUP SLIDES