The Return of Silicon Efficiency

Simon Knowles
Icera

simon@icerasemi.com

AGENDA

- Chip Market
- Design Objectives
- Modern Transistors
- Optimality of Designs
- The Design Hull
- Implications for Design

simon@icerasemi.com

“The Cheap will Outsell the Good”

Customer Priorities

1. Availability
2. Price
3. Features, Power, Size

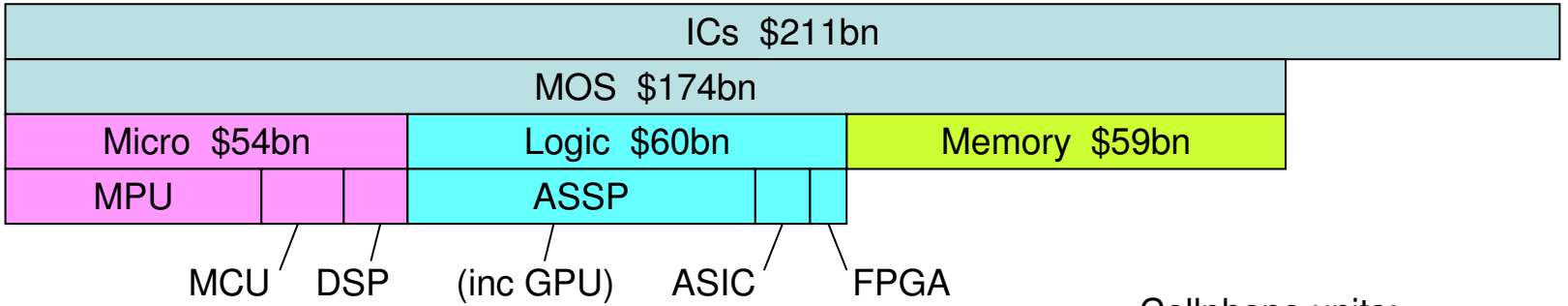
Design Choices

1. Hardware ↔ Software
2. Process Technology
3. MHz and Volts
4. Custom ↔ Synthesized

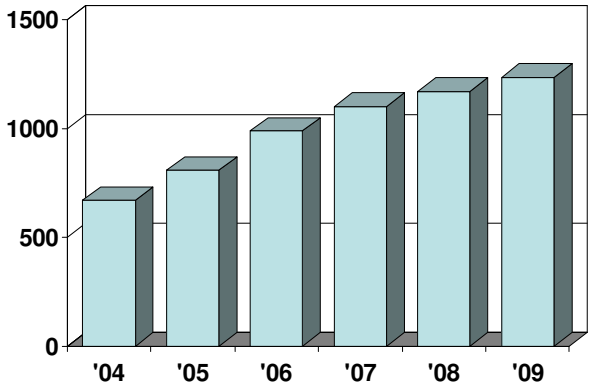
simon@icerasemi.com

The Chip Market

SIA 2006



Cellphone units:



Most logic chips are processors or processor-centric.

Computers have driven MOS chip evolution – the first "product superclass".

The second product superclass will be "cellphones"…

- Consumer driven – more cost sensitive than computers.
- High mobility – more power sensitive than computers.
- High (specialized) performance – eg. HSDPA cellular modem ~50GOPs.

simon@icerasemi.com

What's New(-ish) in Logic Chip Physics?

Small dopant populations (V_T variability)

Subthreshold leakage (V_T reduction is constrained, so Vdd likewise)

Gate leakage (end of SiO_2 T_{ox} scaling)

Channel Stress (engineered and STI-induced)

OPC rules

Well Proximity Effect

Line Edge Roughness (gate and wire)

NBTI (RAM V_T drift)

All important for designers, but not (directly) the main influencers of design evolution

simon@icerasemi.com

Top 2 Design Influencers

1. **Most power-sensitive designs will set their own supply voltage.**
   - Design at typical, turn manufacturing speed spread (functionality risk) into power efficiency spread (battery life).
   - Offset manufacturing leakage spread against speed spread.
   - Dynamically trade performance *vs* power.
   - Disconnect power in standby (efficient regulators are switch-mode).
2. **Processors are evolving to displace fixed-function hardware**
   - Target multiple market sockets per design.
   - Hardware-efficient (dynamic re-purposing of resources).
   - Mitigation of specification uncertainty.
   - Work around bugs by software change.
   - Well-practiced idiom for speed.

simon@icerasemi.com

Modern Transistors

simon@icerasemi.com

Modern Transistors

ITRS distinguishes 3 roadmaps…

- High Performance (HP)
- Low Operating Power (LOP)
- Low Standby Power (LSTP)

TSMC offers 12 logic T's at 65nm…

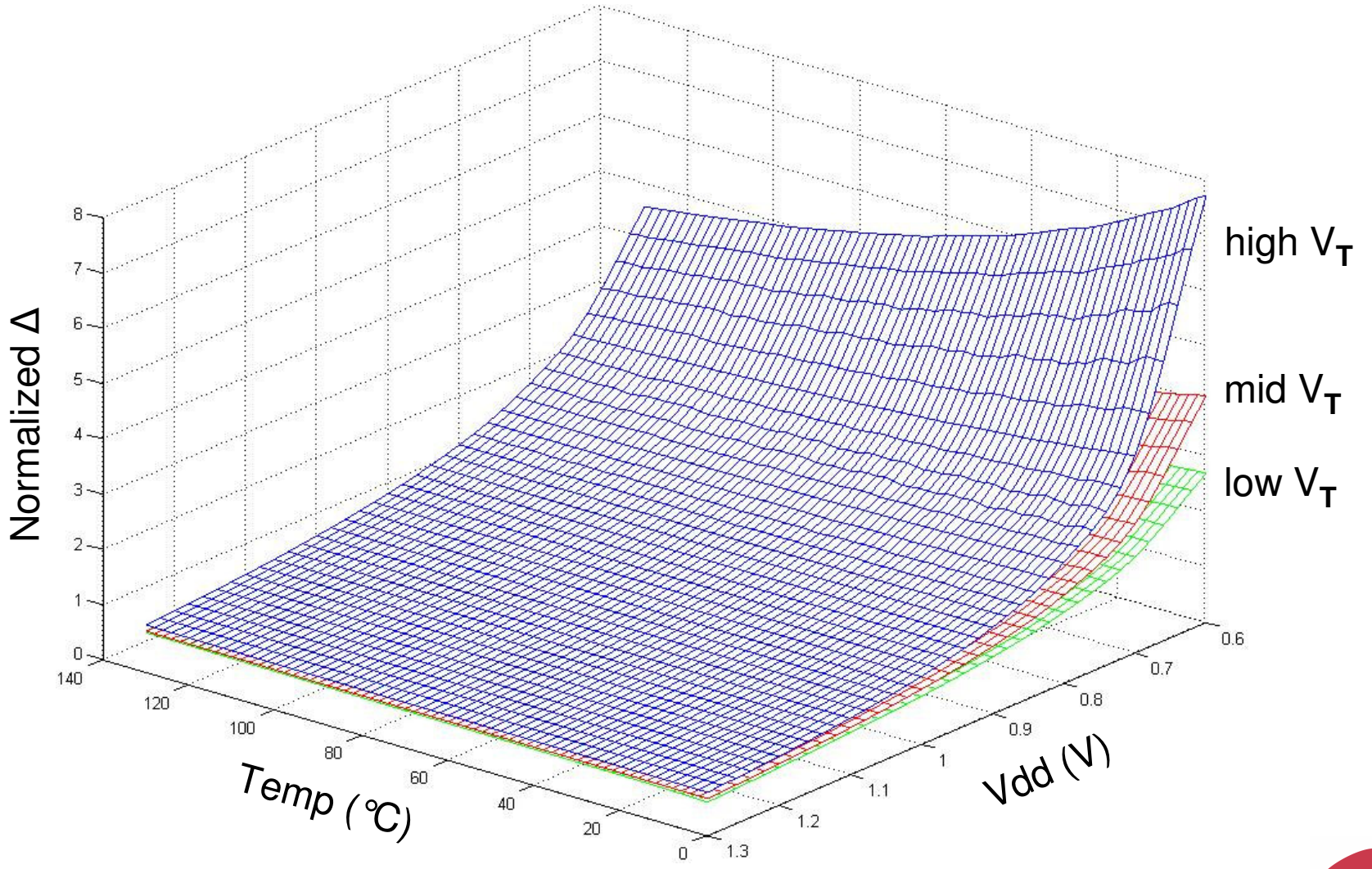
- 2 x HP
- 6 x LOP
- 4 x LSTP

ITRS2005 for 65nm 2007

		L_g (nm)	T_{ox} (nm)	V_T (mV)
computers	HP	25	1.1	165
"cellphones"	LOP	32	1.2	285
	LSTP	45	1.9	524

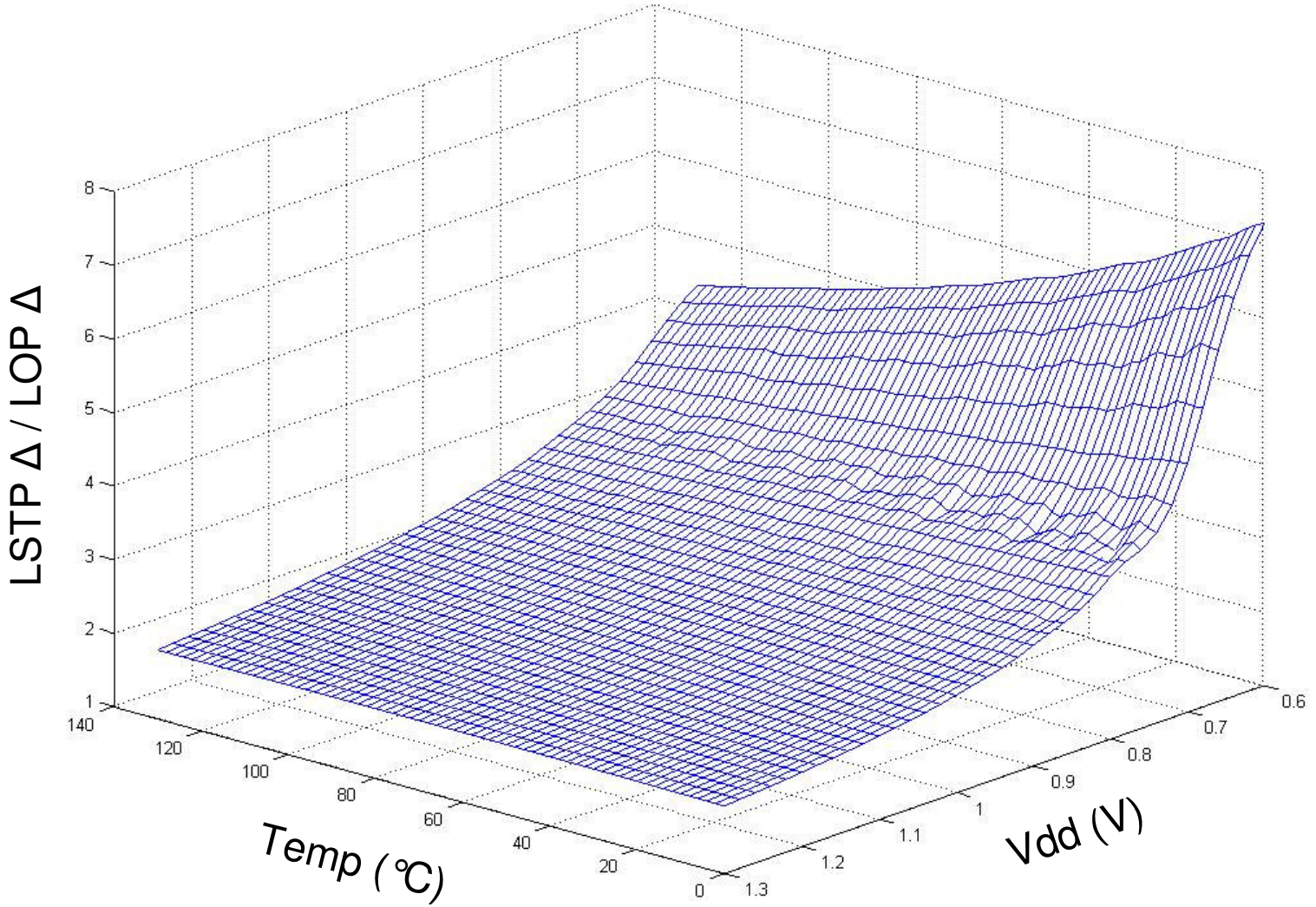
simon@icerasemi.com

65nm LOP Delay



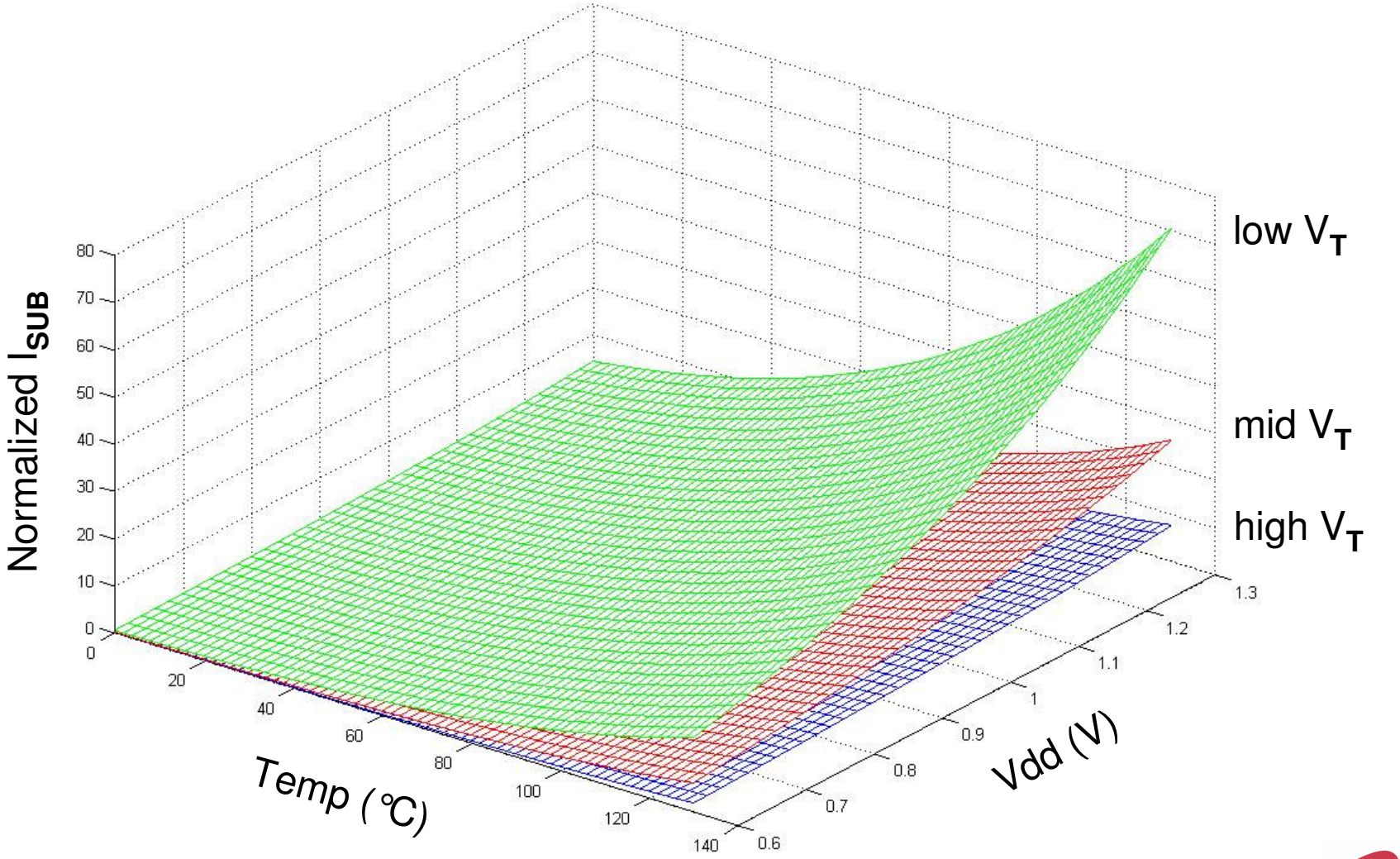
simon@icerasemi.com

65nm LSTP vs LOP Delay



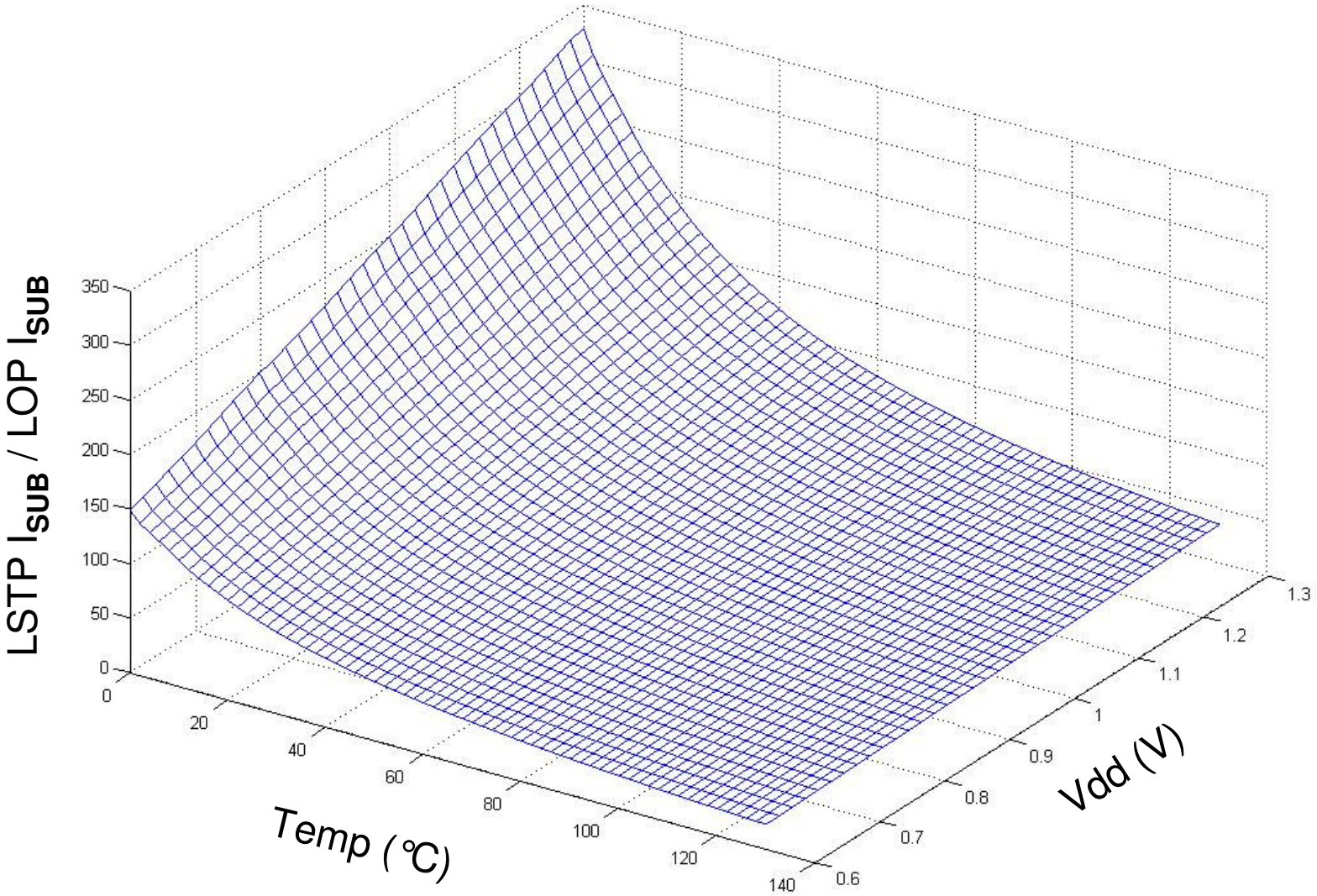
simon@icerasemi.com

65nm LOP Sub-Threshold Leakage



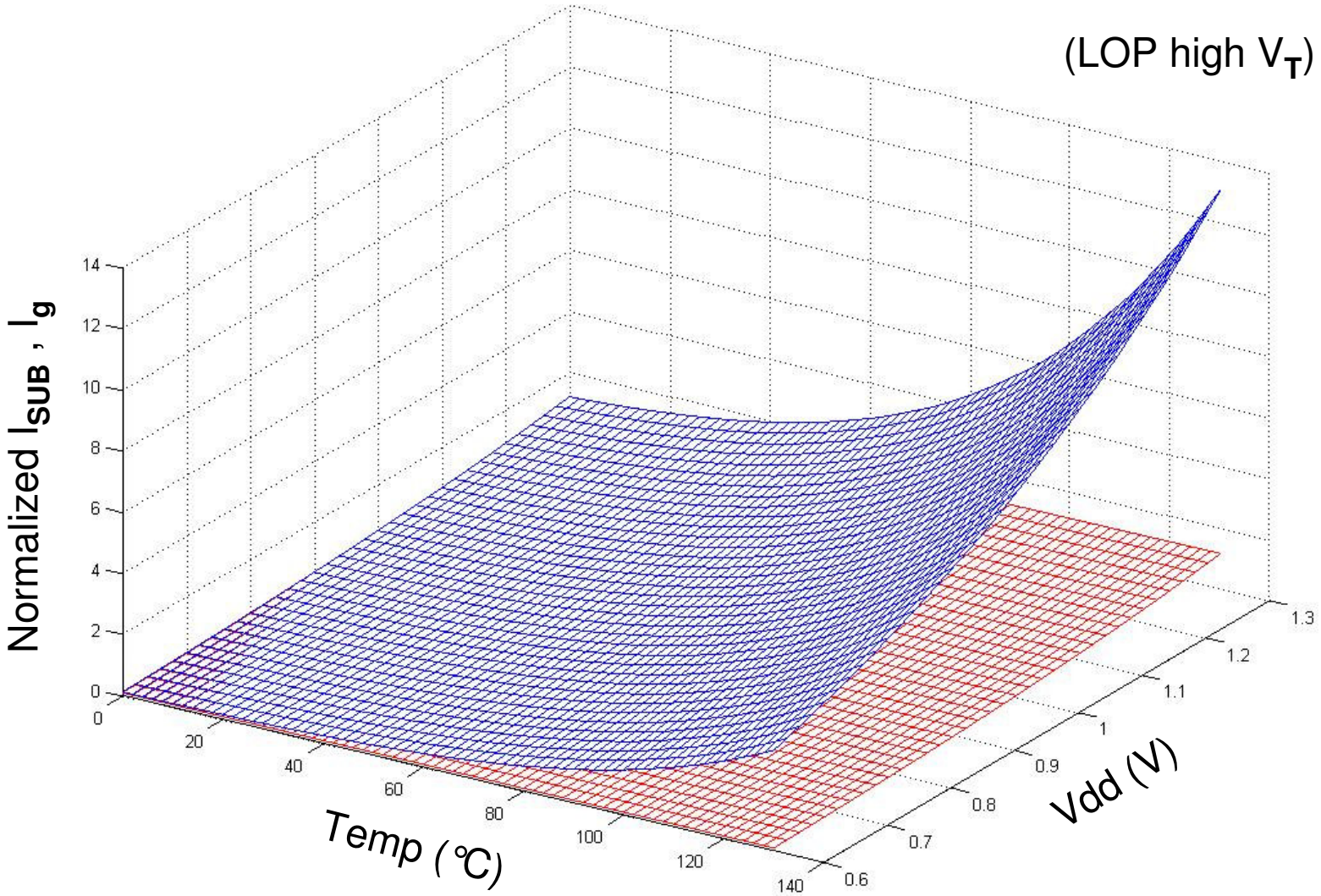
simon@icerasemi.com

65nm LSTP vs LOP Sub-Threshold Leakage



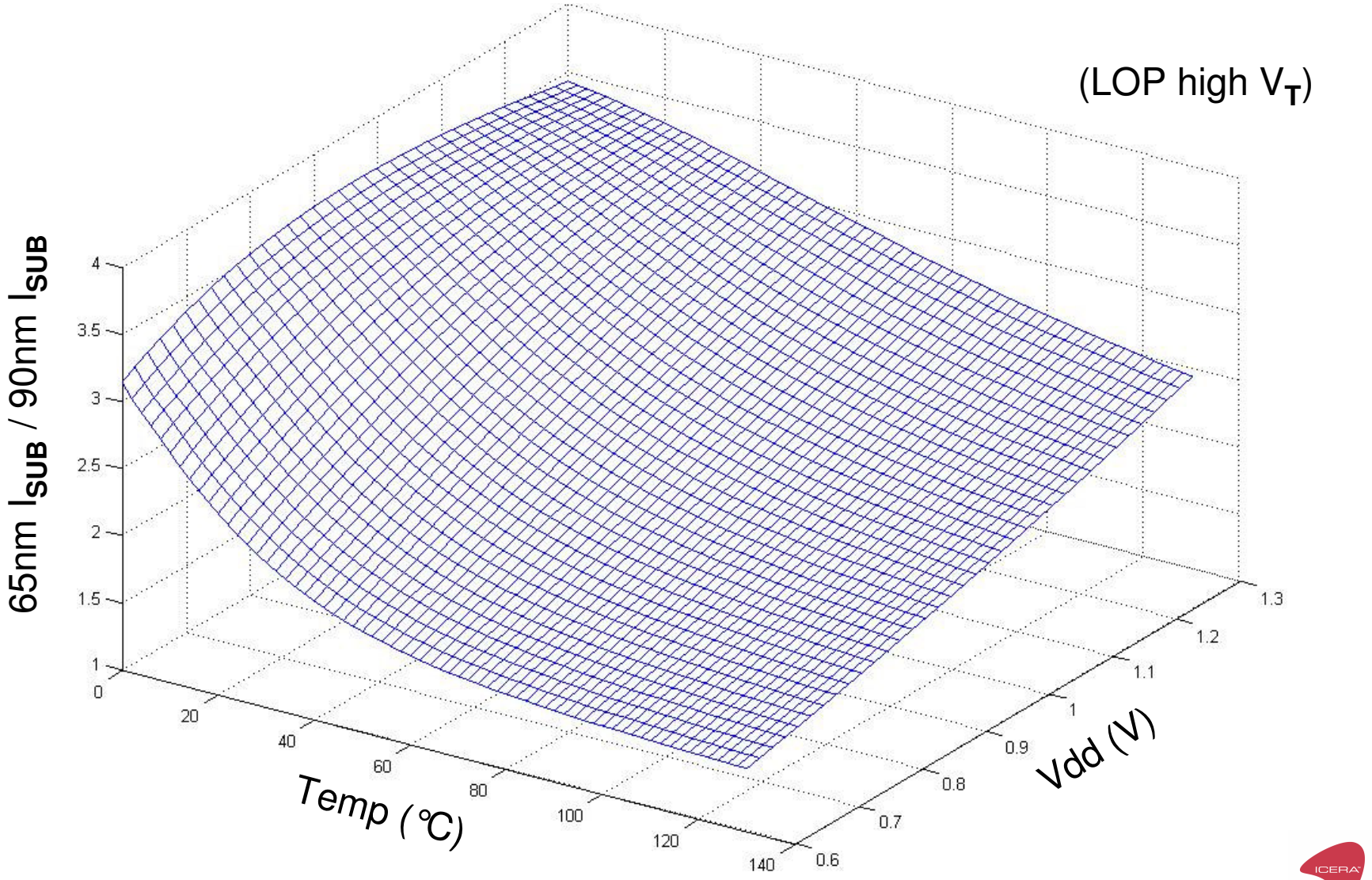
simon@icerasemi.com

High Temp I_{SUB} Still Dominates I_g at 65nm (LOP and LSTP)



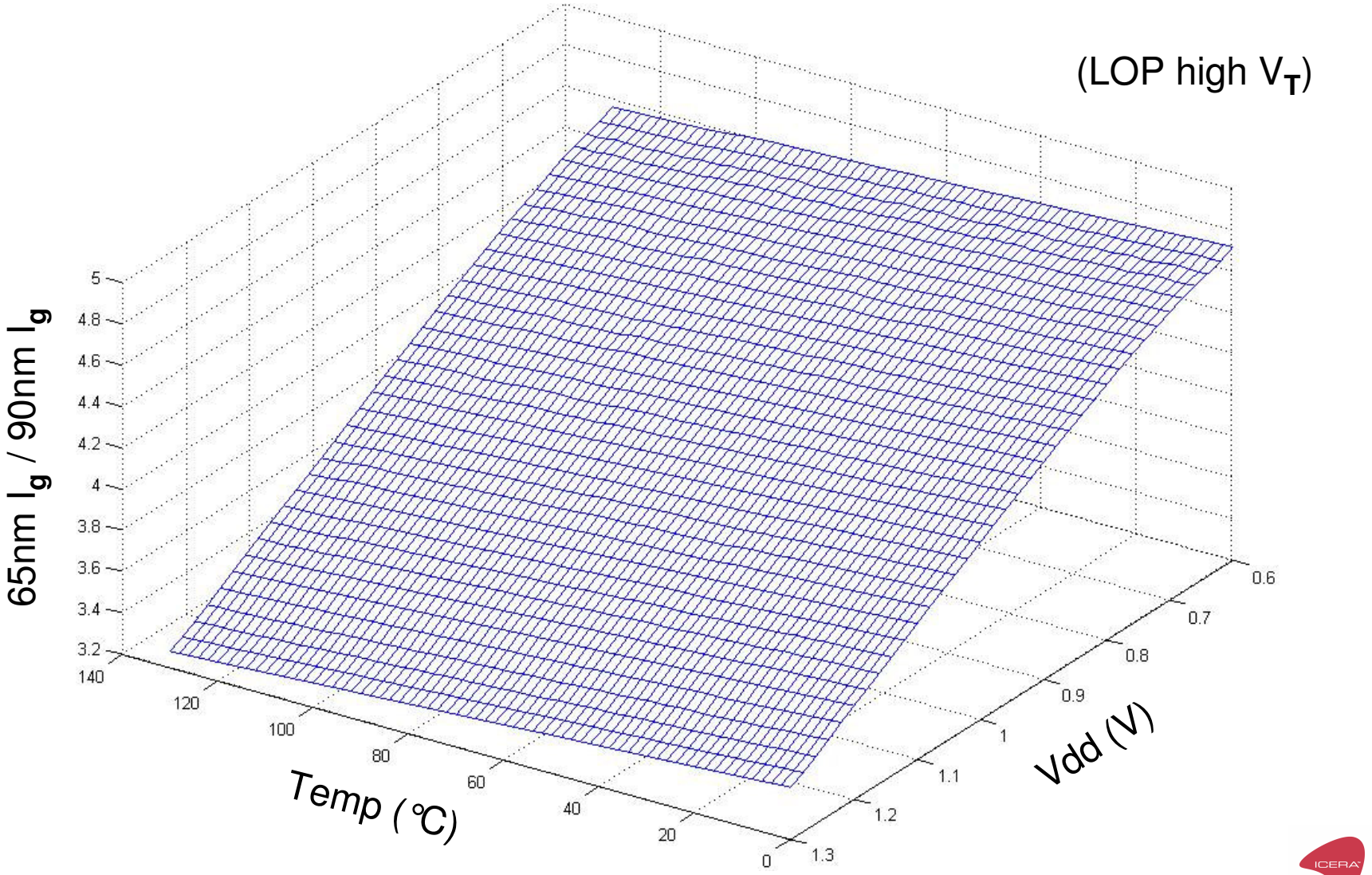
simon@icerasemi.com

LOP 65nm I_{SUB} vs 90nm I_{SUB}



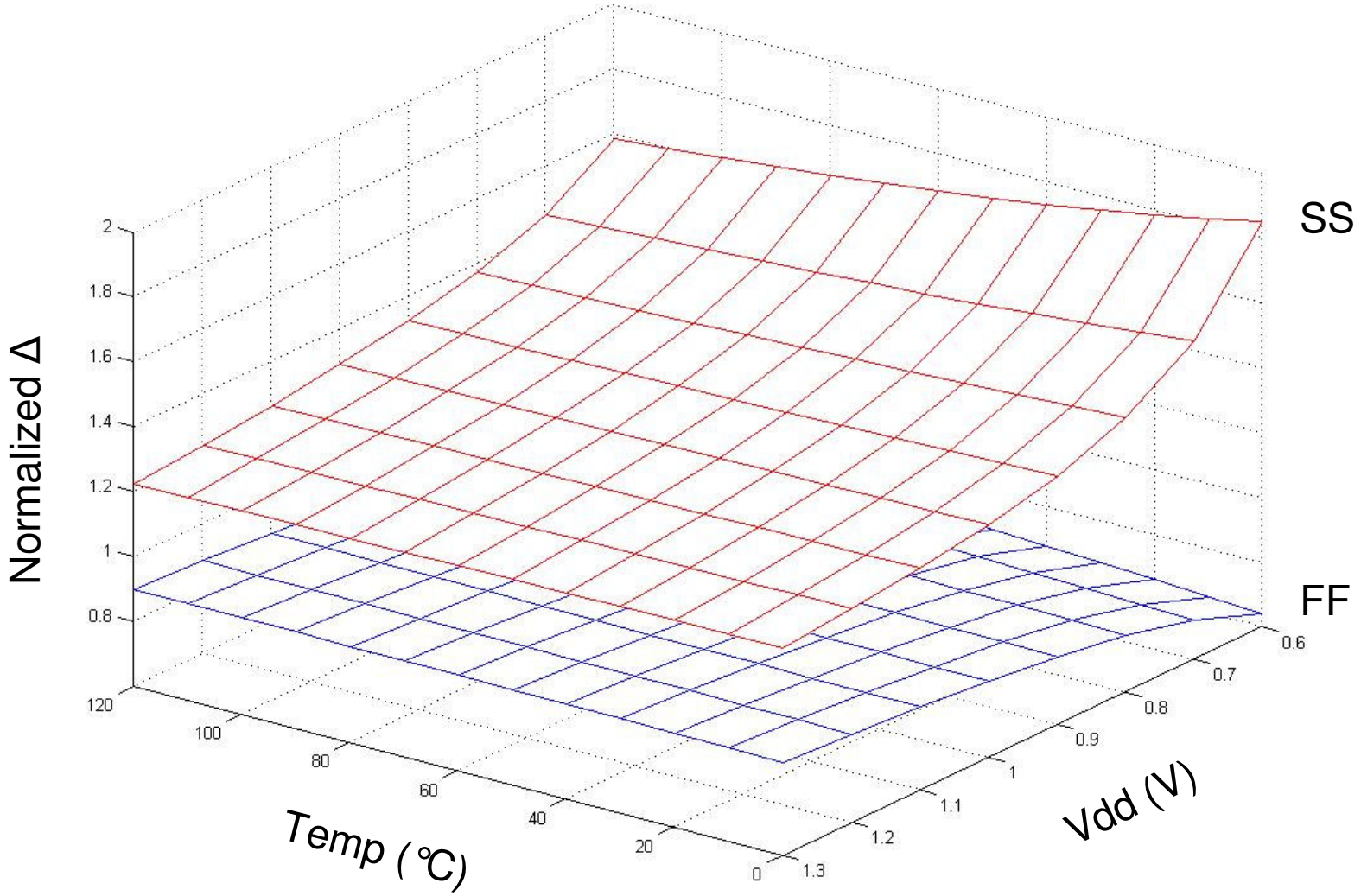
simon@icerasemi.com

LOP 65nm I_g vs 90nm I_g



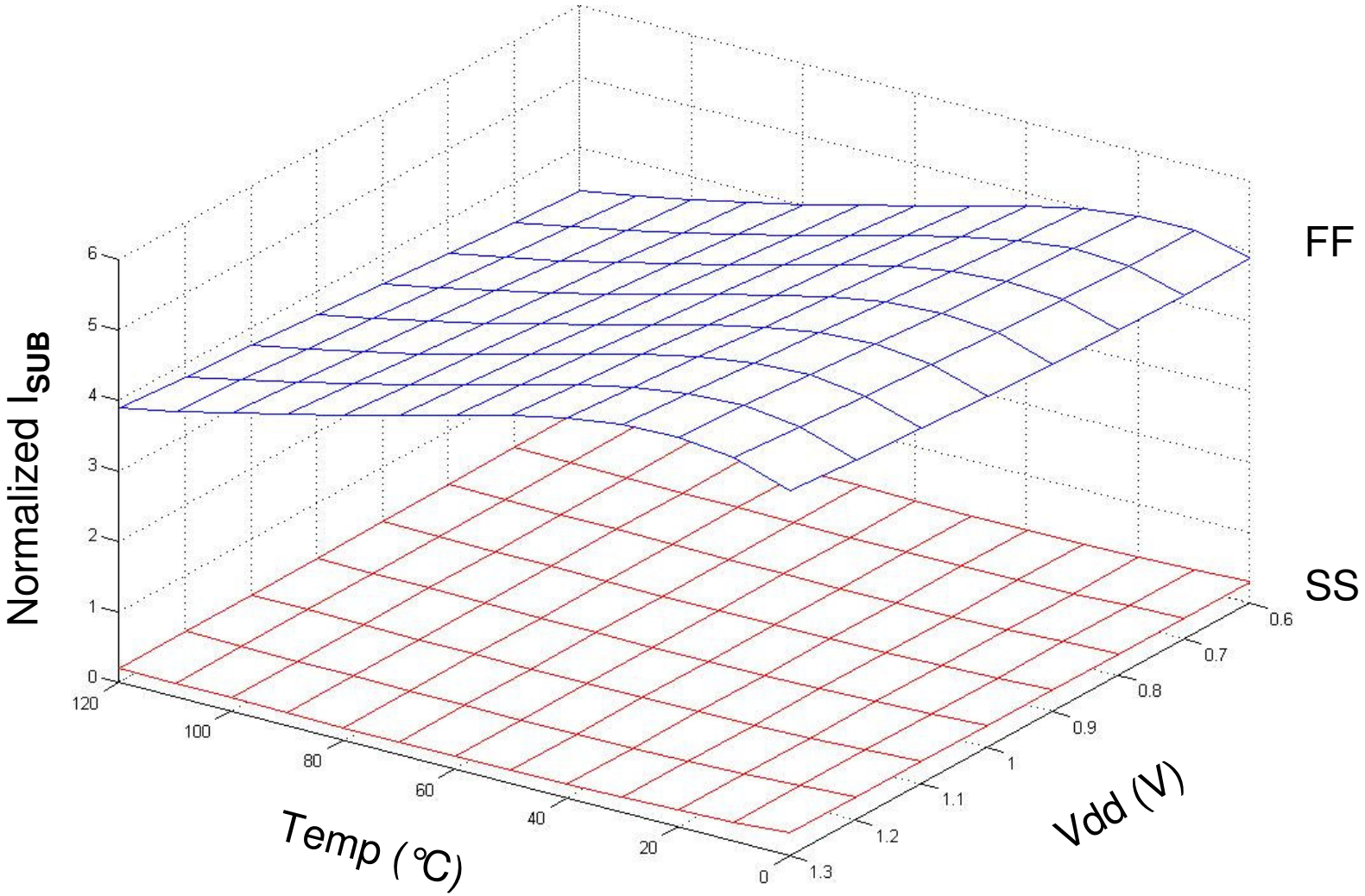
simon@icerasemi.com

Manufacturing Spread: 65nm mid V_T LOP Delay



simon@icerasemi.com

Manufacturing Spread: 65nm mid V_T LOP I_{SUB}



simon@icerasemi.com

Optimality of Design

simon@icerasemi.com

Power, Money and Performance

Excellent recent work on Power vs Performance (among many)…

- Srinivasan, Brooks, Gschwind, Bose, Zyuban, Strenski and Emma, “Optimimum Pipelines for Power and Performance”, Micro ’02.
- Zyuban and Strenski, “Unified Methodology for Resolving Power-Performance Tradeoffs at the Microarchitectural and Circuit Levels”, ISLPED ‘02.
- Harstein, Puzak, “Optimum Power/Performance Pipeline Depth”, Micro ’03.
- Dao, Zeydel, Oklobdzija, “Energy Optimization of Digital Pipelined Systems Using Circuit Sizing and Supply Scaling”, Trans. VLSI Systems ‘05.

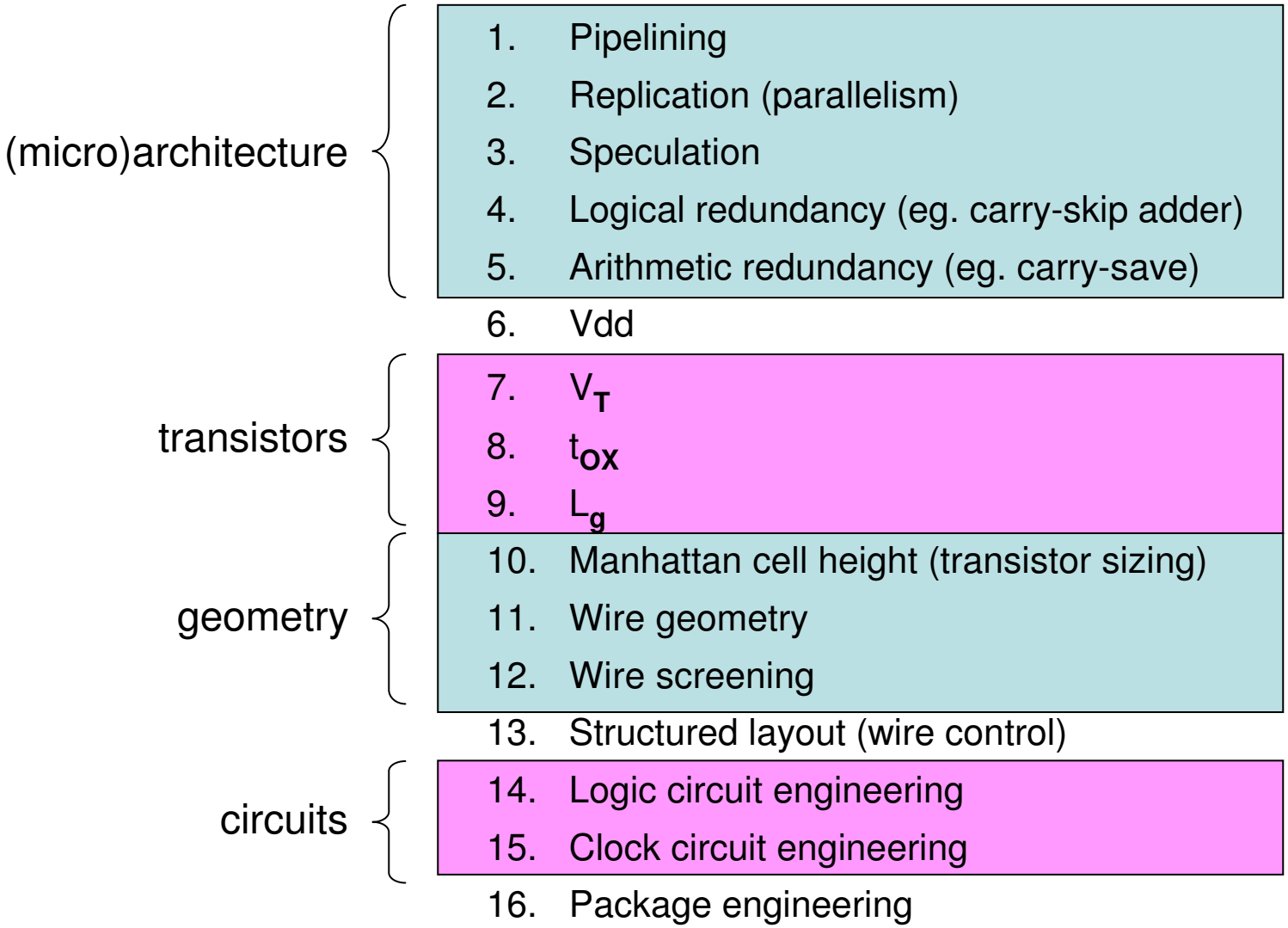
But outside the computer market, few chips sell on MHz.

Most designs set out to minimize cost (and power) at fixed performance…

- Decode an MPEG video without glitching.
- Execute an ADSL modem at 8Mbps.
- etc…

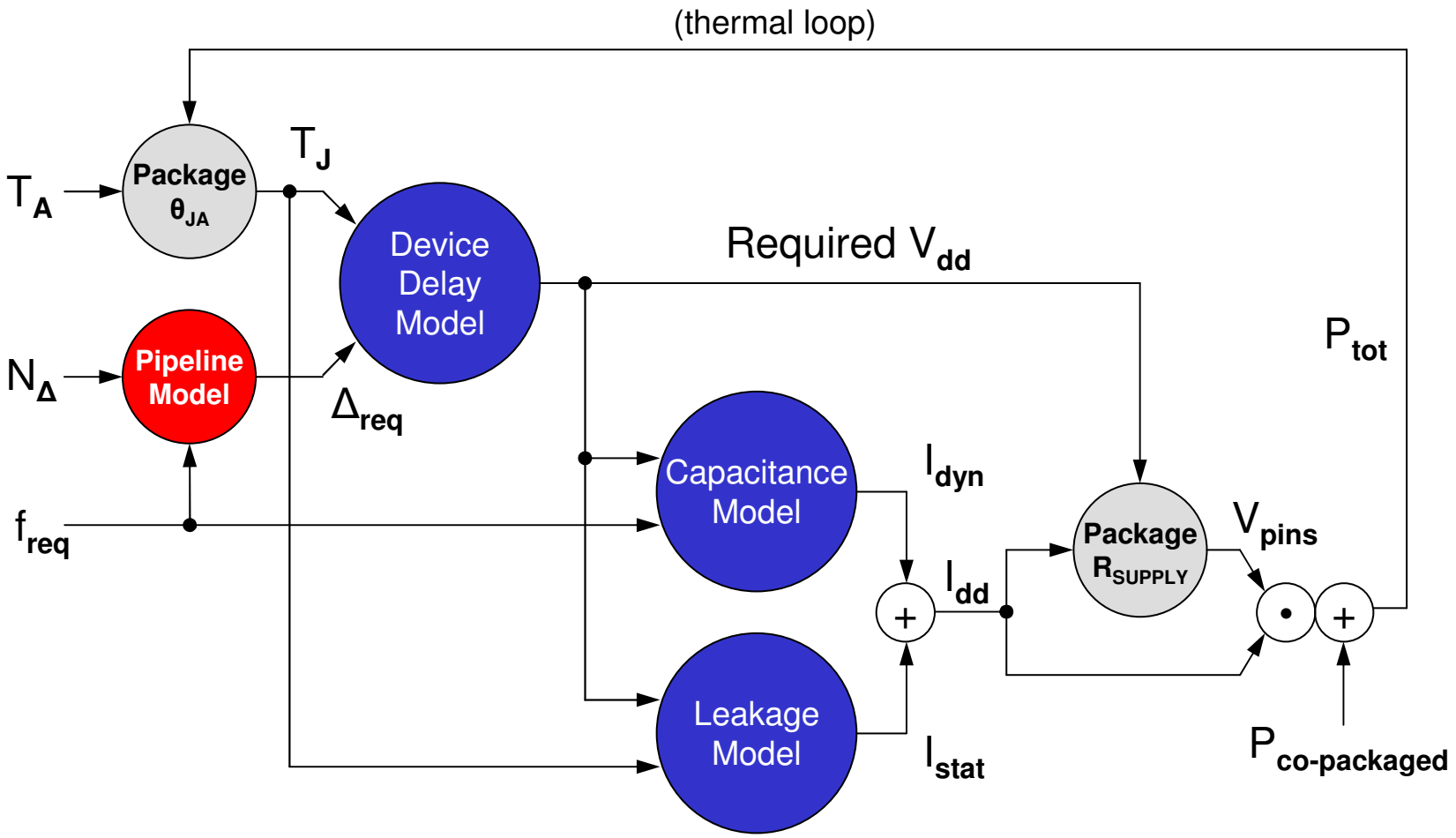
simon@icerasemi.com

Design Levers



simon@icerasemi.com

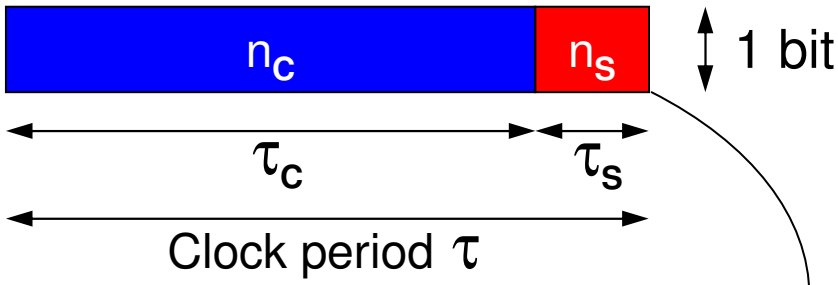
Electro-Thermal Model



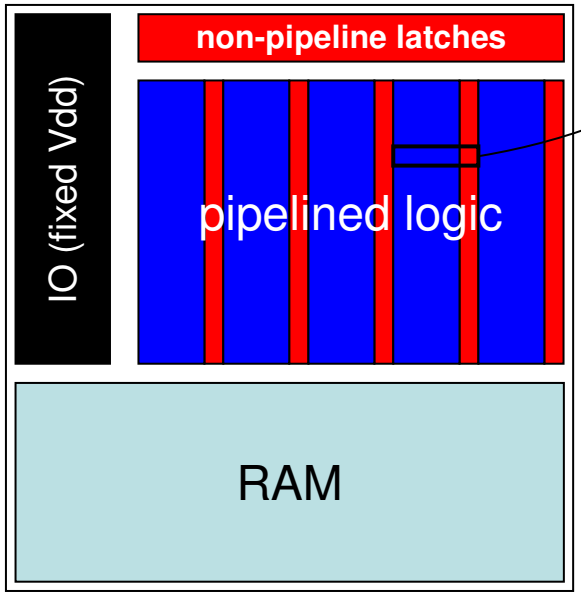
simon@icerasemi.com

Pipeline Model

K “pipettes”…



Chip…



Average n_c transistors of combinatorial logic per pipe stage per pipeline bit.

n_s transistors per pipeline latch.

Units of τ are “i” ≡ “FO4”; 1i = Δ seconds.

Arbitrary width/depth mix of pipettes make up the pipelined logic.

Define…

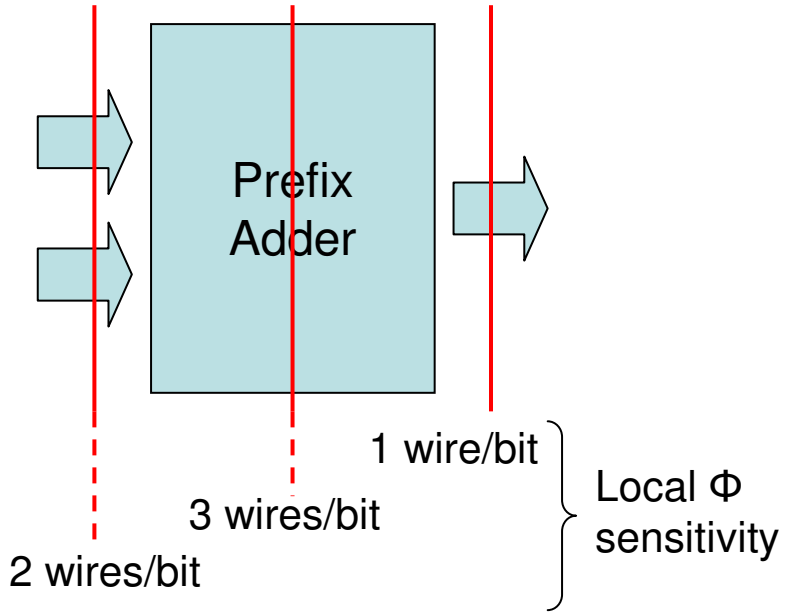
Flux	$\Phi = n_c/\tau_c$	(T/i/b)
Performance	$\Psi = Kn_cf$	(T/s)
Cost	$N = K(n_c+n_s)$	(T)
Frequency	$f = 1/\tau\Delta$	(Hz)

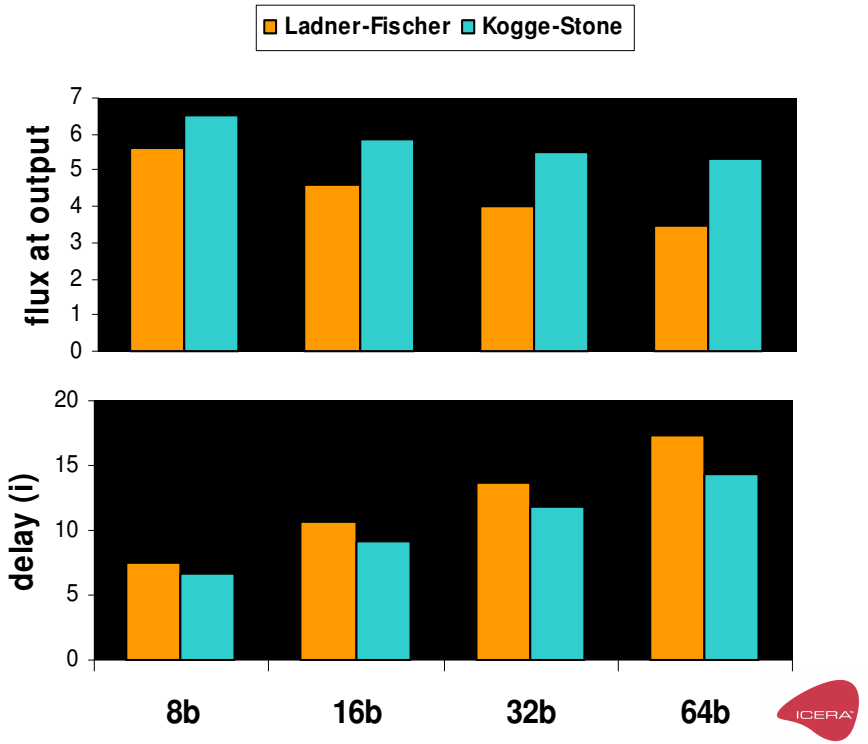
simon@icerasemi.com

Reality not Included…

- Performance Ψ (T_{COMB}/s) tends to over-state effective performance at low τ, because logical redundancy is introduced to speed up logic paths and mitigate latency by speculation.
- Flux Φ (T/i/b) varies a lot locally; tends to rise at low τ for the same redundancy reason, and because pipe cuts have less freedom to avoid high-flux logic.

Eg. adders…



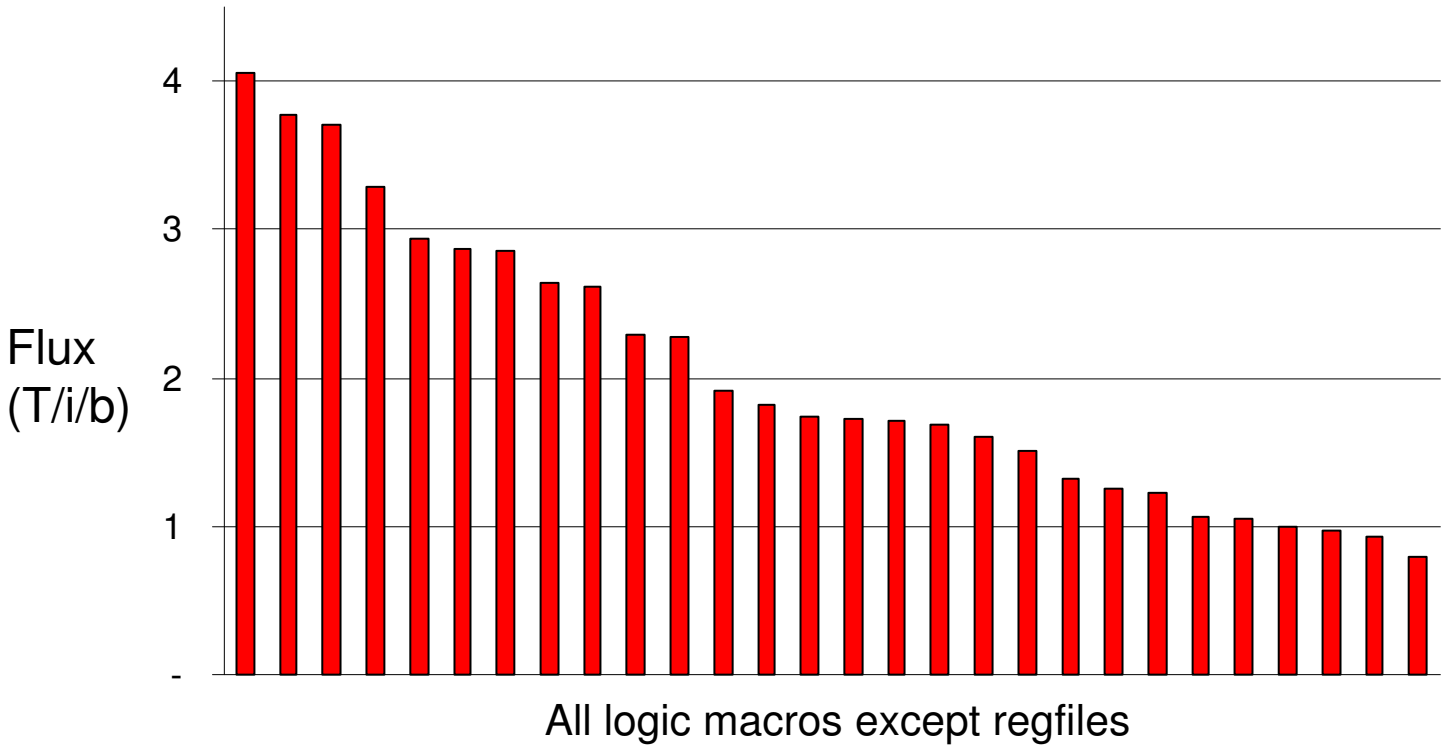


simon@icerasemi.com

Pipeline Flux in Icera's DXP®

Φ ~ 3-4 for arithmetic and random translation logic.

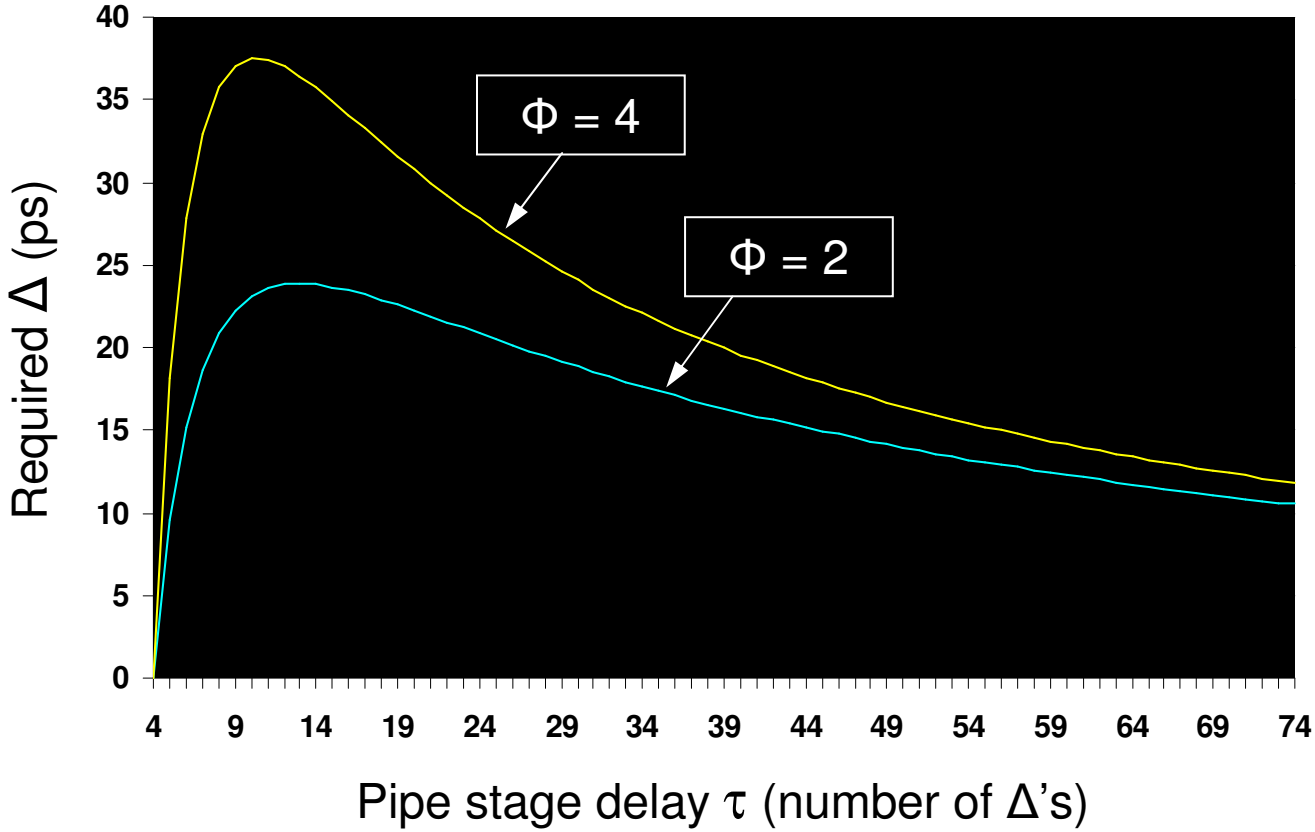
Φ ~ 1-2 for steering logic.



simon@icerasemi.com

Required Δ to meet specified Performance Ψ and Cost N

$$\Delta_{required} = \left(\frac{N}{\Psi\tau}\right) \Big/ \left(1 + \frac{n_S}{\Phi(\tau - \tau_S)}\right)$$

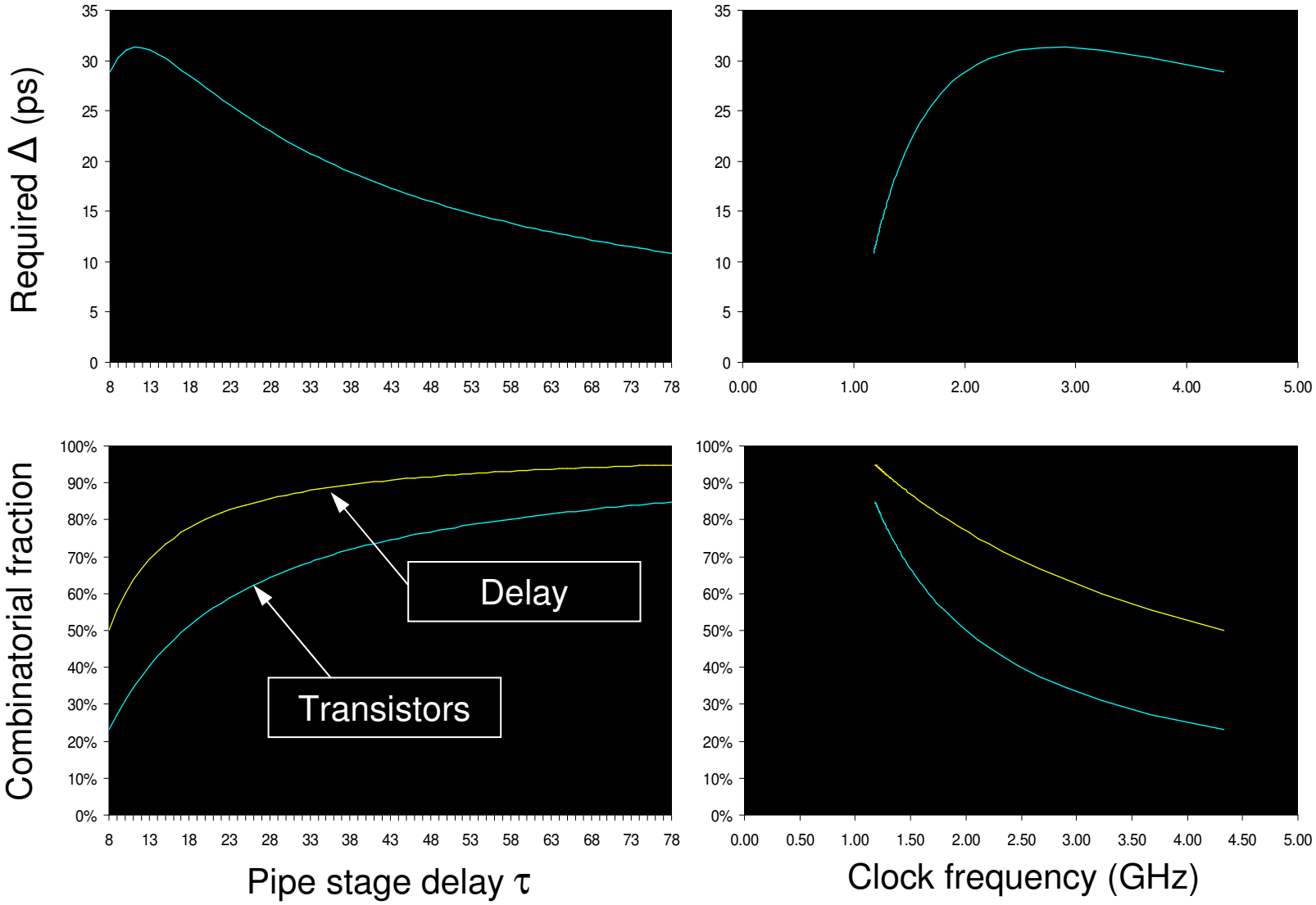


Specification

Latch delay 4i
Latch cost 40T
Performance 10PT/s
Design Cost 10MT

simon@icerasemi.com

Constant Flux Implications ($\Phi = 3$, $\tau_s = 4i$)



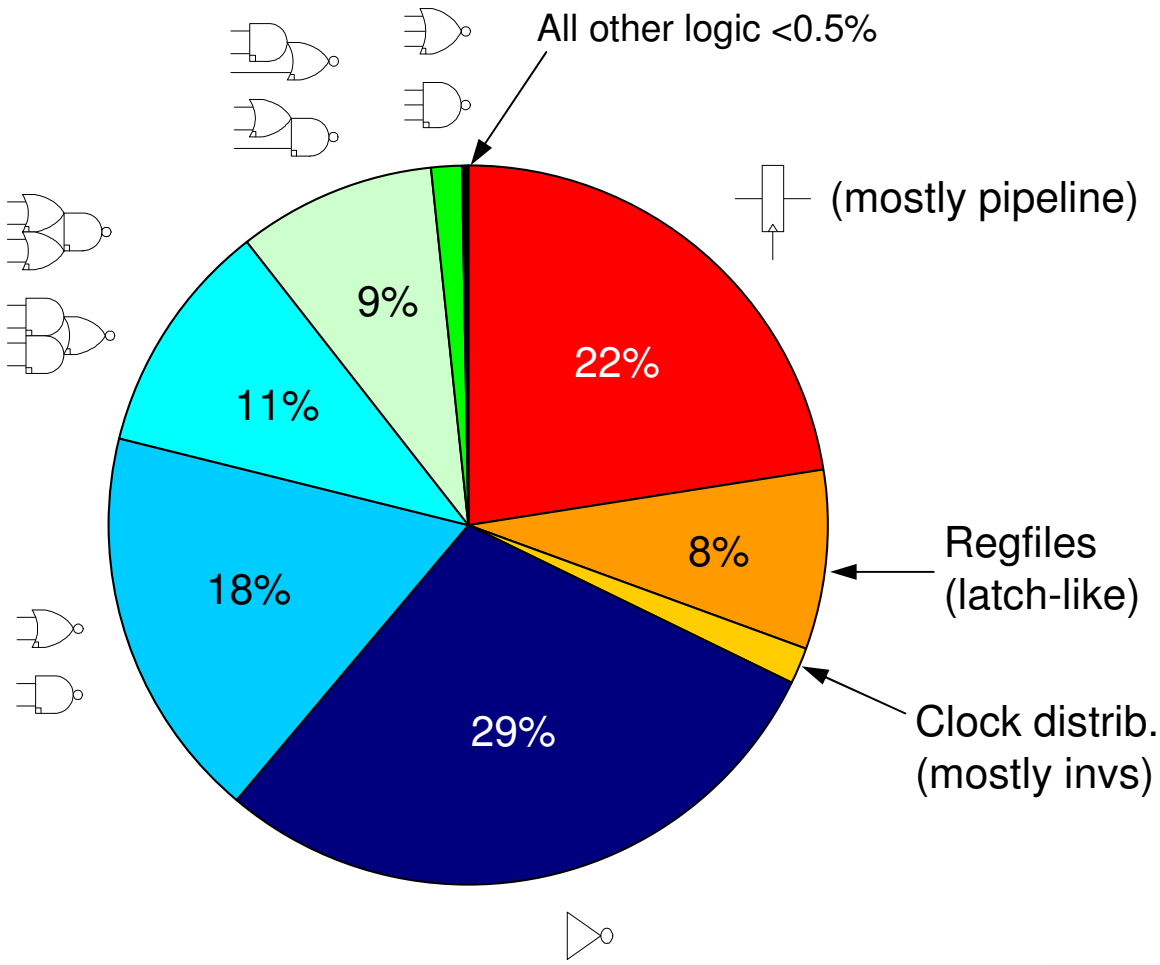
simon@icerasemi.com

DXP® Logic Stage Population, by #Transistors

Manhattan cells are ubiquitous in all digital design methodologies.

Complex cells can be important for performance but account for negligible % of total transistors.

Designs are characterized (capacitance, leakage, delay, area) by a few simple Manhattan stages.



simon@icerasemi.com

The Design Hull

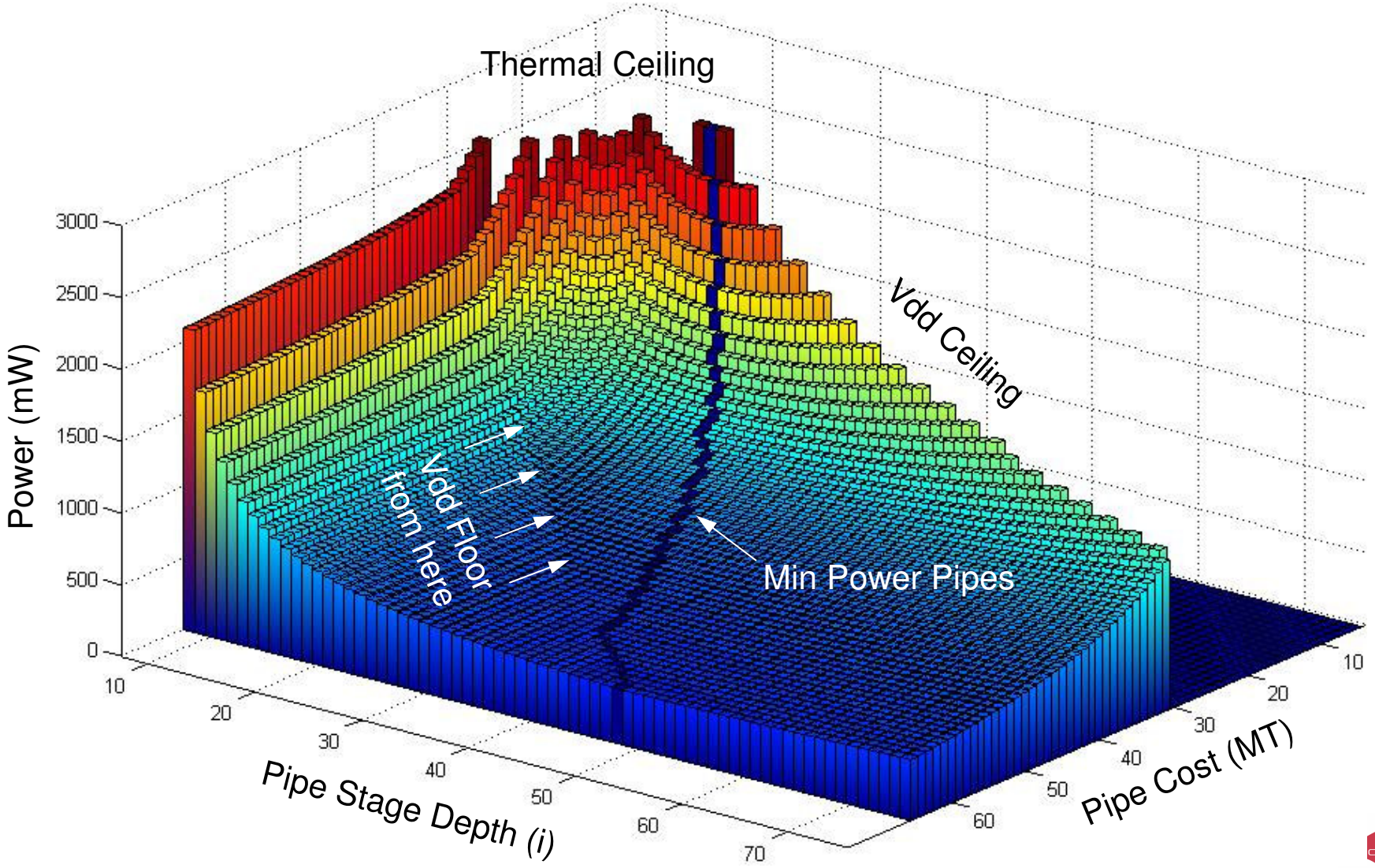
simon@icerasemi.com

Example Design Spec

Performance Ψ = 15PT/s
Flux Φ = 4T/i/b
Latch Cost n_S = 40T
Latch Delay τ_S = 4i
Non-Pipe Logic = 5MT
RAM = 1MB
Latch Activity = 3x Combinatorial Activity
RAM Activity = 0.2x Combinatorial Activity
65nm high V_T LOP Process
Ambient Temp = 85℃
Max Die Temp = 125℃
Package θ_{JC} = 15℃/W
Supply R = 50mΩ
Fixed Parasitic Power = 100mW
Vdd Limit Range = 0.7V – 1.2V

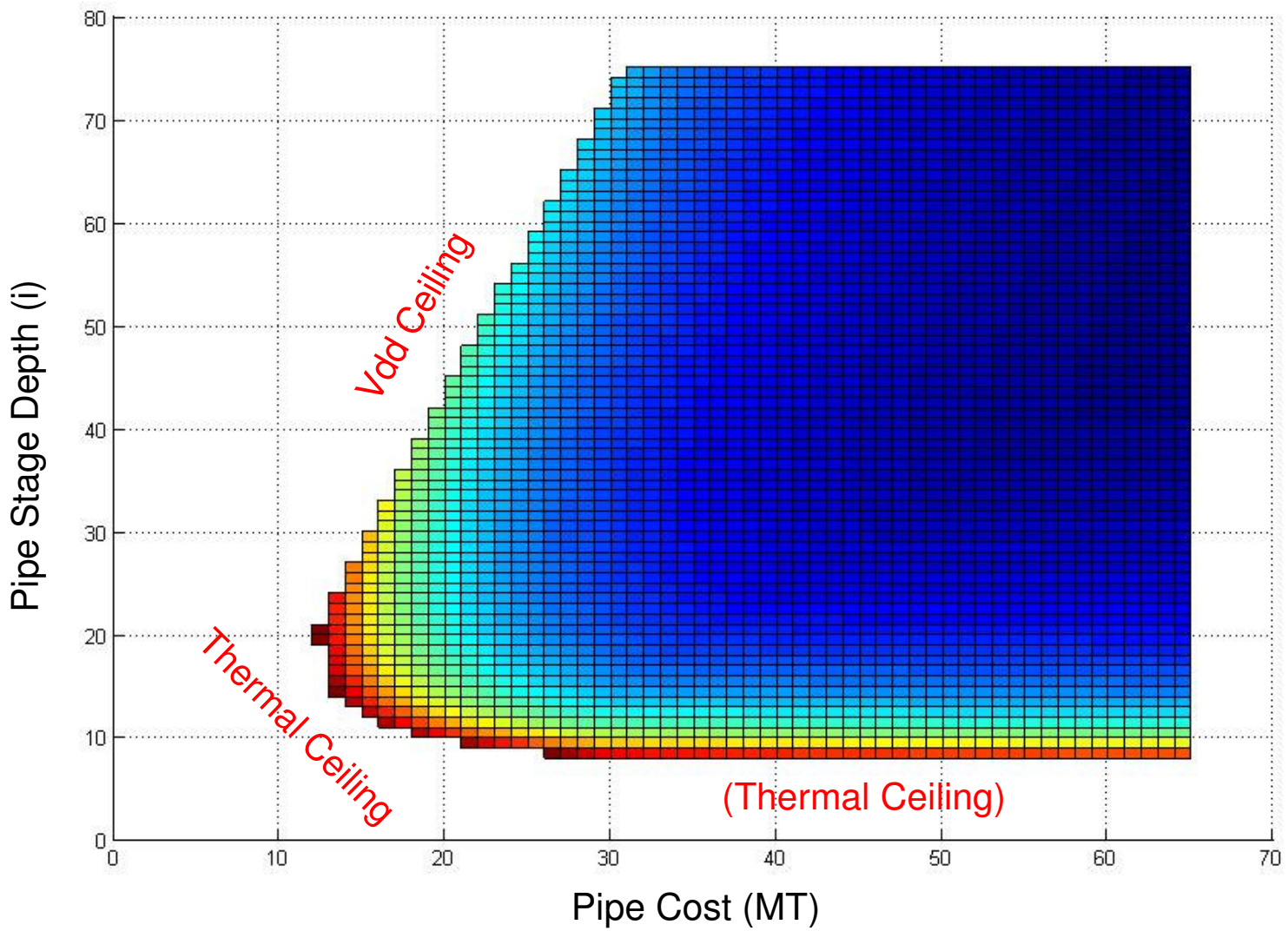
simon@icerasemi.com

Design Hull

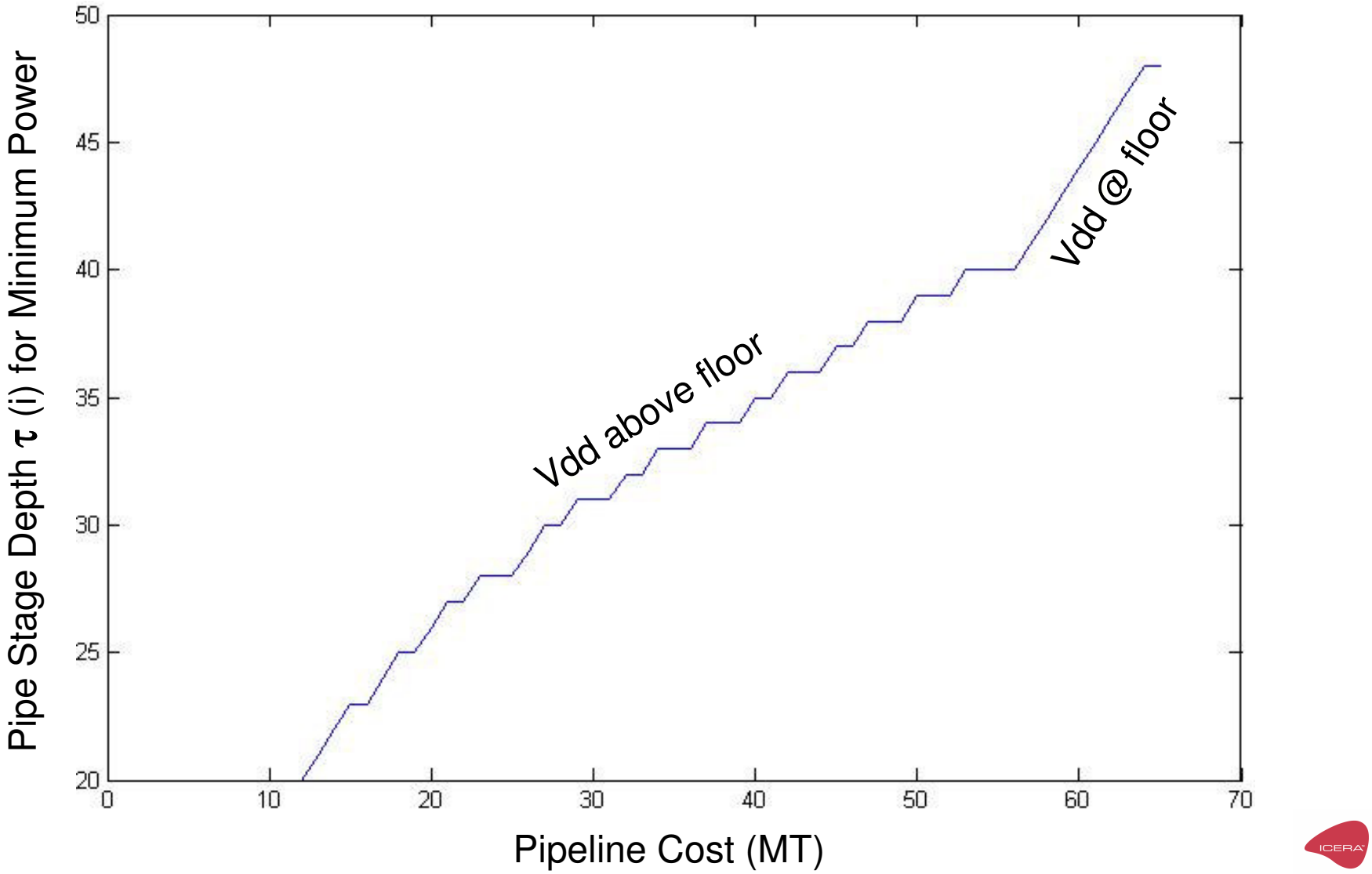


simon@icerasemi.com

Feasible Designs

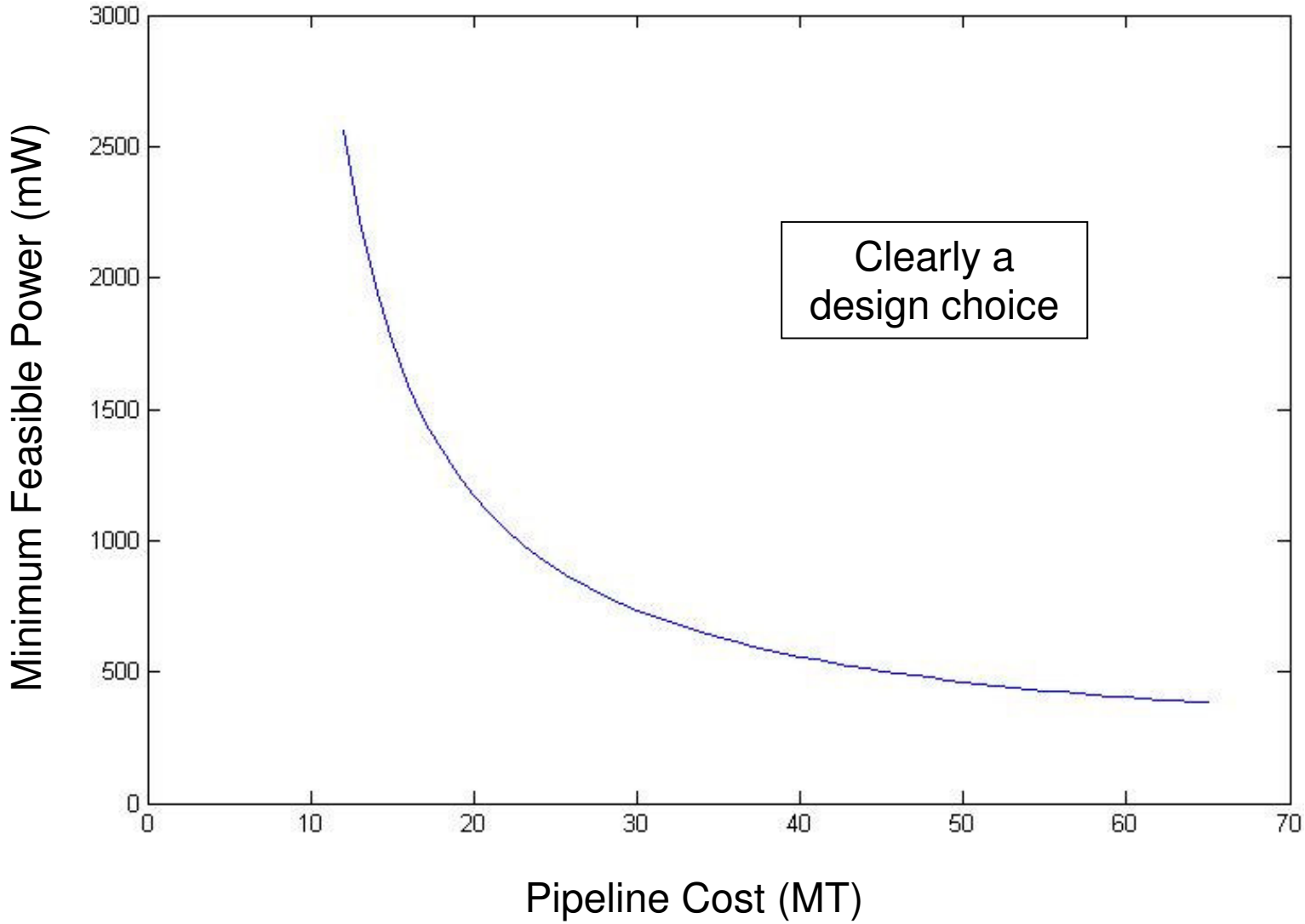


Minimum Power Pipelines



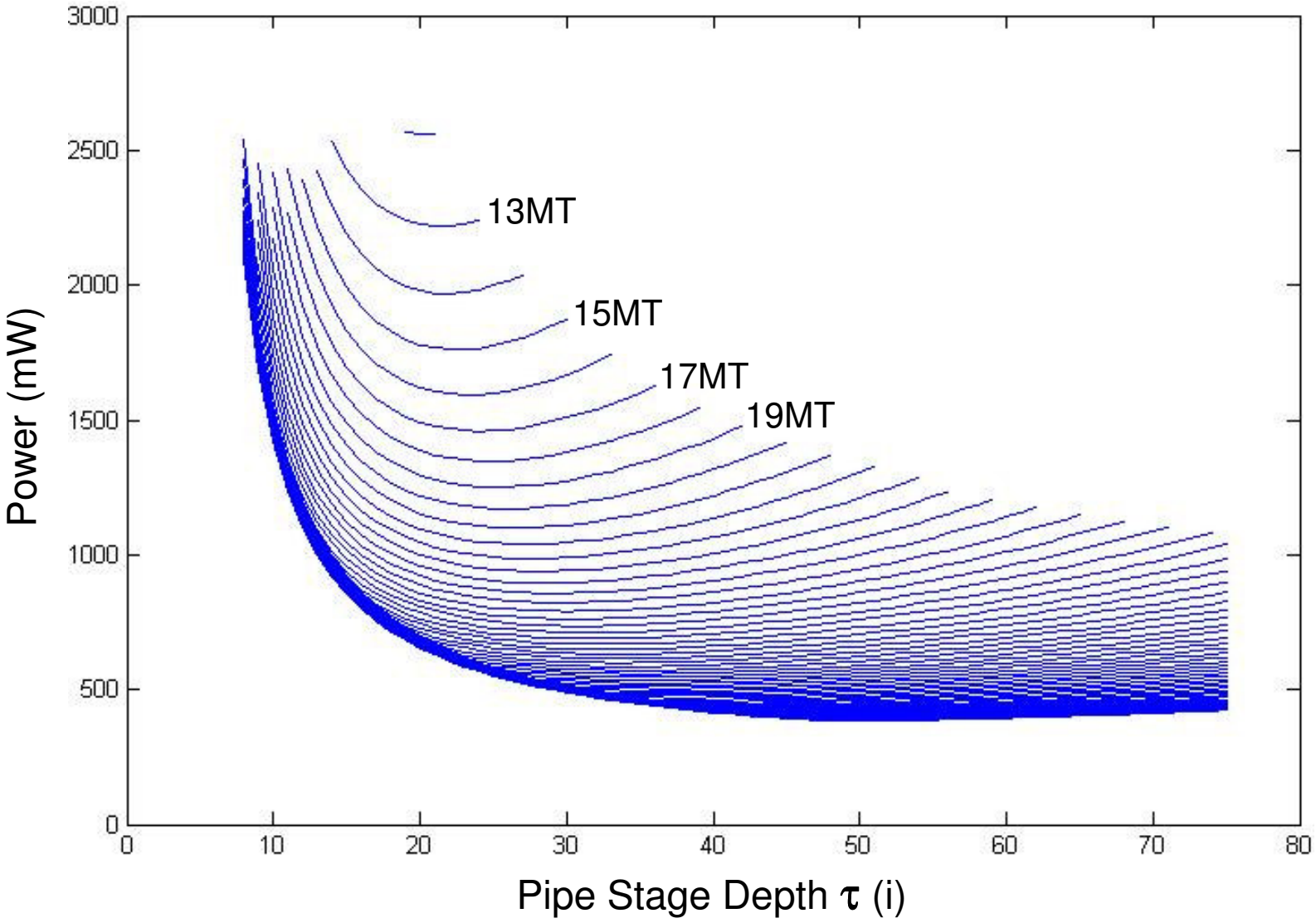
simon@icerasemi.com

The Cost of Minimizing Power



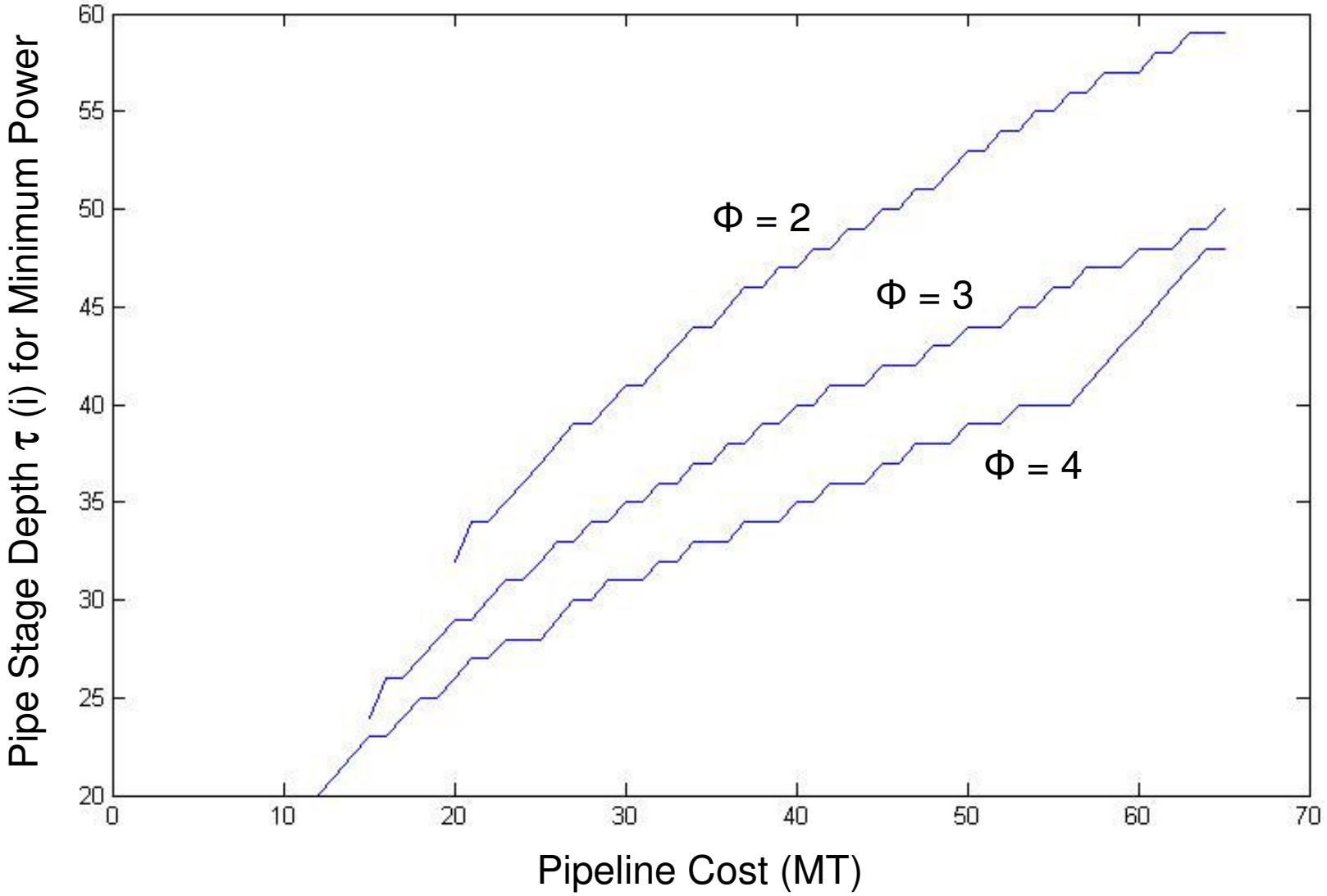
simon@icerasemi.com

Minimum Power Pipelines



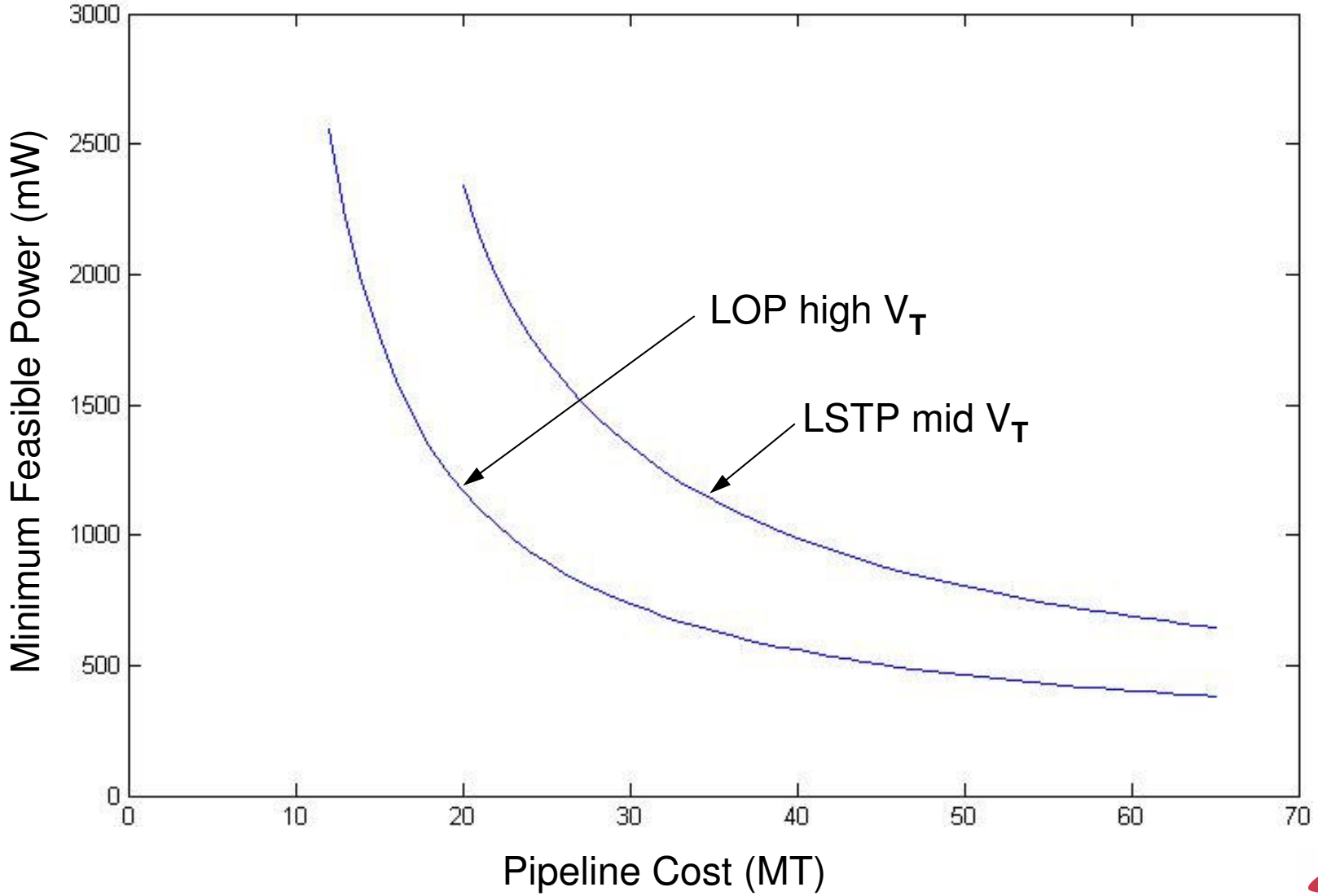
simon@icerasemi.com

Sensitivity to Pipeline Flux



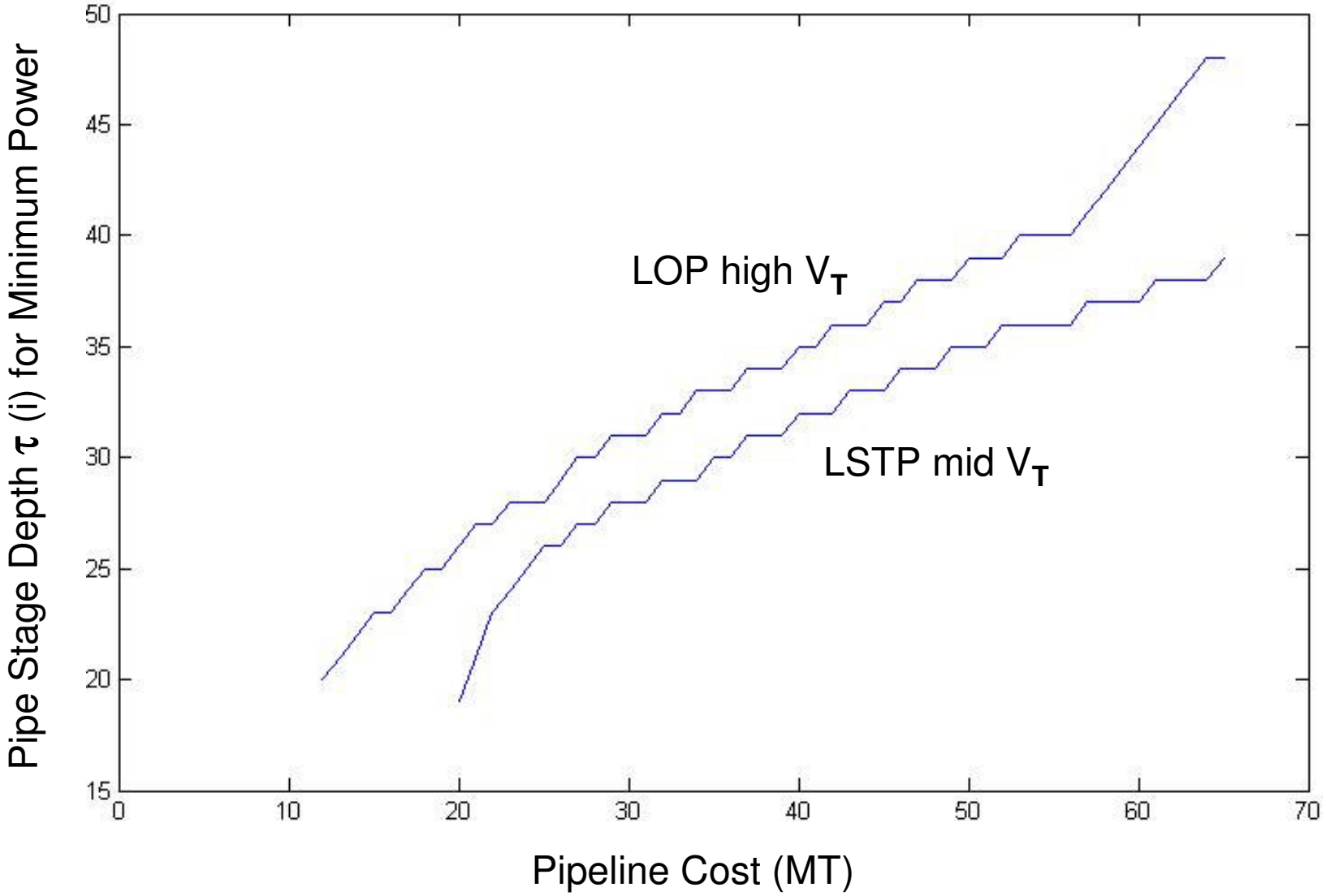
simon@icerasemi.com

LOP vs LSTP



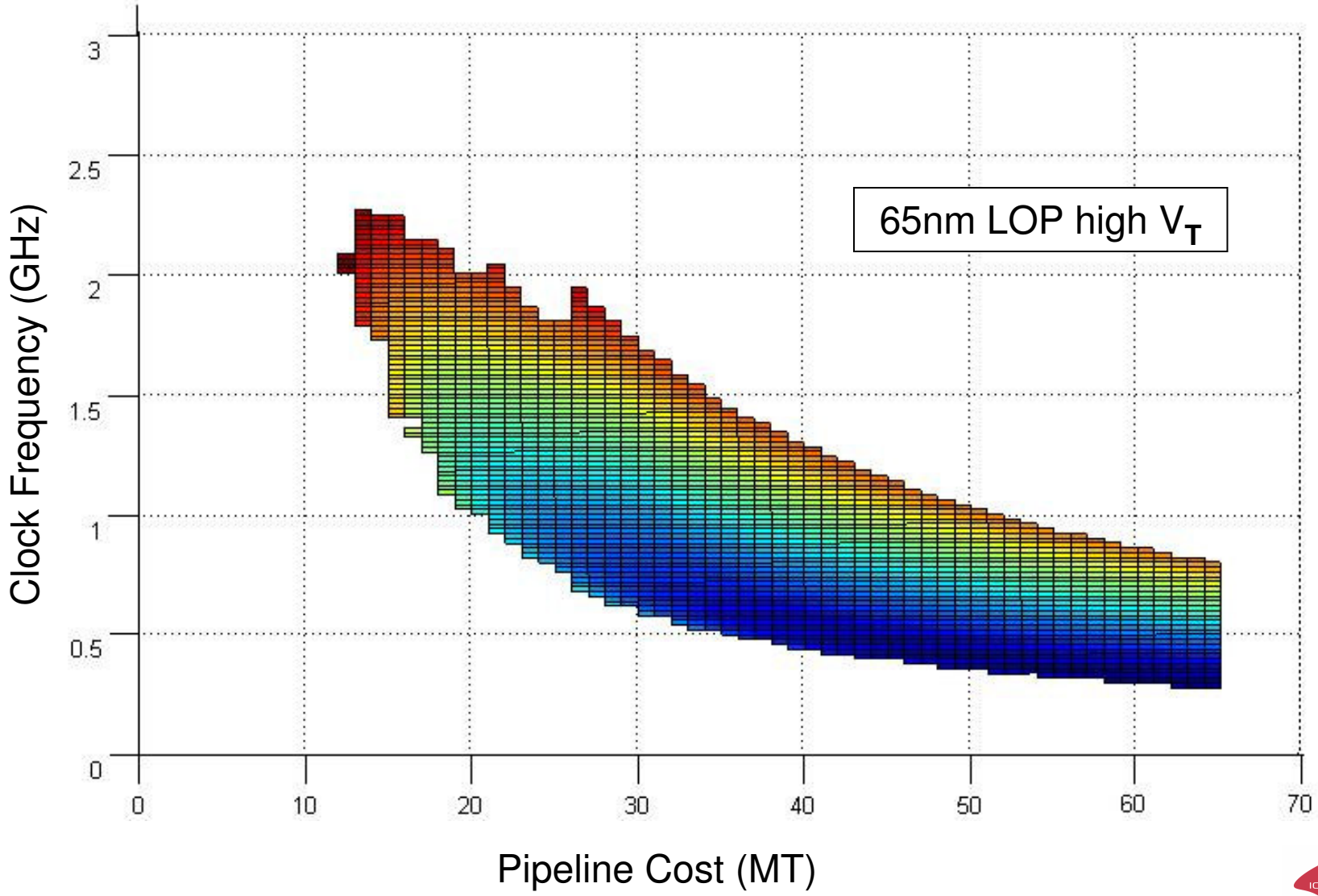
simon@icerasemi.com

LOP vs LSTP



simon@icerasemi.com

Feasible Solutions in Cost-Frequency Space



simon@icerasemi.com

Implications for Design

simon@icerasemi.com

Implications for Chip Design

Every logic chip will have control over its supply voltage.

- Integrated regulators?
- Switch off in standby – limited call for LSTP processes?

We can't minimize cost and power simultaneously.

Good cost-power points require fairly high speed.

- Beyond synthesis and auto-P&R today → more Structured Custom?

Short pipelines expose the basic TA assumption (single path sensitization).

- Wavefront timing analysis tools?
- Voltage agility makes TA sign-off harder.

simon@icerasemi.com

Structured Custom

Natural cell placement by engineers…

…is enough for predictable auto-routing

simon@icerasemi.com

Thank You

Enjoy the Symposium ☺

simon@icerasemi.com