Διαδικαστικά

- Καλό ανάγνωσμα για Low Power:
  - “Low Power Methodology Manual” - Michael Keating et al.
  - “Low Power Design Essentials” - Jan Rabaey

- 2η Εργαστηριακή άσκηση – ALU
  - abs(A) αντί για A*B
  - Οι είσοδοι και έξοδοι είναι προσημασμένοι (two’s complement)
  - Qm.n

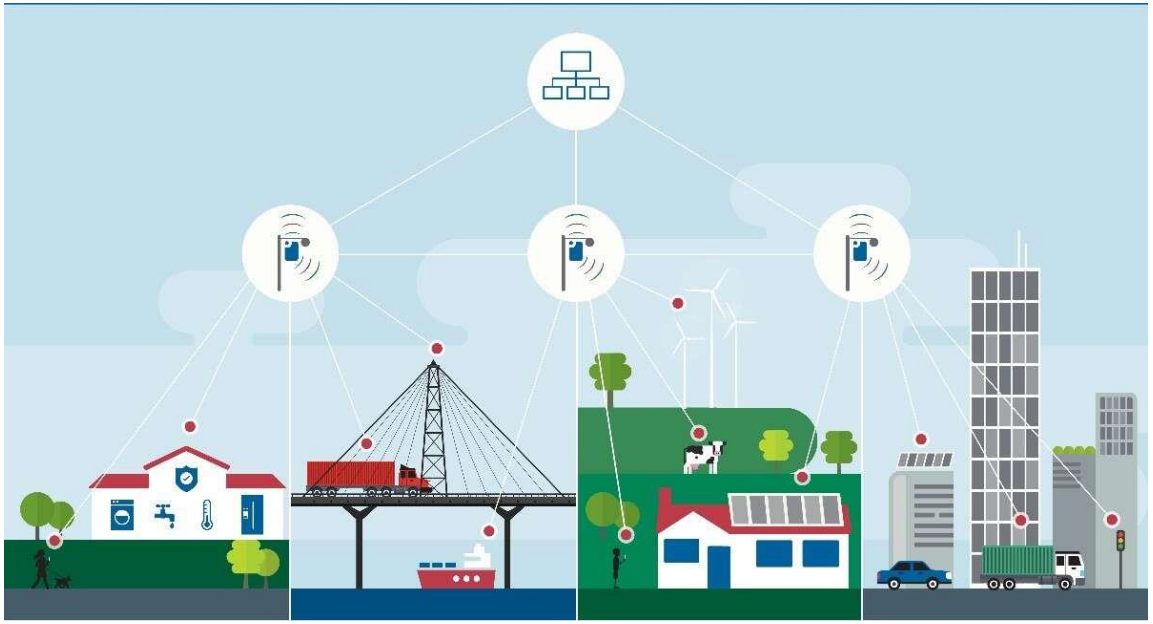
Low Power Design

Increasing Battery Life

- Low-Power IoT will will exceed $2.6 billion by 2024 [1]

src: https://www.dr-hempel-network.com/

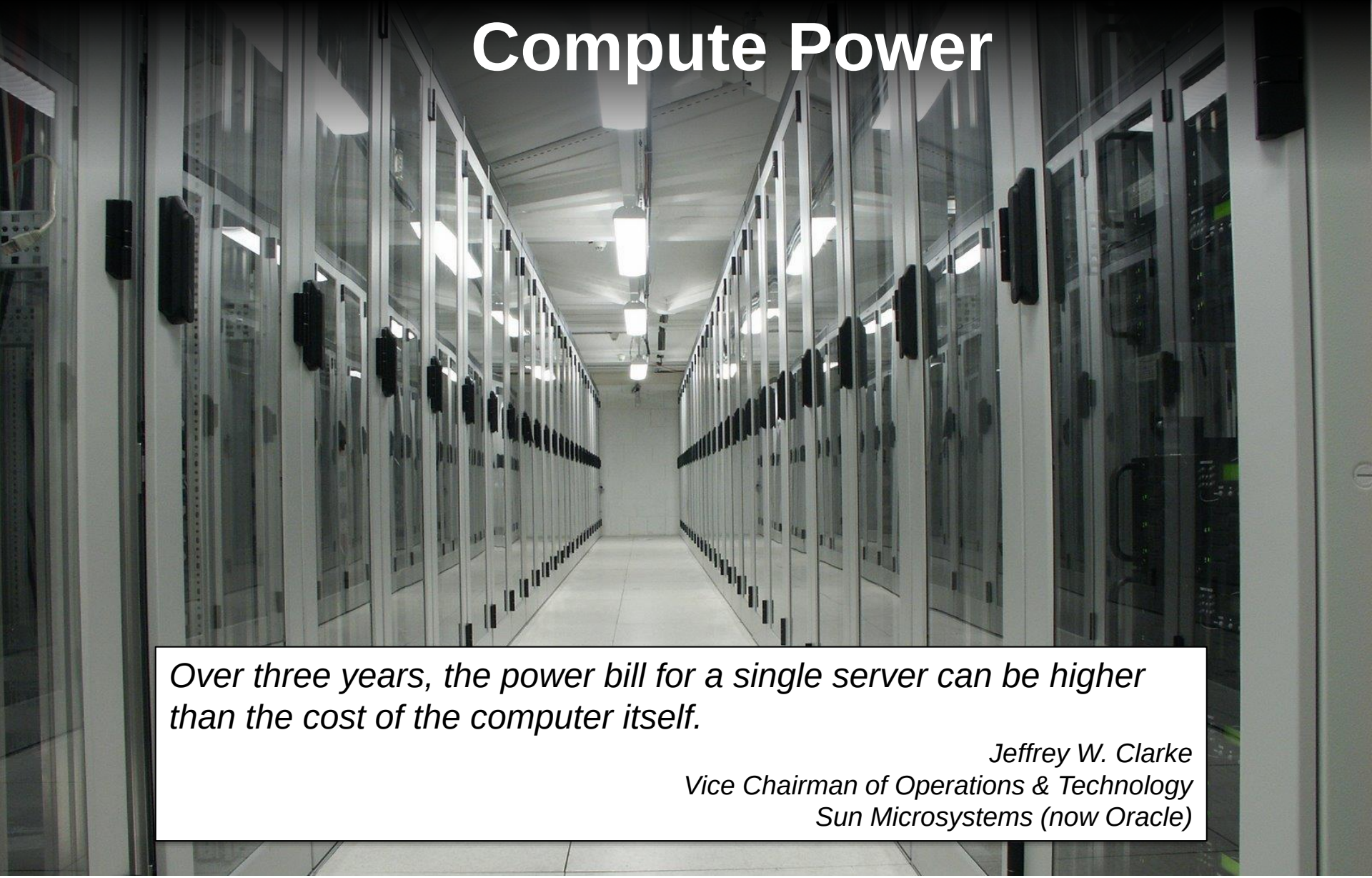
src: https://www.wired.com/



src: https://iot.eetimes.com/

[1] https://www.juniperresearch.com/researchstore/sustainability-technology-iot/esims-research-report

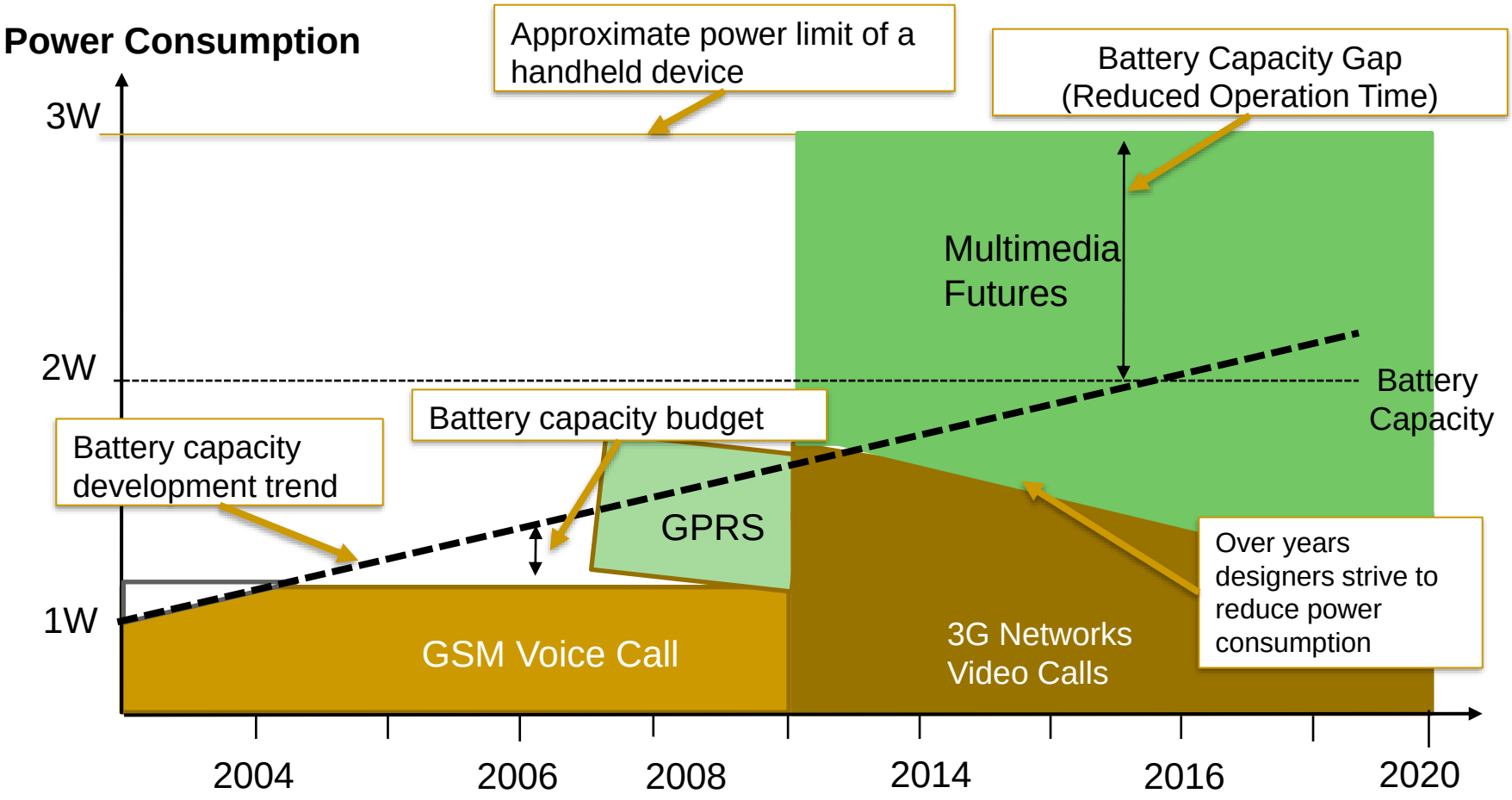
Compute Power



Over three years, the power bill for a single server can be higher than the cost of the computer itself.

Jeffrey W. Clarke
Vice Chairman of Operations & Technology
Sun Microsystems (now Oracle)

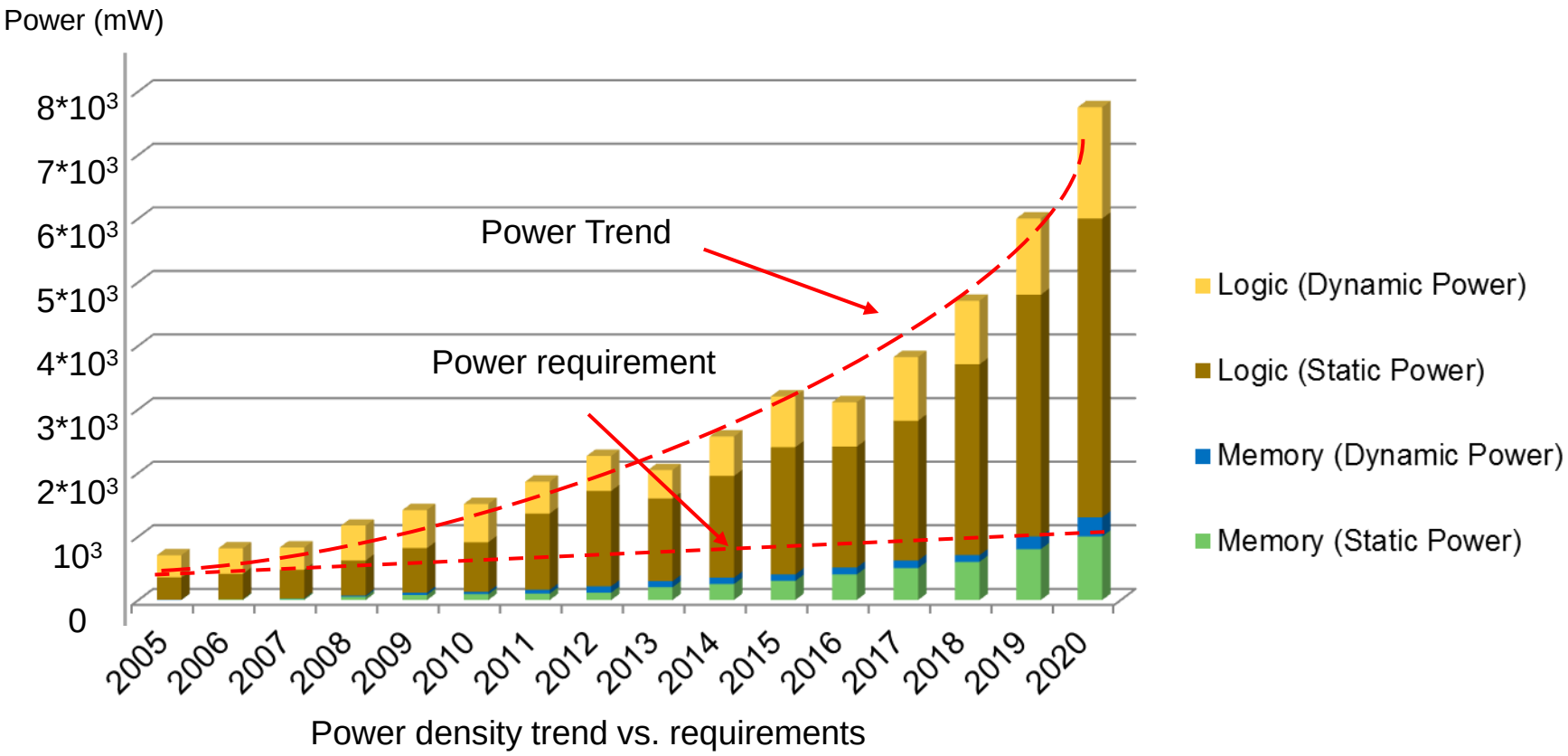
Power Consumption & Battery Capacity Trends



Requirements will Continue to Increase

YEAR OF PRODUCTION	2006	2009	2010	2011	2015	2018	2020
Process Technology (nm)	90	65	45	32/28	12	7	5
Supply Voltage (V)	1	0.8	0.6	0.5	0.3	0.3	0.3
Clock Frequency (MHz)	2000	2500	2900	3200	4000	4500	4800
Application (maximum required performance) Application (other)	Real Time Video Codec (MPEG4/CIF) TV Telephone (1:1) Voice Recognition (Input) Authentication(Crypto Engine)		Real Time Interpretation TV Telephone (>3:1) Voice Recognition (Operation)		Mobile	Wearable	Wearable
Processing Performance (GOPS)	2	14	77	461	2458	8500	8500
Required Average Power (W)	0.1	0.1	0.1	0.1	0.1	0.05	0.05
Required Standby Power (mW)	2	2	2	2	2	1.5	1.5
Battery Capacity (Wh/Kg)	200	200	400	400	400	500	500

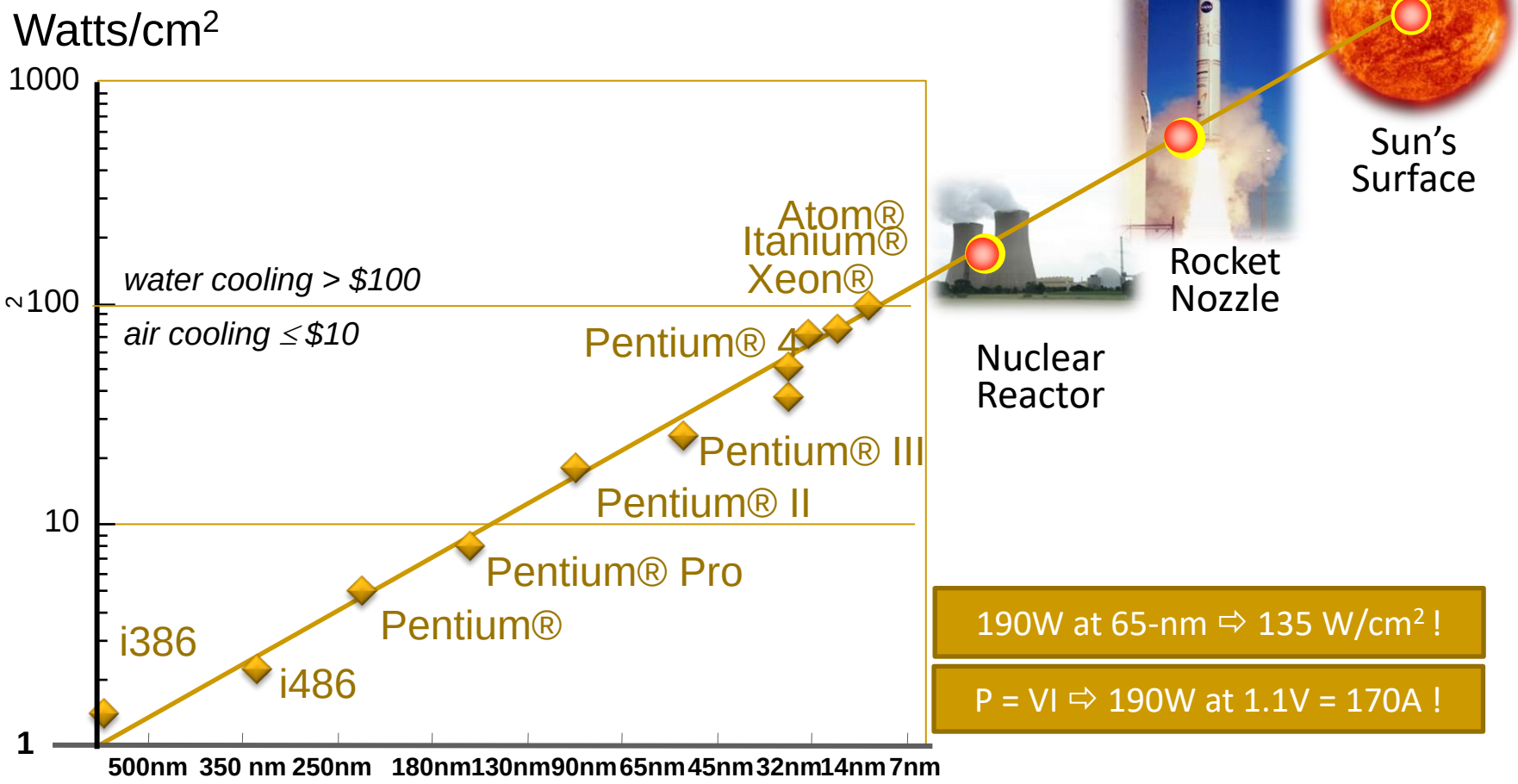
Power Consumption of Integrated Circuits



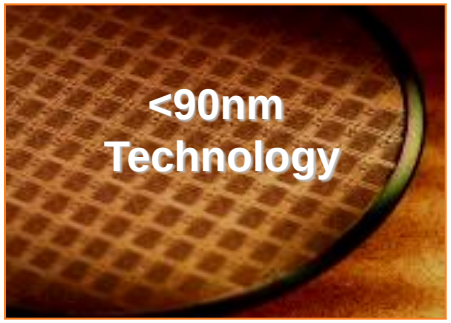
Power density trend vs. requirements

http://www.eetimes.com/document.asp?doc_id=1278448

Power Consumption Trends

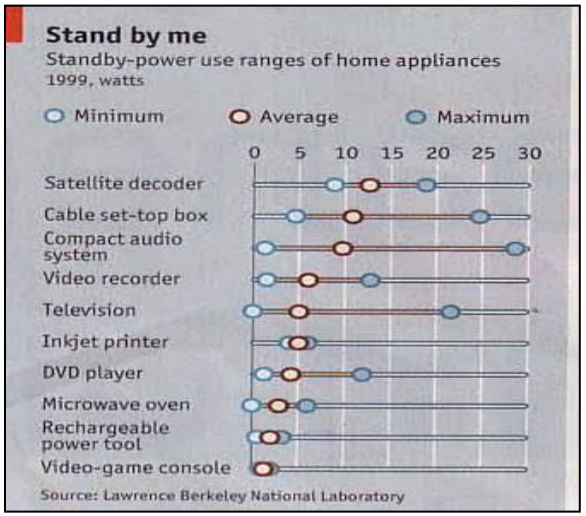


Power Affected Problems

	Low Power	Power Efficiency	Reliability
			
Application	• Wireless • Handheld • Embedded systems	• Microprocessors • Graphics/multimedia • Networking/telecom	• All design<90nm
Concern	• Battery life • Leakage power • Dynamic power	• Thermal management • Packaging, cooling cost	• Leakage power • IR-drop • Electromigration

SYNOPSYS®

The World Demanding Low Power

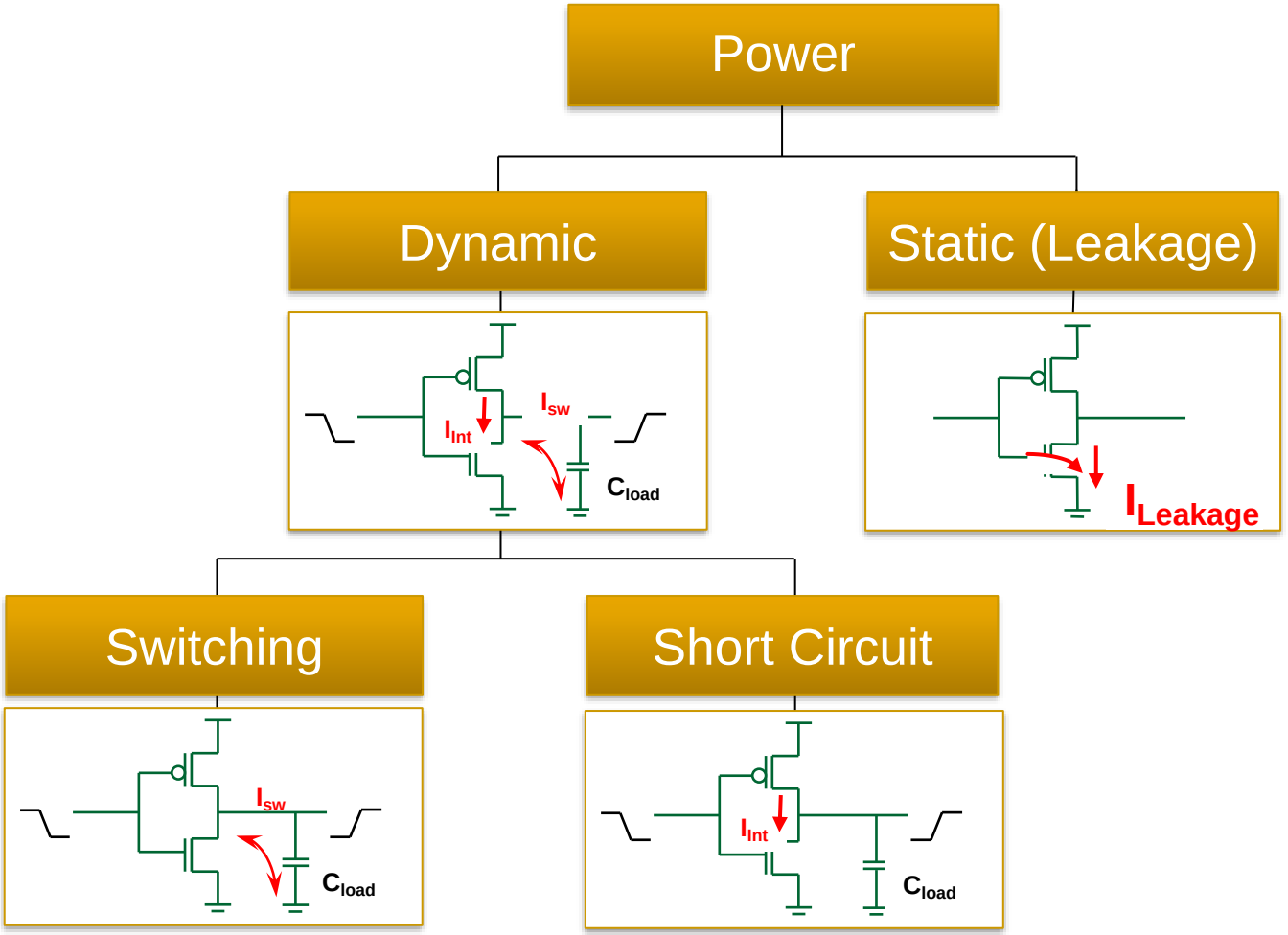


Silent Thievery

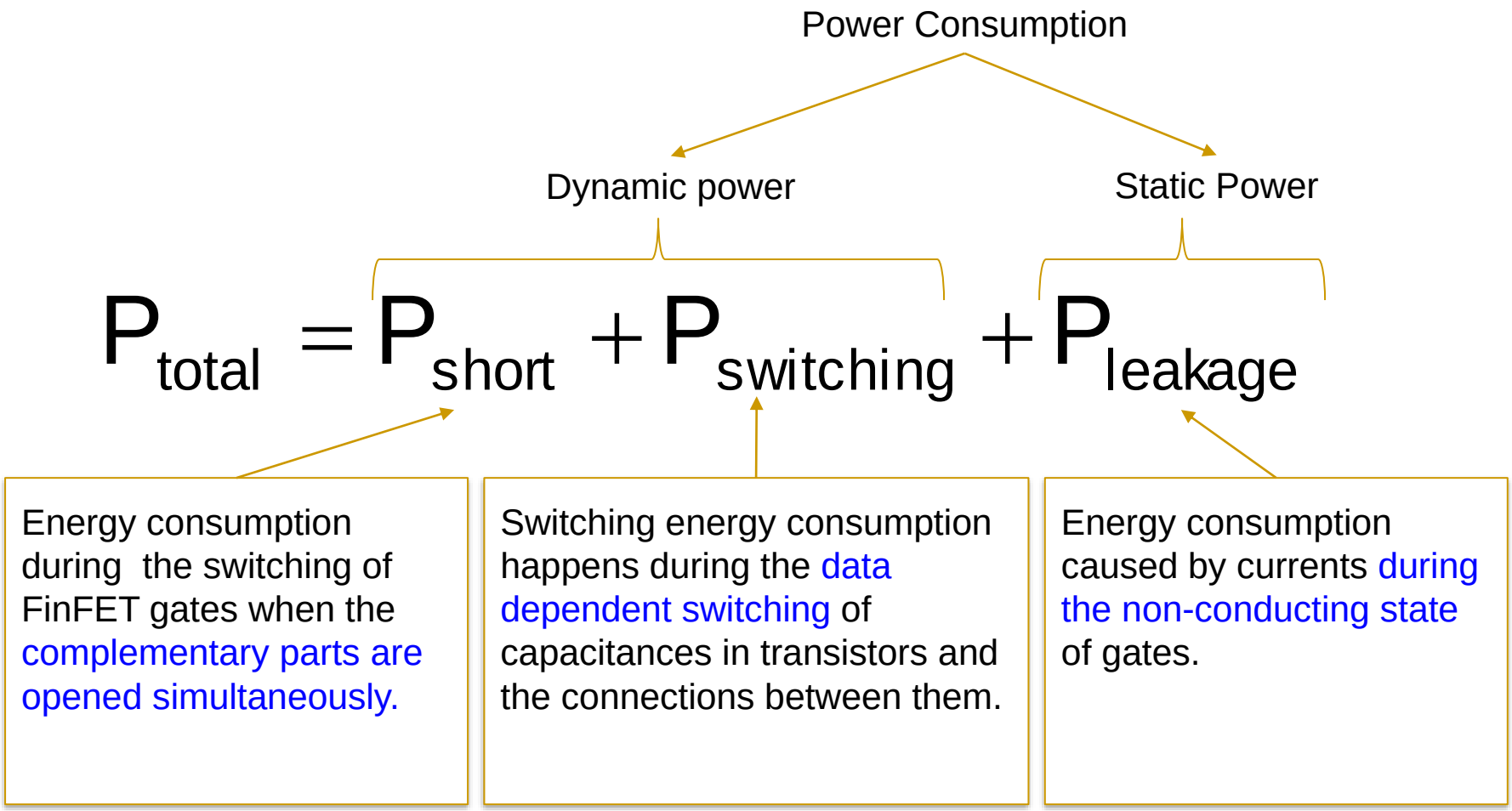
ergy, in other words, is equivalent to the output of 18 typical power stations.

SYNOPSYS®

Sources of Power Dissipation in FinFET

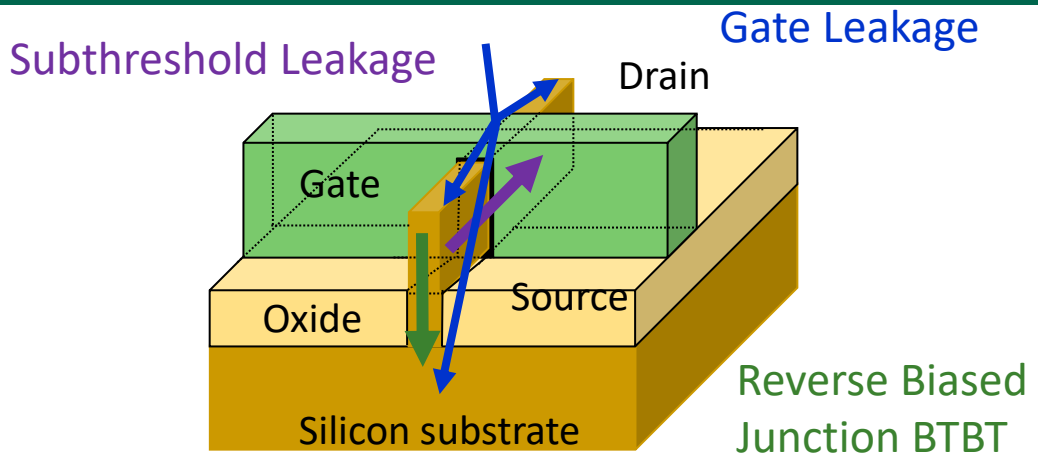


Sources of Power Dissipation in FinFET



SYNOPSYS®

Leakage Current Components



Subthreshold leakage

$$I_{sub} = I_0 \exp^{q\frac{V_{gs}-V_{th}}{nkT}} \cdot \left(1 - \exp^{q\frac{-V_{ds}}{kT}}\right)$$

$$I_O \approx \beta\left(\frac{kT}{q}\right)^2 e^{1.8} \qquad \beta = \mu C_{ox}\frac{W}{L}$$

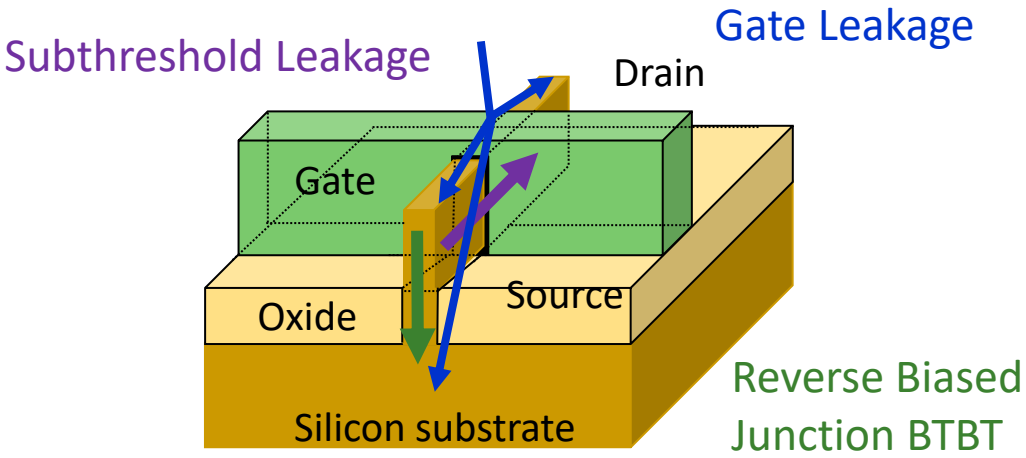
Gate Leakage

$$J_G = J_0 \exp\left\{-\frac{8\pi\sqrt{2q}}{3h}\left(m_{eff}\,\Phi_b\right)^{0.5} Kt_{ox,eq}\right\}$$

Reverse Biased Junction (Band-To-Band-Tunneling) BTBT Leakage

$$I_{reverse} = A \cdot J_S\left(e^{\frac{qVapp}{kT}} - 1\right)$$

The Role of Leakage Components

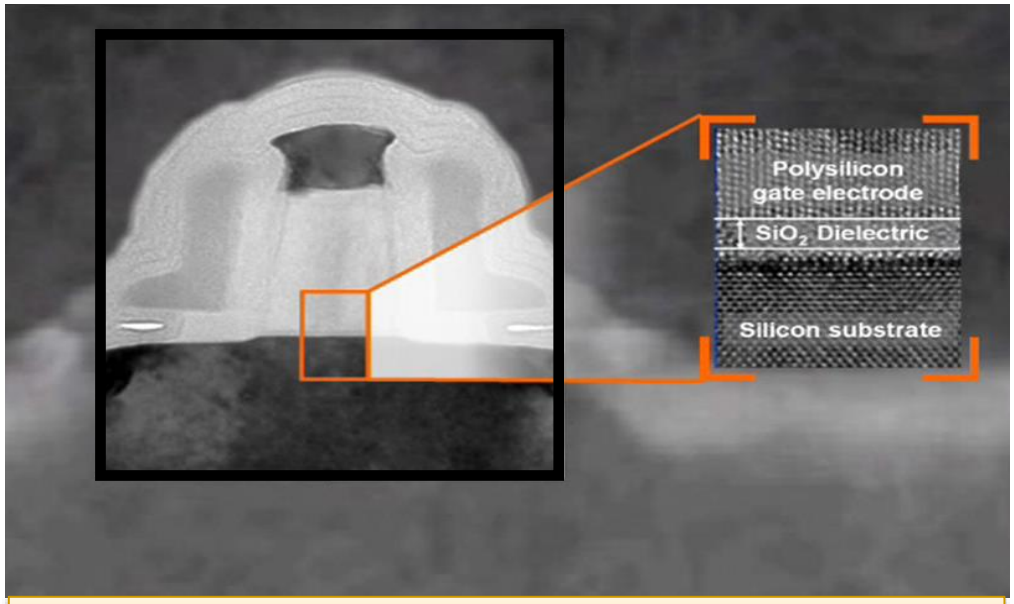


Total Transistor Leakage = $I_{overall} = I_{BTBT} + I_{sub} + I_{gate}$

Long Channel	Short Channel	Very Short channel	Nano-scaled
(L > 1µm)	(L > 180 nm, T_{ox} > 30Å)	(L > 14nm, T_{ox} > 20Å)	(L < 14 nm, T_{ox} < 20Å)
Negligible leakage	Subthreshold leakage	Subthreshold + Gate Leakage	Subthreshold + Gate + BTBT Leakage

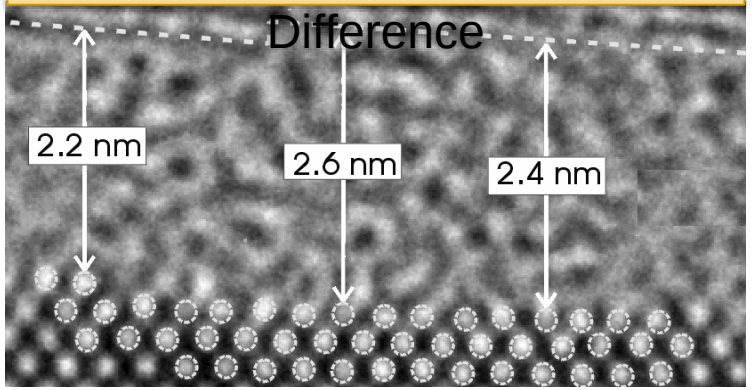
SYNOPSYS

Leakage Sources



3 to 5 Molecules Of SiO_2 Dielectric

±1 Molecule Makes The Difference



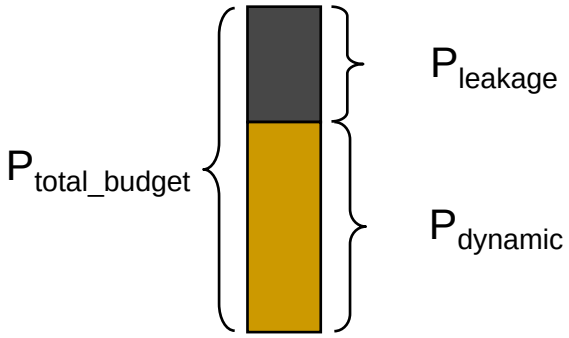
SYNOPSYS®

Leakage Definition

- Leakage Current
  - Current that flows due to inability to short-off transistor
  - Transistors are harder to turn off with very low voltages
- Standby Leakage
  - Leakage that occurs while overall circuit is sleeping
    - Example: Laptop sleep mode
- Active Leakage
  - Leakage that occurs while overall circuit is opening
    - Example: FPU while PC performs integer operations

Leakage Problem

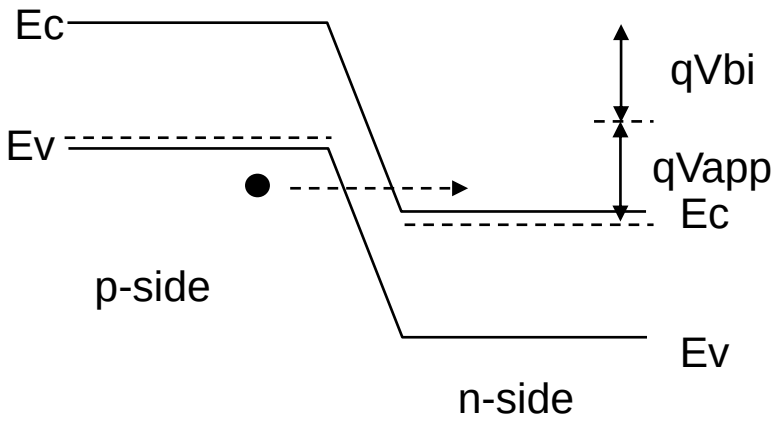
- Technology trends result in increased leakage
  - Advanced processes and reduced supply voltages lead to increased leakage currents
- Market demands decreased leakage
  - Longer battery life for wireless and consumer markets
  - More performance for high performance markets (CPU, DSP)



SYNOPSYS®

P-n Junction Reverse-Bias Current

- Drain and source to well junctions are typically reverse biased, causing p-n junction leakage current. A reverse-bias p-n junction leakage has two main components:
  - Minority carrier diffusion/drift near the edge of the depletion region
  - Due to electron-hole pair generation in the depletion region of the reverse-biased junction
- P-n junction reverse-bias leakage is a function of junction area and doping concentration. If both n and p regions are heavily doped (this is the case for advanced MOSFETs using heavily doped shallow junctions and halo doping for better SCE), band-to-band tunneling (BTBT) dominates the p-n junction leakage.



Band-to-Band Tunneling Current

- High electric field 10 V/cm across the reverse-biased p-n junction causes significant current to flow through the junction due to tunneling of electrons from the valence band of the p region to the conduction band of the n region.

BTBT current density
- Depends on junction field (ξ)
- Junction voltage (Vapp)
- Band-gap (Eg)

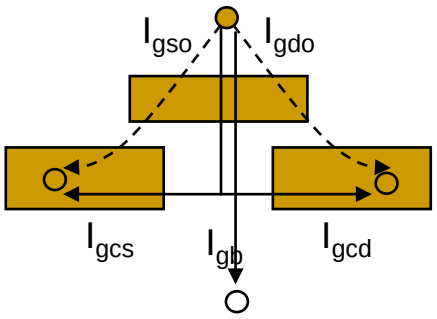
High BTBT in scaled devices
- High junction doping: “Halo” profiles
- Small depletion width
- Large electric field

$$J_{b-b} = A \frac{\xi V_{app}}{E_g^{1/2}} \exp\left(- B \frac{E_g^{1/2}}{\xi} \right)$$

$$A = \frac{\sqrt{2m^*}q^3}{4\pi^3 h^2}, B = \frac{4\sqrt{2m}}{3qh}$$

Gate Leakage

- Gate direct tunneling current is due to the tunneling of electrons (or holes) from the bulk silicon and source/drain (S/D) overlap region through the gate oxide potential barrier into the gate (or vice-versa).
- The tunneling current increases exponentially with the decrease in the oxide thickness and the increase in the potential drop across oxide.
- Major components of gate tunneling in a scaled MOSFET device are:
  - Gate to S/D overlap region current (Edge Direct Tunneling (EDT)) components (I_{gso} and I_{gdo})
  - Gate to channel current (I_{gc}), part of which goes to source (I_{gcs}) and rest goes to drain (I_{gcd})
  - Gate to substrate leakage current (I_{gb}). The overlap tunneling dominates the gate leakage in an 'off' (V_{gs}=0) transistor, whereas, gate-to-channel tunneling control the gate current in an 'on' device.



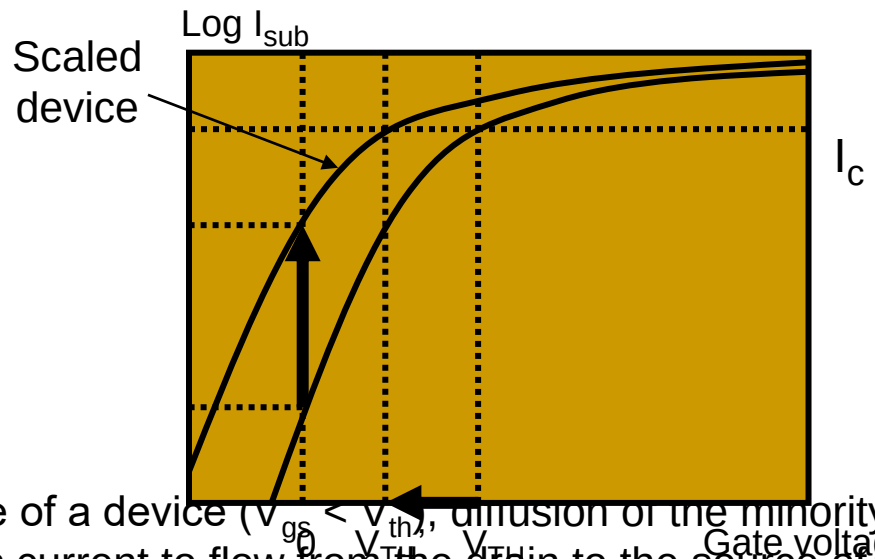
Transistor off (V_g="0") – I_{gdo} and I_{gso} dominates

Transistor on (V_g="1") – I_{gc} (I_{gcs} and I_{gcd}) dominates

I_{gb} is small compared to others

SYNOPSYS®

Subthreshold Leakage



- In the “off” state of a device ($V_{gs} < V_{th}$), diffusion of the minority carriers through the channel causes current to flow from the drain to the source of a transistor.
- The subthreshold current depends exponentially on both gate-to-source voltage and V_{th}. A reverse bias applied at the substrate increases V_{th}, thereby reducing the subthreshold current (Body effect).

Subthreshold Leakage (I_{sub})

- Exponential dependence on V_{gs} and V_{th} W= nf*nfin*(2* H_{fin} + tf)

$$I_{sub} = \frac{w_{off}}{L_{off}} \mu \sqrt{\frac{q\varepsilon_{si} N_{cheoff}}{2\Phi_s}} v_r^{\,2} \exp\left(\frac{V_{gs} - V_{th}}{n v_r}\right)\left(1 - \exp\frac{-V_{ds}}{v_r}\right)$$

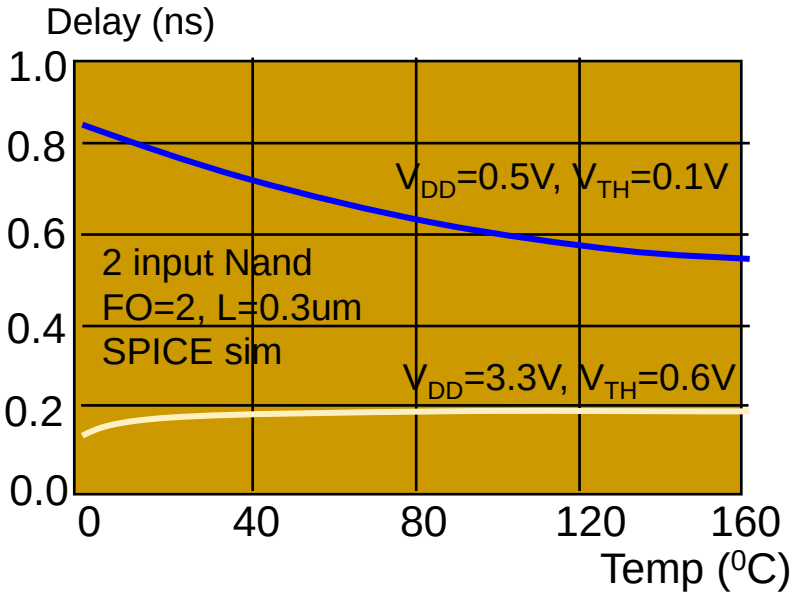
- V_{th} modulation – thickness of fin
- Short channel effect – V_{th} reduction due to
  - Increase in V_{ds} (DIBL)
  - Reduction in channel length (V_{th} roll off)
- Body effect – negative V_{bs} increases V_{th}
- Quantum confinement effect – increases V_{th}

$$V_{th} = V_{FB} + \left(\Phi_{s0} - \Delta\Phi_s\right) + \gamma\sqrt{\Phi_{s0} - V_{bs}}\left(1 - \lambda\frac{W_{dm}}{L_{off}}\right) + V_{nce} + V_{QM}$$

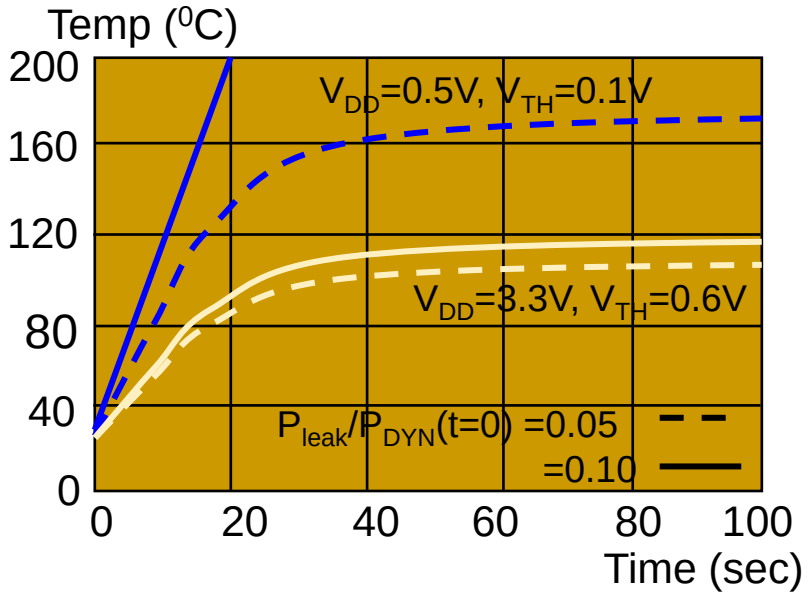
$$\Delta\Phi_s = \left[2\left(V_{bi} - \phi_{S0}\right) + V_{ds}\right] \times \left[e^{-L/2l_c} + 2e^{-L/l_c}\right] \text{ and } l_C = \sqrt{\left(\varepsilon_{si}/\varepsilon_{ox}\eta\right) T_{ox} W_{dm}}$$

Increasing Challenges With Leakage

- Leakage temperature sensitivity
  - Worst case delay occurs at low temperature
  - Thermal runaway possible if leakage is not minimized



Delays decrease at Low V, High T_j



Potential for thermal instability

Leakage Reduction

- Minimize:
  - Leakage in standby mode
  - Leakage in active mode
  - Performance penalty
  - Area penalty
  - Dynamic power overhead
  - Additional circuit design / cell libraries
  - Impact on design flow

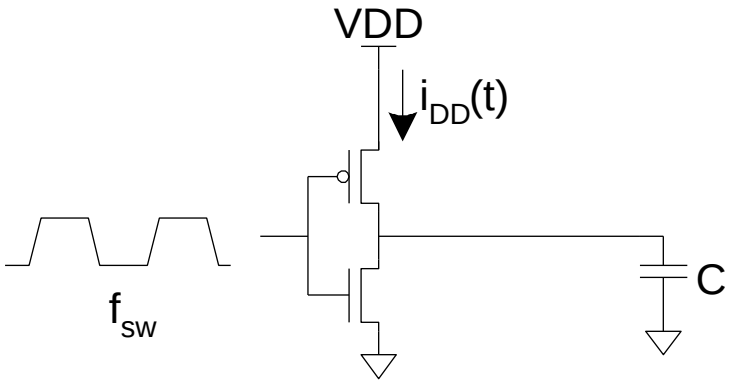
Dynamic Power

$$P_{\text{dynamic}} = \frac{1}{T}\int_0^T i_{DD}(t)V_{DD}\,dt$$

$$= \frac{V_{DD}}{T}\int_0^T i_{DD}(t)\,dt$$

$$= \frac{V_{DD}}{T}\left[Tf_{\text{sw}}CV_{DD}\right]$$

$$= CV_{DD}{}^2 f_{\text{sw}}$$



Activity Factor

- Suppose the system clock frequency = f
- Let $f_{sw} = P_{trans} f_{clock}$, where P_{trans} = activity factor
  - If the signal is a clock, $P_{trans} = 1$
  - If the signal switches once per cycle, $P_{trans} = ½$
  - Static gates:
    - Depends on design, but typically $P_{trans} = 0.1$
- Dynamic power: $P_{\text{dyn}} = P_{trans} C V_{DD}^{\ 2} f_{clock}$

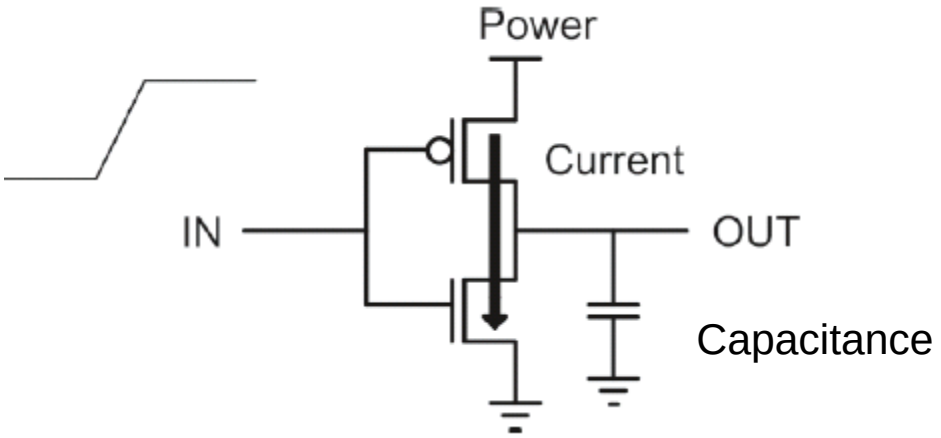
$$\Rightarrow P_{\text{dyn}} = C_{eff} V_{DD}^{\ 2} f_{clock} \quad \text{where } C_{eff} = P_{trans} C$$

- Switching power is not a function of the transistor size, but only a function of the load capacitance and switching activities

Power consumption in Transistors

$$P_{dyn} = C_{eff} \times V_{dd}^2 \times f_{clock} + t_{sc} \times V_{dd} \times I_{peak} \times f_{clock}$$

where t_{sc} is the time duration of the short circuit, and I_{peak} is the total internal switching current.



- As long as the ramp time of the input signal is kept small the short circuit current, also called the crowbar current, occurs only for a very short time and therefore the total dynamic power will be dominated by the switching power, i.e.: $P_{dyn} = C_{eff} \times V_{dd}^2 \times f_{clock}$

SYNOPSYS®

How to reduce Power?

- There are architectural, logic design, and circuit design techniques
- Focus on the voltage and frequency components of the equation, and reducing the data dependent switching activities
- Reduction of voltage has a significant effect on power but need to consider speed degradation

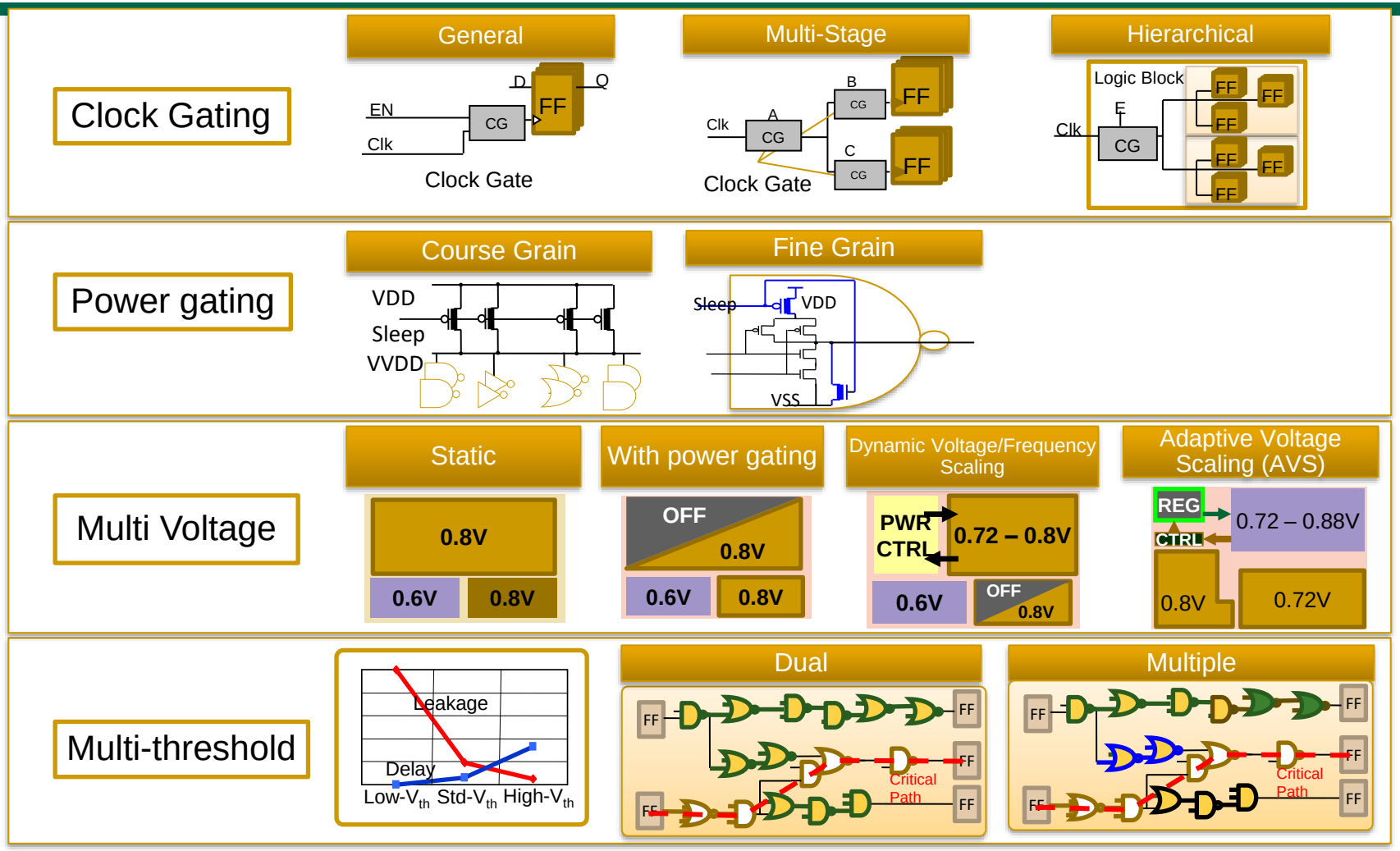
Few Techniques to reduce power

- Use lower supply voltage for slow blocks, such as peripherals (aka multi-voltage method)
- Use variable supply voltage for processors, higher voltage for tasks requiring peak performance, and lower supply voltage for lower performance tasks (aka voltage scaling method)
- Clock gating to reduce the frequency to zero

Conflict between Dynamic and Static Power

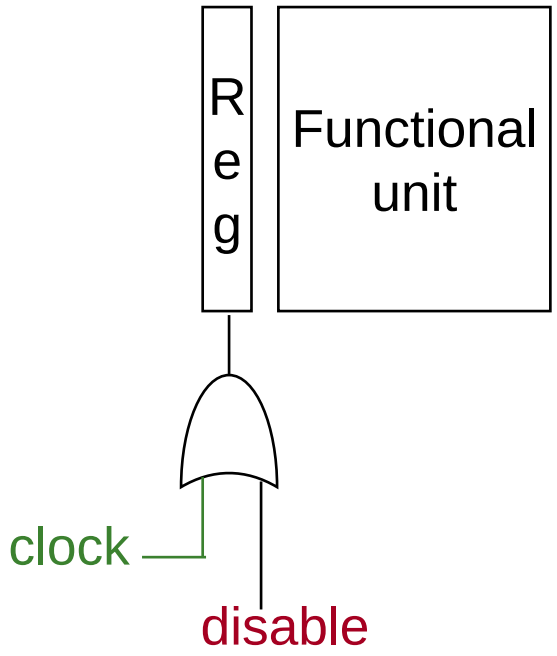
- Sub-threshold leakage occurs when a CMOS gate is not turned completely off
- Reducing voltage leads to reduction in dynamic power but to make up for performance reduction, threshold should be recued as well
- However reduction in threshold leads to exponential increase in leakage current and therefore the static power consumption
- There is therefore a conflict between reducing the dynamic power and increase in static power

LPD Techniques



Clock Gating

- Most popular method for power reduction of clock signals and functional units
- Gate off clock to idle functional units
  - e.g., floating point units
  - need logic to generate disable signal
    - increases complexity of control logic
    - consumes power
    - timing critical to avoid clock glitches at OR gate output
  - additional gate delay on clock signal
    - gating OR gate can replace a buffer in the clock distribution tree



Clock Gating

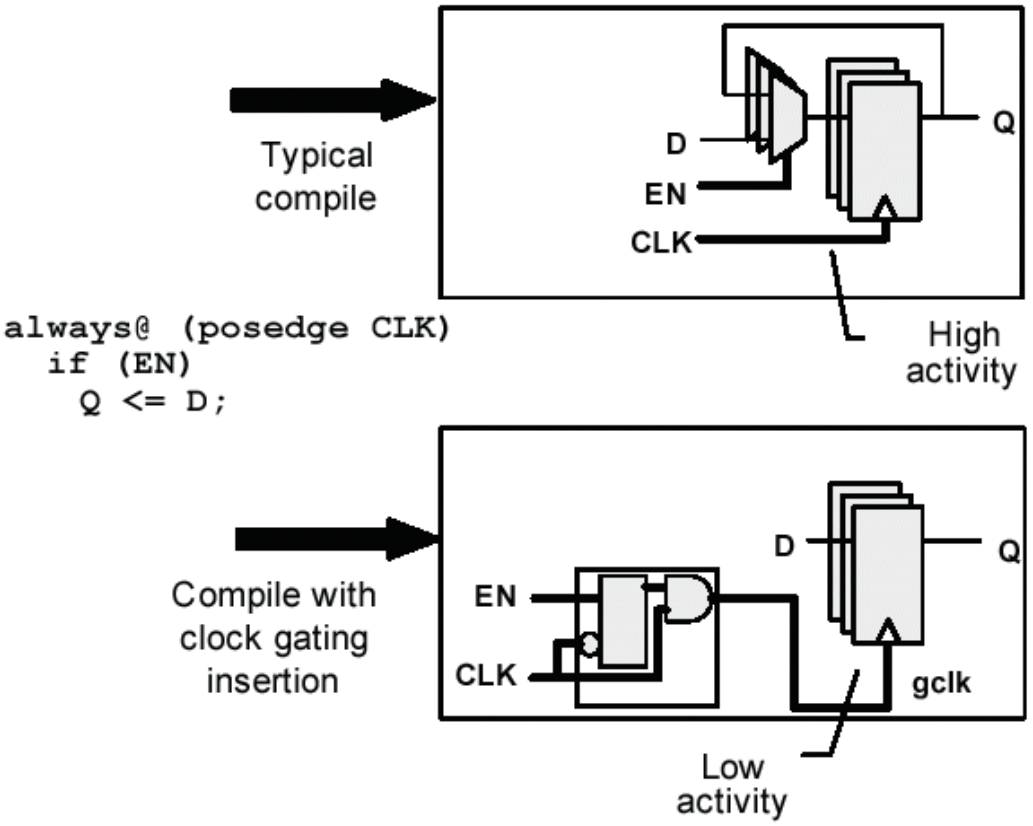
- A significant fraction of the dynamic power in a chip is in the distribution network of the clock.
- Up to 50% or even more of the dynamic power can be spent in the clock buffers. The reason is:
  - Clock buffers have the highest toggle rate in the system
  - Typically there are lots of clock buffers in a design
  - Clock buffers often have a high drive strength to minimize clock delay
- Additionally the flops receiving the clock dissipate some dynamic power even if the input and output remain the same.

The most common technique to reduce this power is to turn clocks off when they are not required. This approach is known as clock gating.

How Clock Gating Works?

- Modern design tools support automatic clock gating: i.e. they can identify circuits where clock gating can be inserted without changing the function of the logic.

In the original RTL, the register is updated or not depending on a variable (EN). The same result can be achieve by gating the clock based on the same variable.



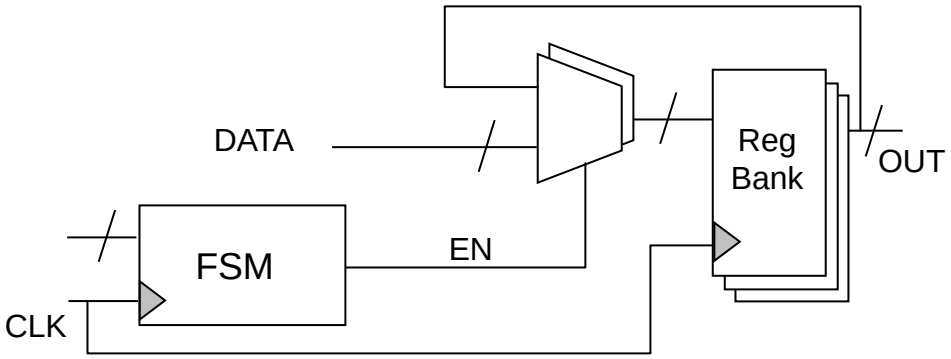
SYNOPSYS

How Clock Gating Works?

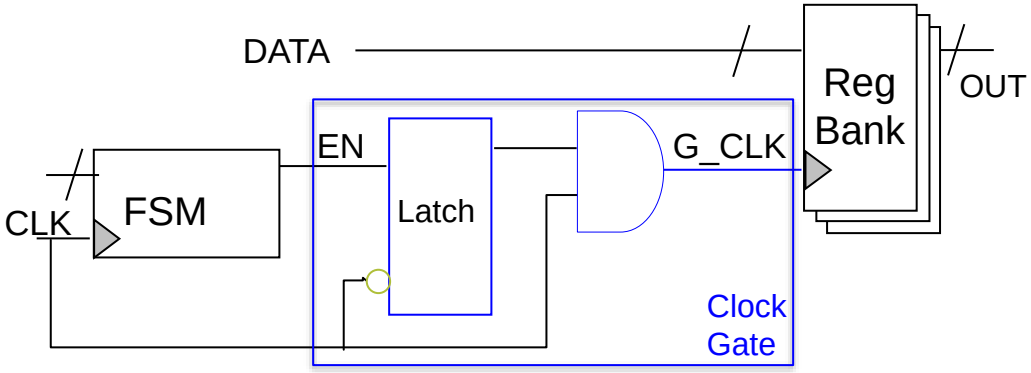
- If the registers involved are single bits, then a small savings occurs. If they are, say, 32 bit registers, then one clock gating cell can gate the clock to all 32 registers (and any buffers in their clock trees). This can result in considerable power savings.
- In the early days of RTL design, engineers would code clock gating circuits explicitly in the RTL. This approach is error prone – it is very easy to create a clock gating circuit that glitches during gating, producing functional errors.
- Today, most libraries include specific clock gating cells that are recognized by the synthesis tool. The combination of explicit clock gating cells and automatic insertion makes clock gating a simple and reliable way of reducing power. No change to the RTL is required to implement this style of clock gating.

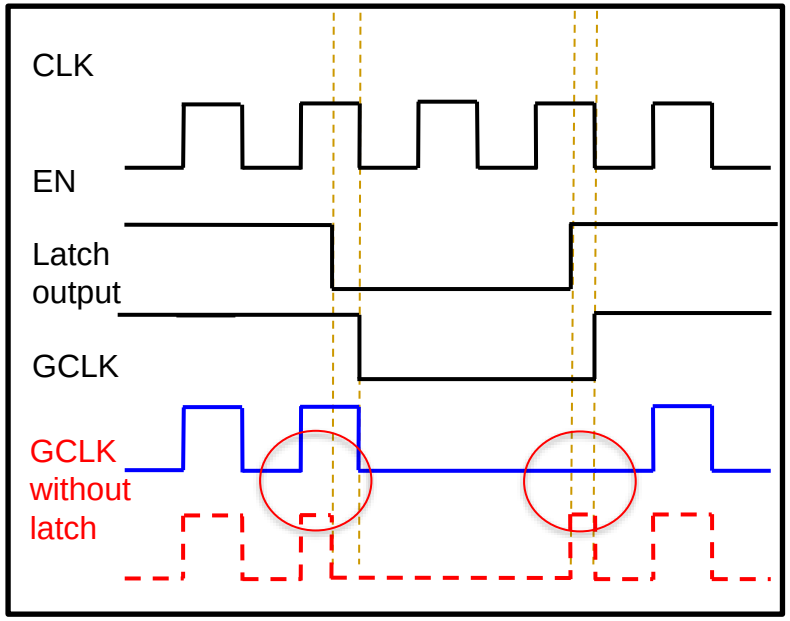
Clock Gating Bases

Original Circuit

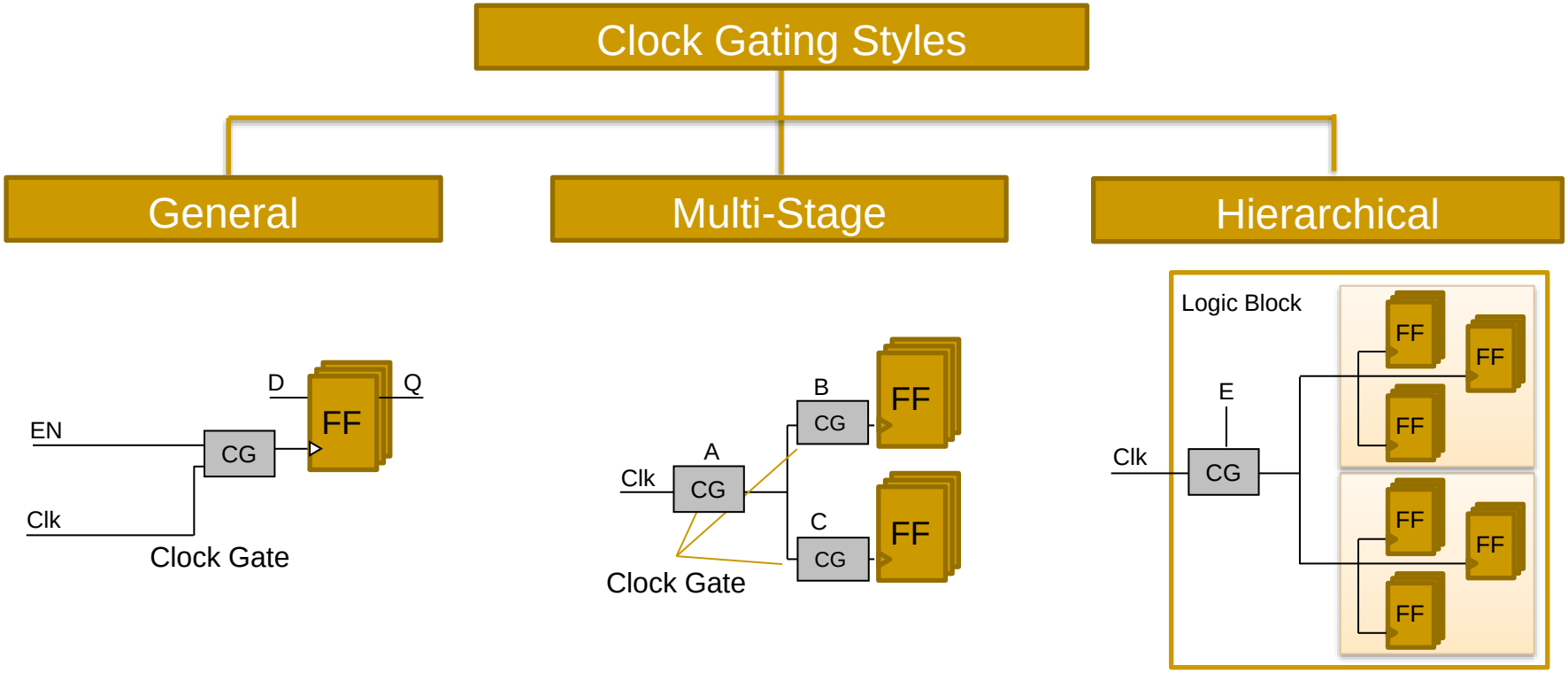


Circuit with Clock Gating





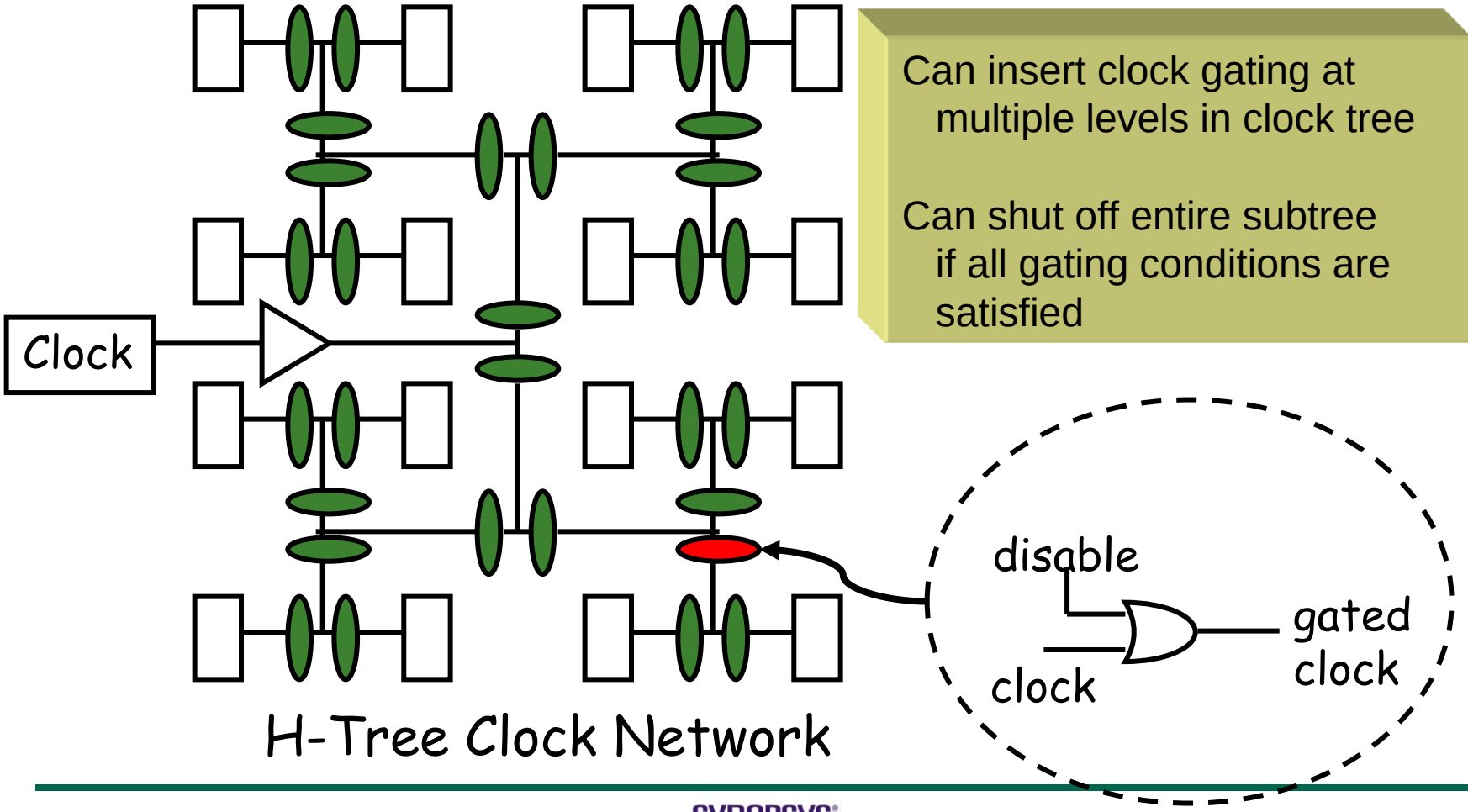
Clock Gating Styles



SYNOPSYS®

Gated Clock Distribution

If the paths are perfectly balanced, clock skew is zero



Can insert clock gating at multiple levels in clock tree

Can shut off entire subtree if all gating conditions are satisfied

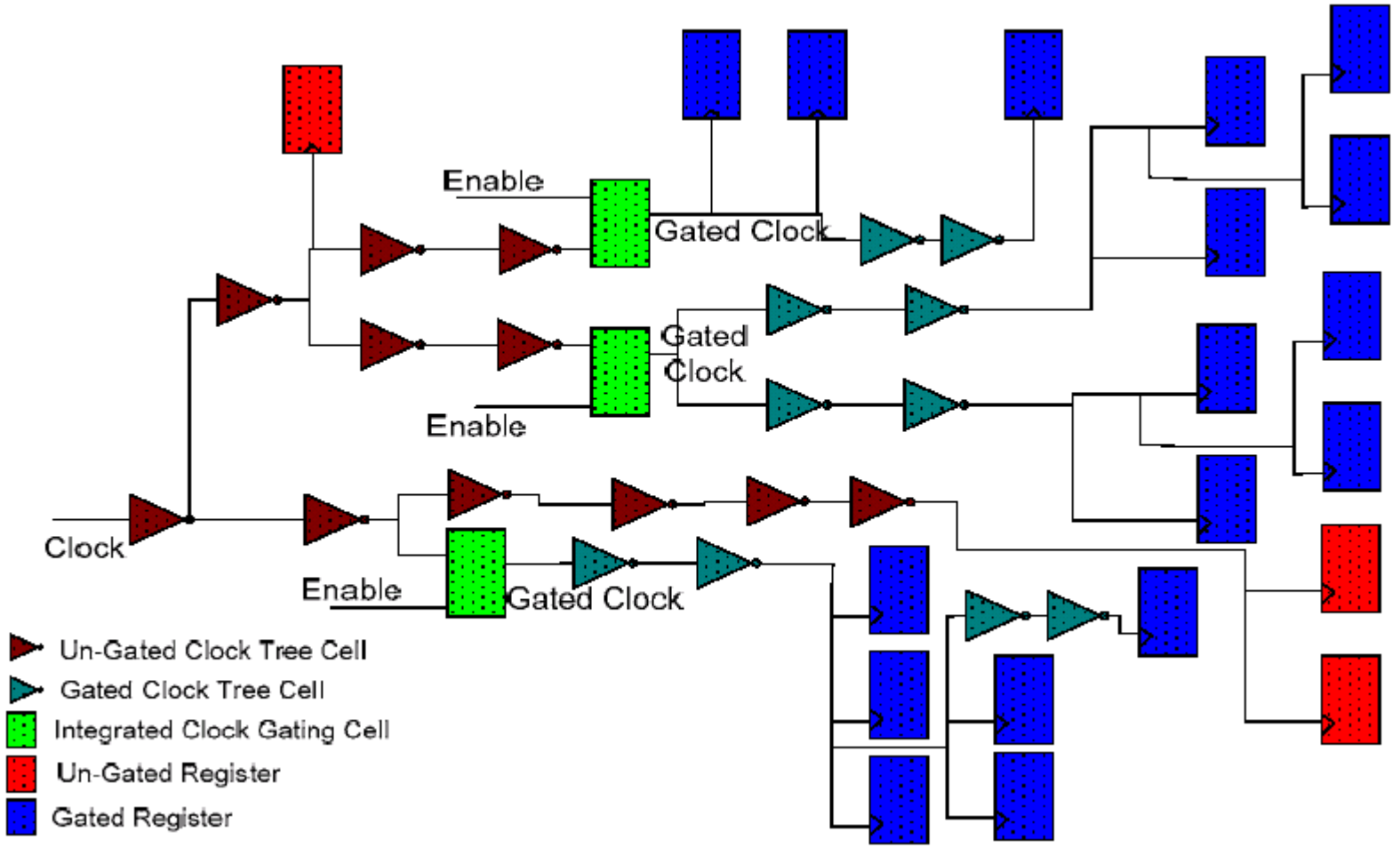
SYNOPSYS®

Example

- In a power reduction project, an existing 180nm chip without clock gating was re-implemented in the same technology with clock gating. Only minor changes in the logic were implemented (some small blocks were removed and replaced by other blocks, for a small net increase in functionality).
- Pokhrel reports an area reduction of 20% and a power savings of 34% to 43%, depending on the operating mode.
- Power measurements were made on the whole chip when the processor was in IDLE mode; that is, the processor was turned off.) The power measurements are from actual silicon.
- The area savings is due to the fact that a single clock gating cell takes the place of multiple muxes.

Ref: Pokhrel, K. “Physical and Silicon Measures of Low Power Clock Gating Success: An Apple to Apple Case Study”, SNUG, 2007 http://www.snug-universal.org/cgibin/search/search.cgi?San+Jose,+2007.

Clock Gating Example



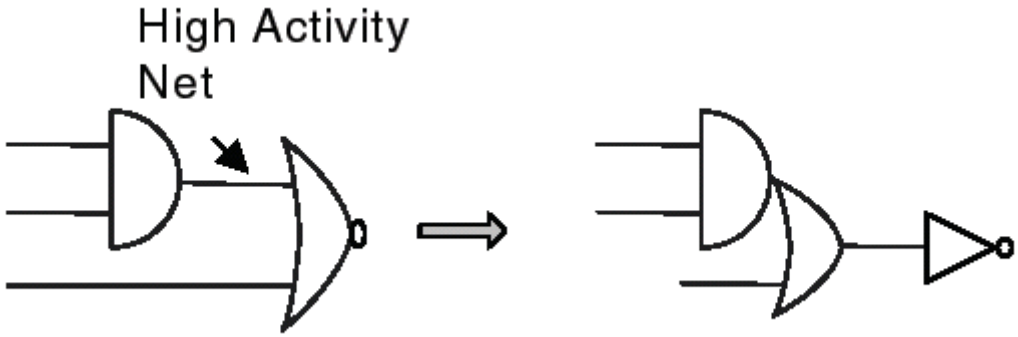
Ref: Pokhrel, K. “Physical and Silicon Measures of Low Power Clock Gating Success: An Apple to Apple Case Study”, SNUG, 2007 http://www.snug-universal.org/cgibin/search/search.cgi?San+Jose,+2007.

More Observations

- After some analysis and experiments, the team decided to use clock gating only on registers with a bit-width of at least three. They found that clock gating on one-bit registers was not power or area efficient.
- Much of the power savings was due to the fact that the clock gating cells were placed early in the clock path. Approximately 60% of the clock buffers came after the clock gating cell, and so had their activity reduce to zero during gating.

Gate Level Power Optimization

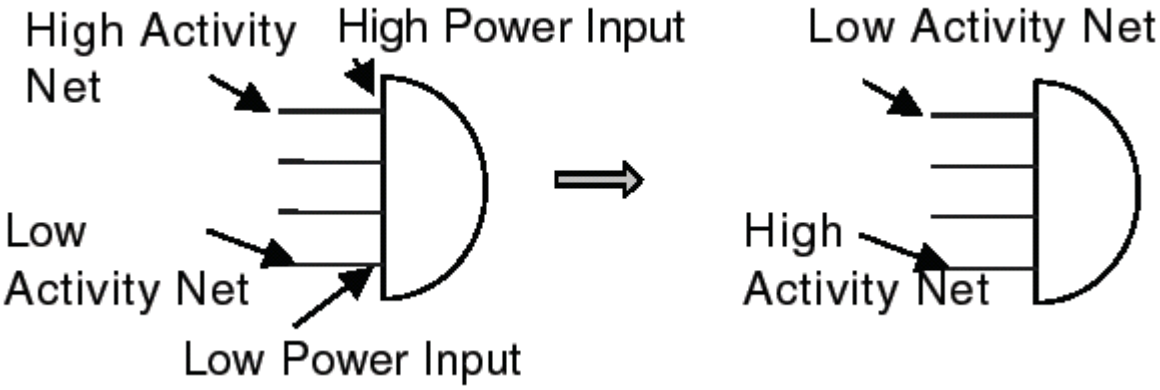
- In addition to clock gating, there are a number of logic optimizations that the tools can perform to minimize dynamic power.



- Example: In above figure, an AND gate output has a particularly high activity. It is possible to re-map the two gates to an AND-OR gate plus an inverter, so the high activity net becomes internal to the cell. Now the high activity node (the output of the AND gate) is driving a much smaller capacitance, reducing dynamic power.

SYNOPSYS®

Gate Level Power Optimization



- Example: In the above figure, an AND gate has been initially mapped so that a high activity net is connected to a high power input pin, and a low activity net has been mapped to a low power pin. For multiple input gates *there can be a significant difference in the input capacitance* - and hence the power - for different pins. By remapping the inputs so the high activity net is connected to the low power input, the optimization tool can reduce dynamic power.

Other Gate Level Power Optimization Techniques

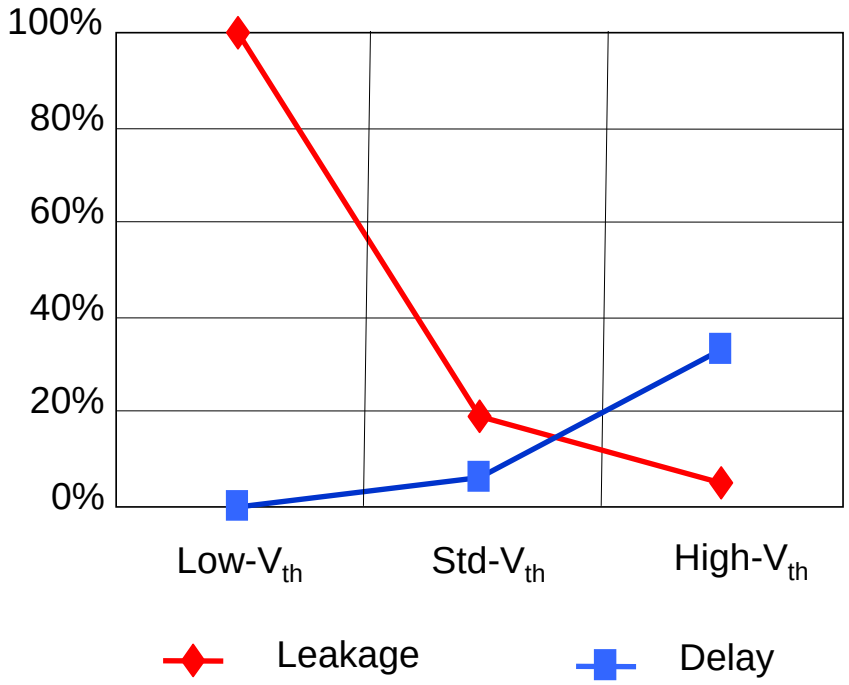
- Other examples of gate level power optimization include cell sizing and buffer insertion.
- In cell sizing, the tool can selectively increase and decrease cell drive strength throughout the critical path to achieve timing and then reduce dynamic power to a minimum.
- In buffer insertion, the tool can insert buffers rather than increasing the drive strength of the gate itself. If done in the right situations, this can result in lower power.
- Like clock gating, gate level power optimization is performed by the implementation tools, and is transparent to the RTL designer.

Multi-Threshold Logic

- As geometries have shrunk to 130nm, 90nm, and below, using libraries with multiple V_T has become a common way of reducing leakage current.
- Many libraries today offer two or three versions of their cells: Low V_T, Standard V_T, and High V_T.
- The implementation tools can take advantage of these libraries to optimize timing and power simultaneously.

Multi-Threshold Design Bases

- Voltage Thresholds affect FinFET leakage
  - Higher Vth cells have low leakage power but are slow
  - Lower Vth cells have high leakage power but are fast



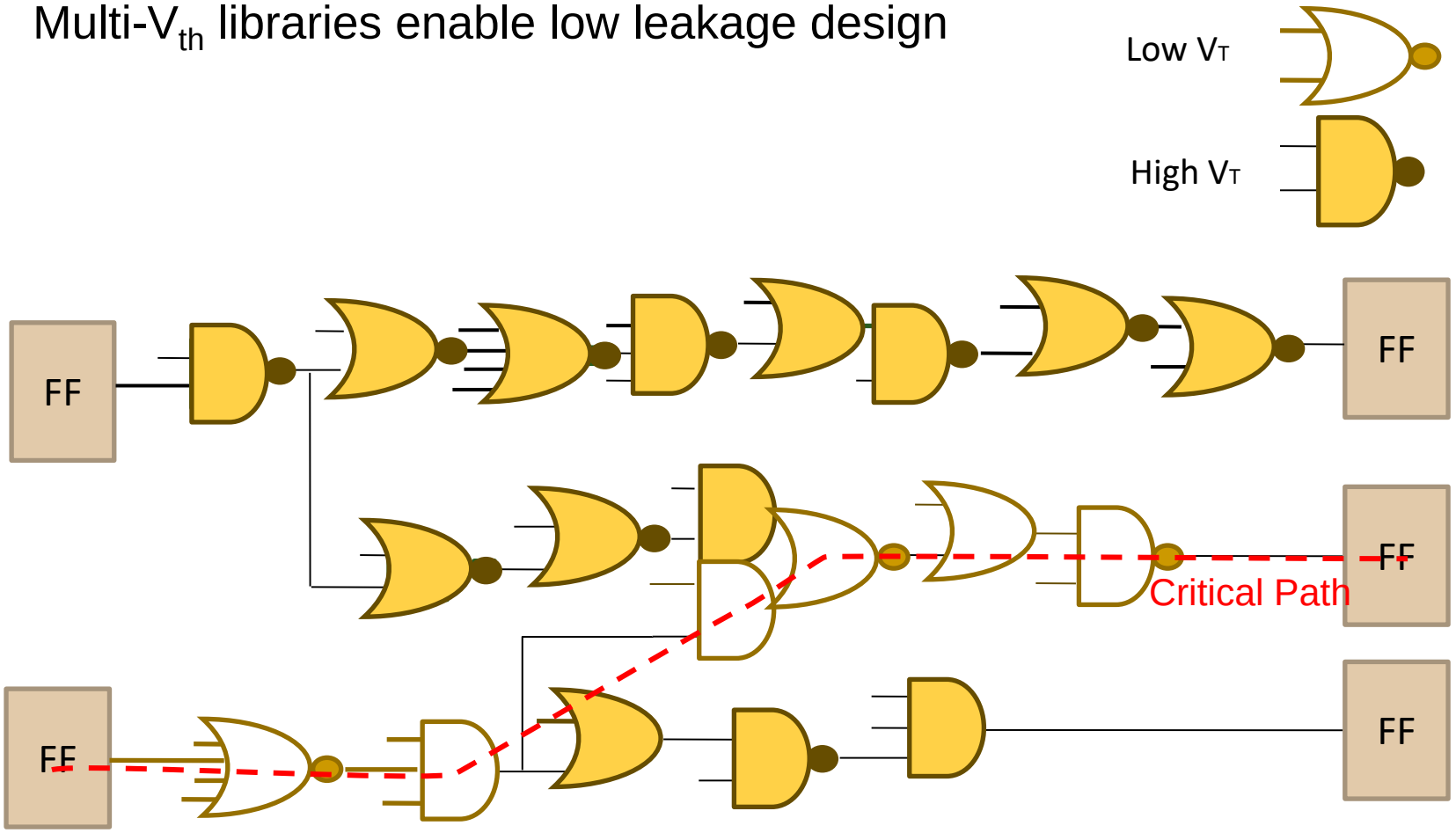
SYNOPSYS®

How the scheme works?

- Usually there is a minimum performance which must be met before optimizing power.
- In practice this usually means synthesizing with the high performance, high leakage library first and then relaxing back any cells not on the critical path by swapping them for their lower performing, lower leakage equivalents.
- If minimizing leakage is more important than achieving a minimum performance then this process can be done the other way around: we can target the low leakage library first and then swap in higher performing, high leakage equivalents in speed critical areas.

Multi-Threshold Design Example

- Multi-V_{th} libraries enable low leakage design



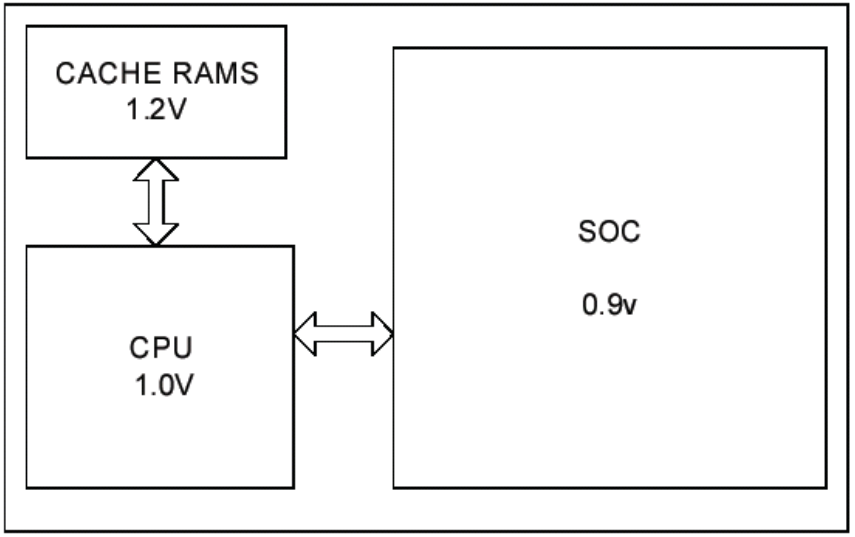
SYNOPSYS®

Multi VDD

- Since dynamic power is proportional to V_{DD}^2, lowering V_{DD} on selected blocks helps reduce power significantly.
- Unfortunately, lowering the voltage also increases the delay of the gates in the design.

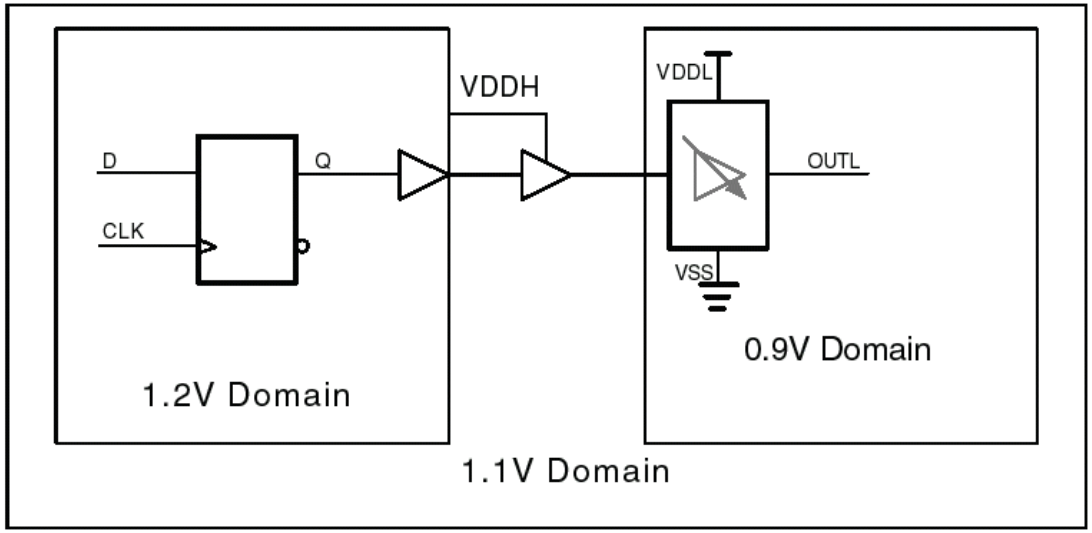
Example

- Cache RAMS are run at the highest voltage because they are on the critical timing path.
- The CPU is run at a high voltage because its performance determines system performance. But it can be run at a slightly lower voltage than the cache and still have the overall CPU subsystem performance determined by the cache speed.
- The rest of the chip can run at a lower voltage still without impacting overall system performance.



Challenges

- Mixing blocks at different VDD supplies adds some complexity to the design:
  - need to add IO pins to supply the different power rails
  - need a more complex power grid
  - need level shifters on signals running between blocks



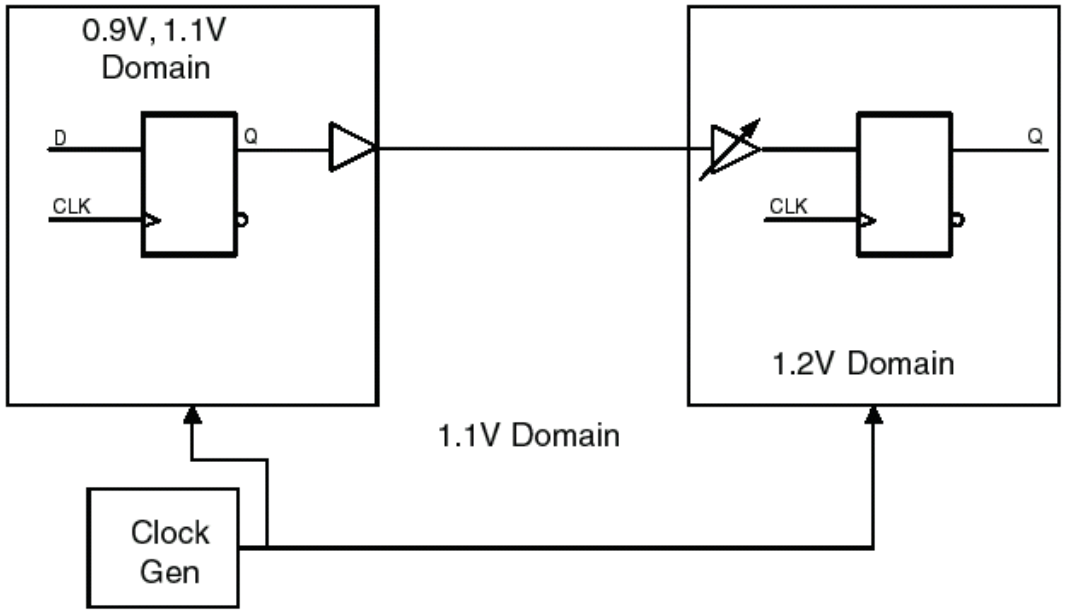
SYNOPSYS®

Timing Issues in Multi-Voltage Designs; Clocks

- Routing clocks across different power domains means that they have to go through level shifters.
- This clearly complicates automation – the clock tree synthesis tools need to understand level shifters and automatically insert them in the appropriate places.
- With multiple level voltage scaling, clock distribution gets even more complex.

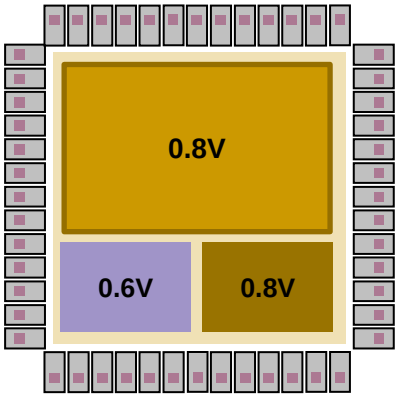
Timing Issues in Multi-Voltage Designs;
Clocks

- The clock buffers in the multi-level domain will sometimes be powered at 0.9V and sometimes at 1.1V. Under which conditions do we attempt to minimize clock skew relative to the clock in the 1.2V domain?

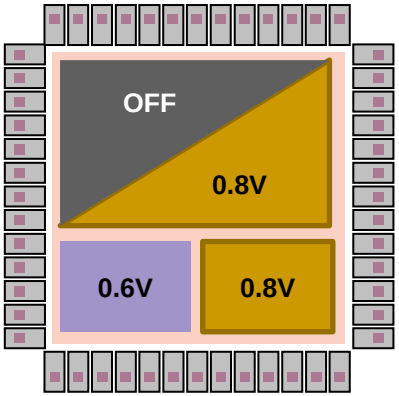


- The solution is that optimization and timing analysis must be done simultaneously for both situations, to assure that timing will be met for both conditions.

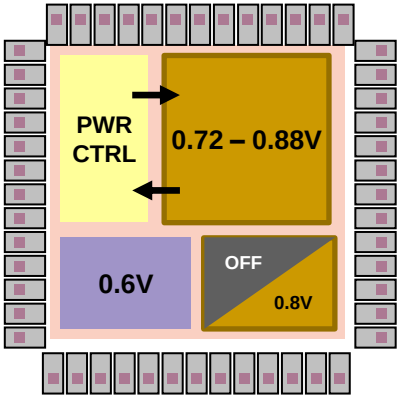
Multi-Voltage Design Techniques



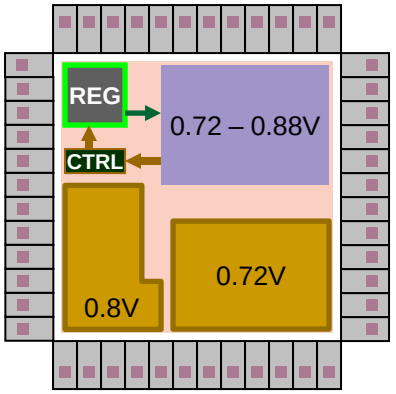
Static Multi Voltage (MV)



MV with power gating



Dynamic Voltage/Frequency Scaling (DVFS)



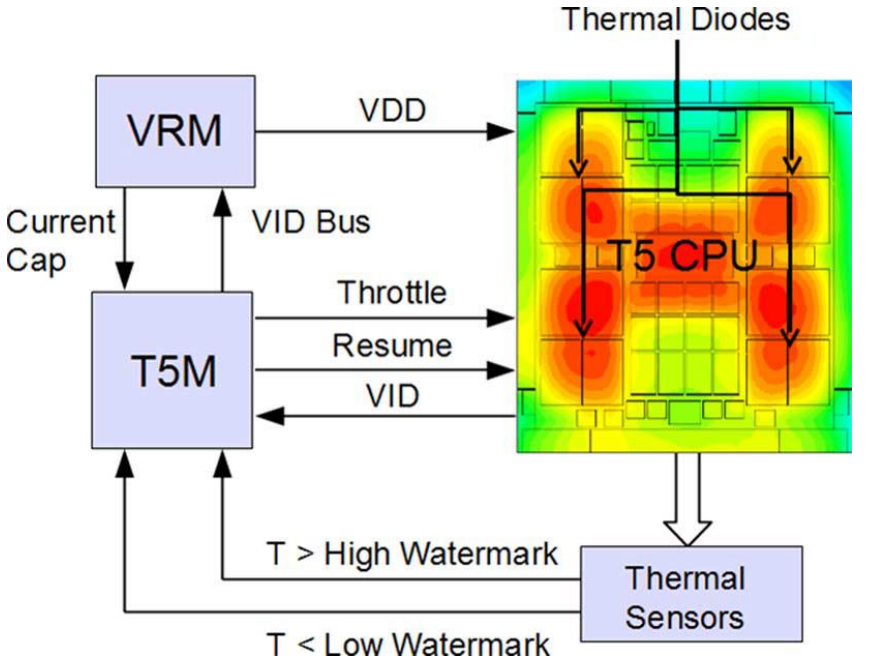
Adaptive Voltage Scaling (AVS)

SYNOPSYS®

A 3.6 GHz 16-Core SPARC SoC Processor in 28 nm

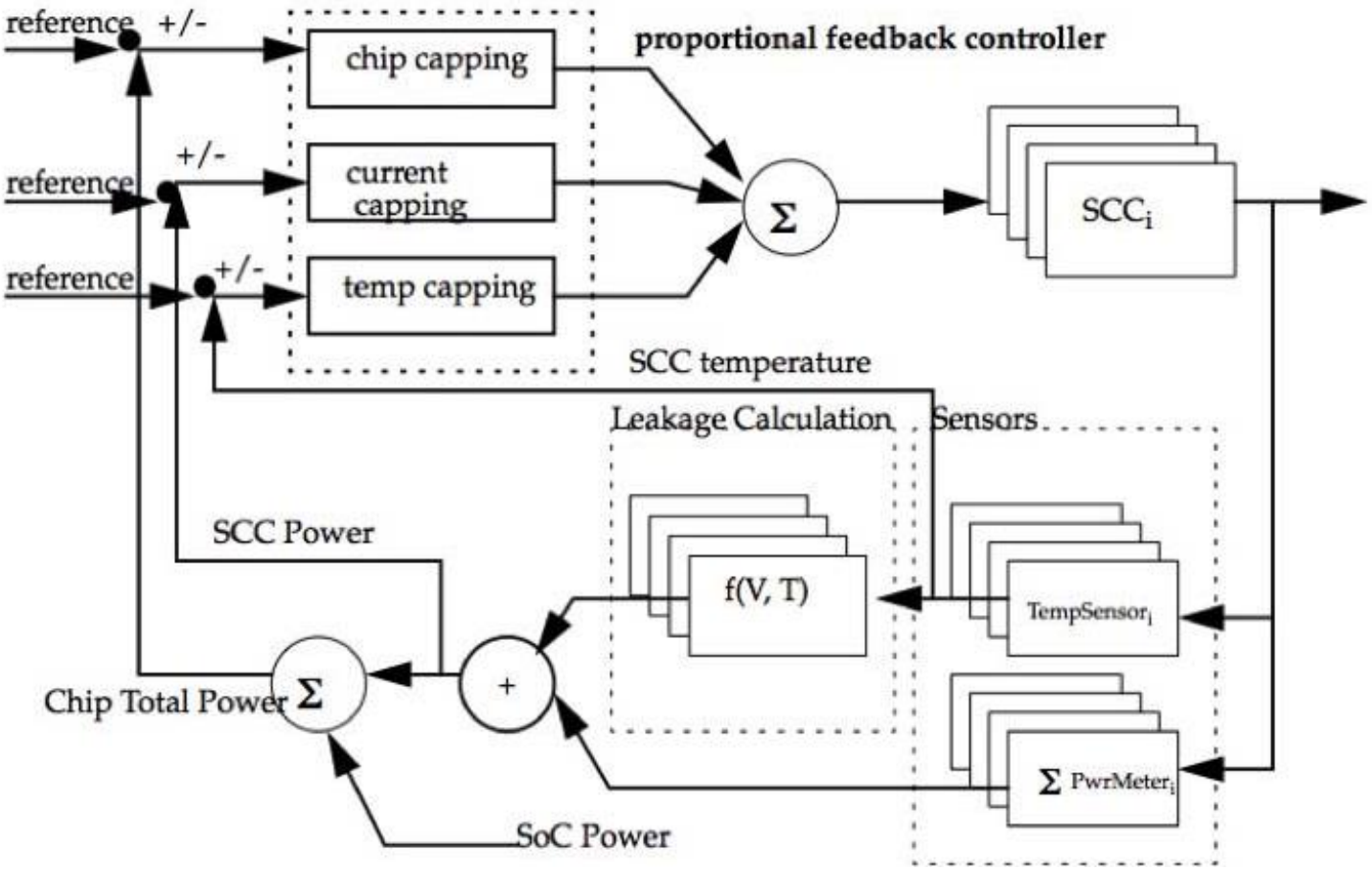
- SPARC cores, L3 cache and crossbar are synchronously clocked and are all subject to DVFS equally
- Hardware P-state table of Voltage IDs (VID) 0..31(frequency modes)
- Digital voltage regulator module (VRM) changes VDD
- Frequency dividers are inside core
- The T5M asserts VIDs to the VRM on a dedicated bus

Dynamic voltage frequency scaling system control architecture



Hart, Jason M., et al. "A 3.6 GHz 16-Core SPARC SoC Processor in 28 nm."*Solid-State Circuits, IEEE Journal of* 49.1 (2014): 19-31.

Fine-Grained Adaptive Power Management of the SPARC M7 Processor



Krishnaswamy V. et al. 4.3 Fine-grained adaptive power management of the SPARC M7 processor //Solid-State Circuits Conference-(ISSCC), 2015 IEEE International. – IEEE, 2015. – C. 1-3.

Dynamic and Leakage Power Profiles

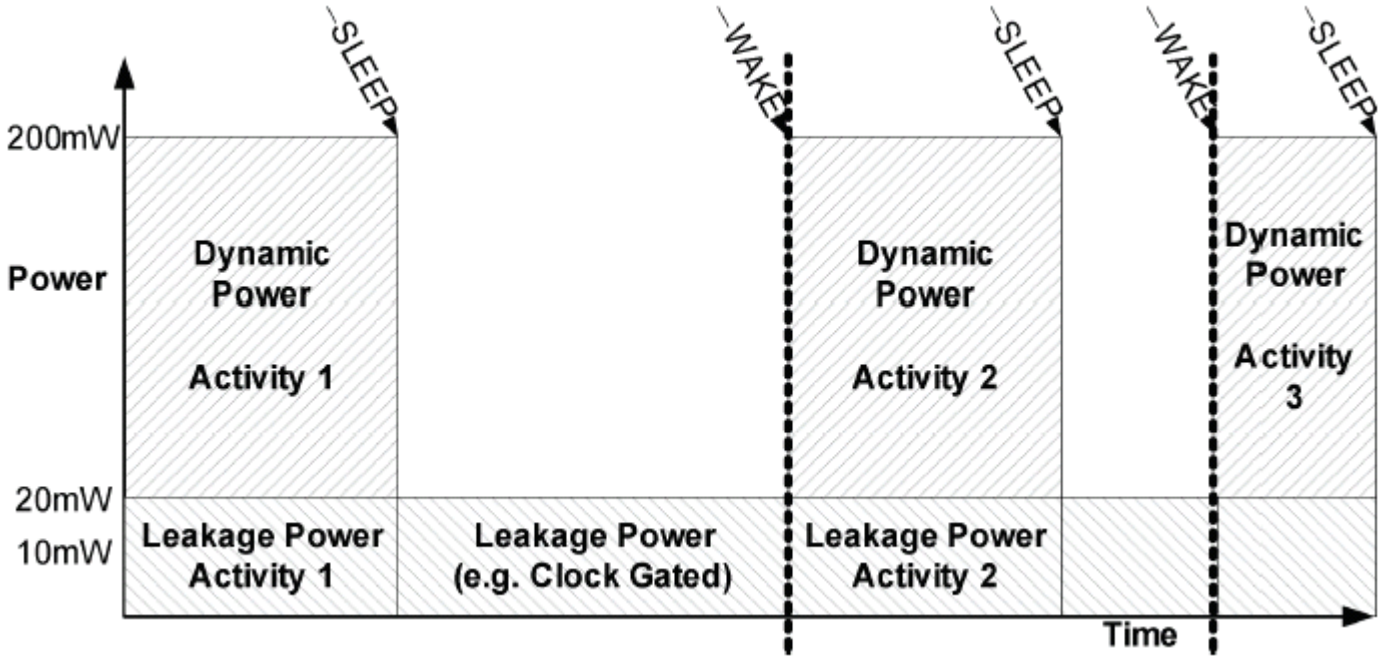
- The basic strategy of power gating is to provide two power modes: a low power mode and an active mode. The goal is to switch between these modes at the appropriate time and in the appropriate manner to maximize power savings while minimizing the impact to performance.
- The power reduction techniques described earlier do not affect the functionality of the design and do not require changes to the RTL. They can be handled fairly transparently from a design and implementation and perspective; power gating is more invasive than clock-gating in that it affects inter-block interface communication and adds significant time delays to safely enter and exit power gated modes.
- Shutting down power to a block of logic may be scheduled explicitly by control software as part of device drivers or operating system idle tasks. Alternatively it may be initiated in hardware by timers or system level power management controllers.

Architectural Trade-offs in Power Gating

- Amount of possible saving in leakage power
- Entry and exit time penalties incurred
- The energy dissipated due to entering and leaving such leakage saving modes
- The activity profile (proportion and frequency of times asleep or active)

Terminology for the Entry and Exit from Power Modes

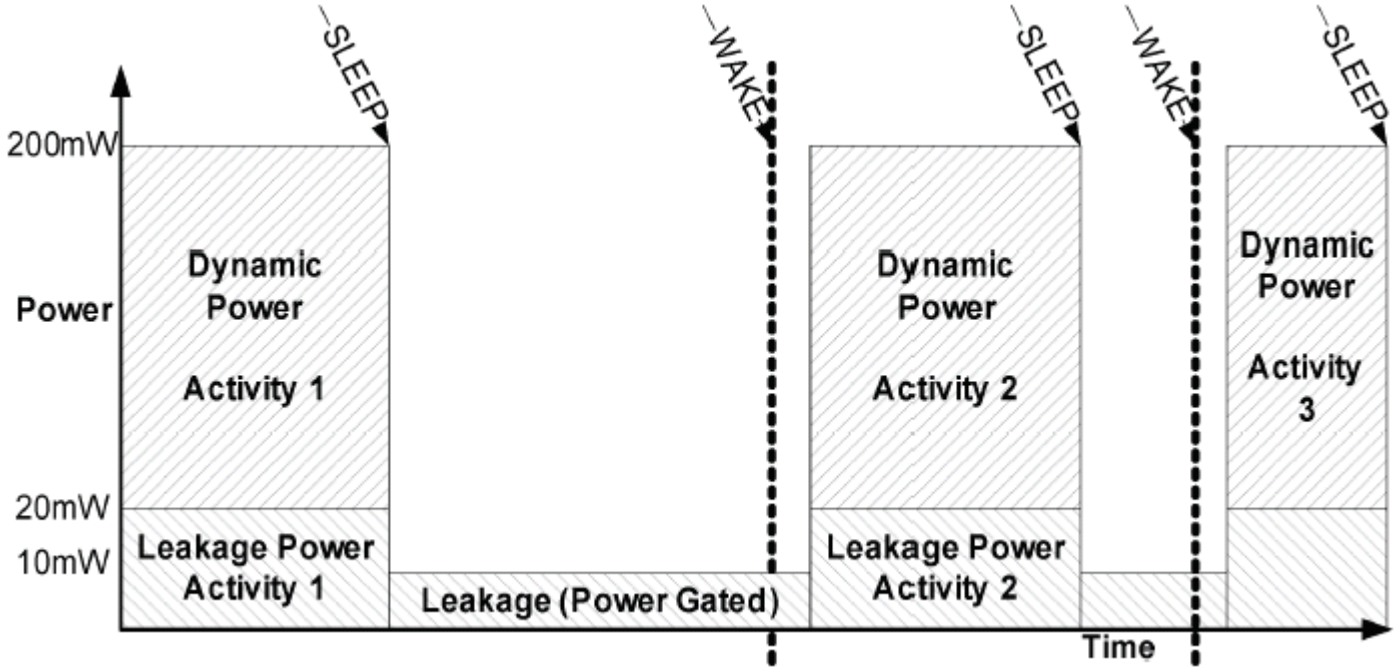
- **SLEEP** events initiate entry to the low power mode
- **WAKE** events initiate return to active mode



An example activity profile for a sub-system using clock gating to reduce power.

SYNOPSYS®

Power Gating Example



An example activity profile for the same sub-system with basic power gating implemented. The response time between the WAKE event and having clocks running may be significant and cannot be ignored at the system design level.

Realistic Profile with Power Gating

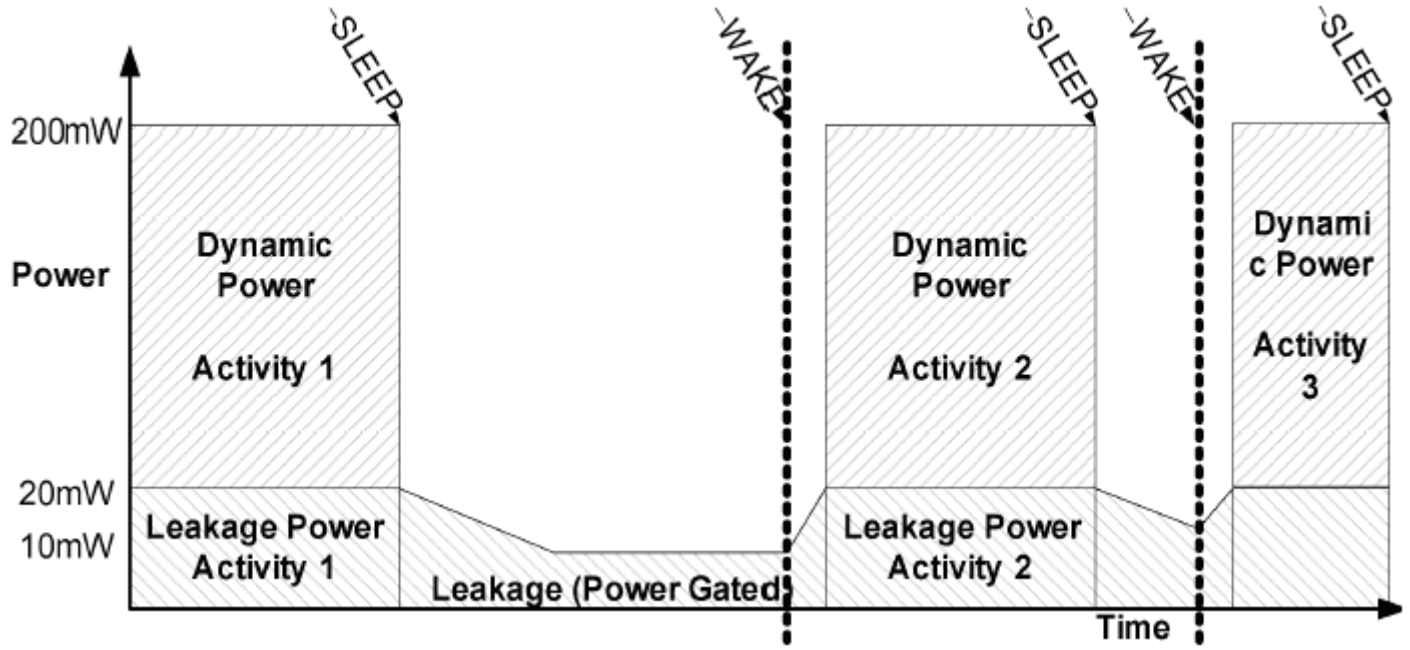
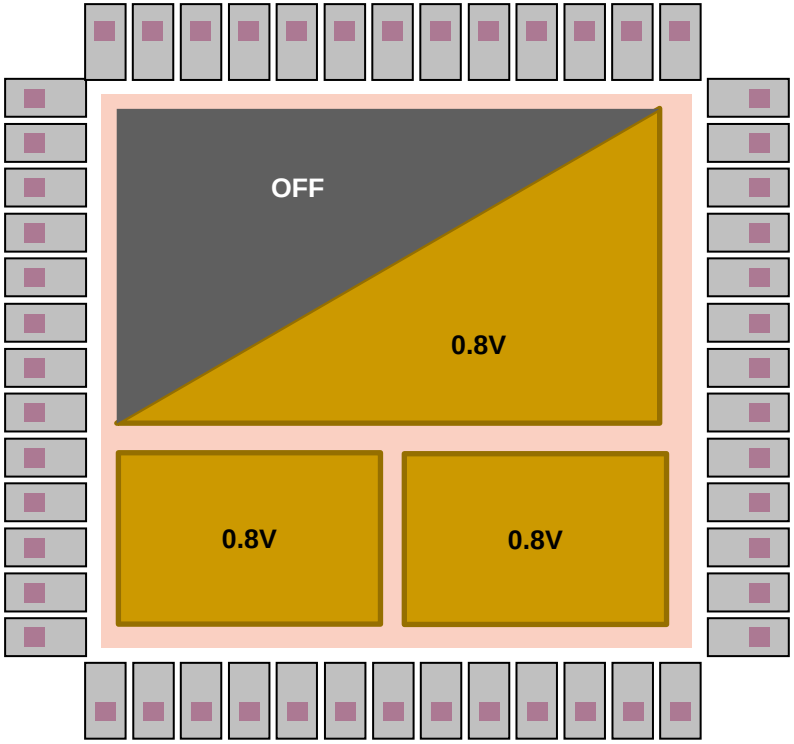


Figure shows Realistic Profile with Power Gating. As shown the leakage power savings are not perfect and instantaneous; the full leakage power savings take some time to reach target levels. This is due partly to the (hotter) thermal profile of the preceding activity and partly to the non-ideal nature of the power-gating technology. Therefore the achievable savings are compromised to some extent.

Power Gating Bases

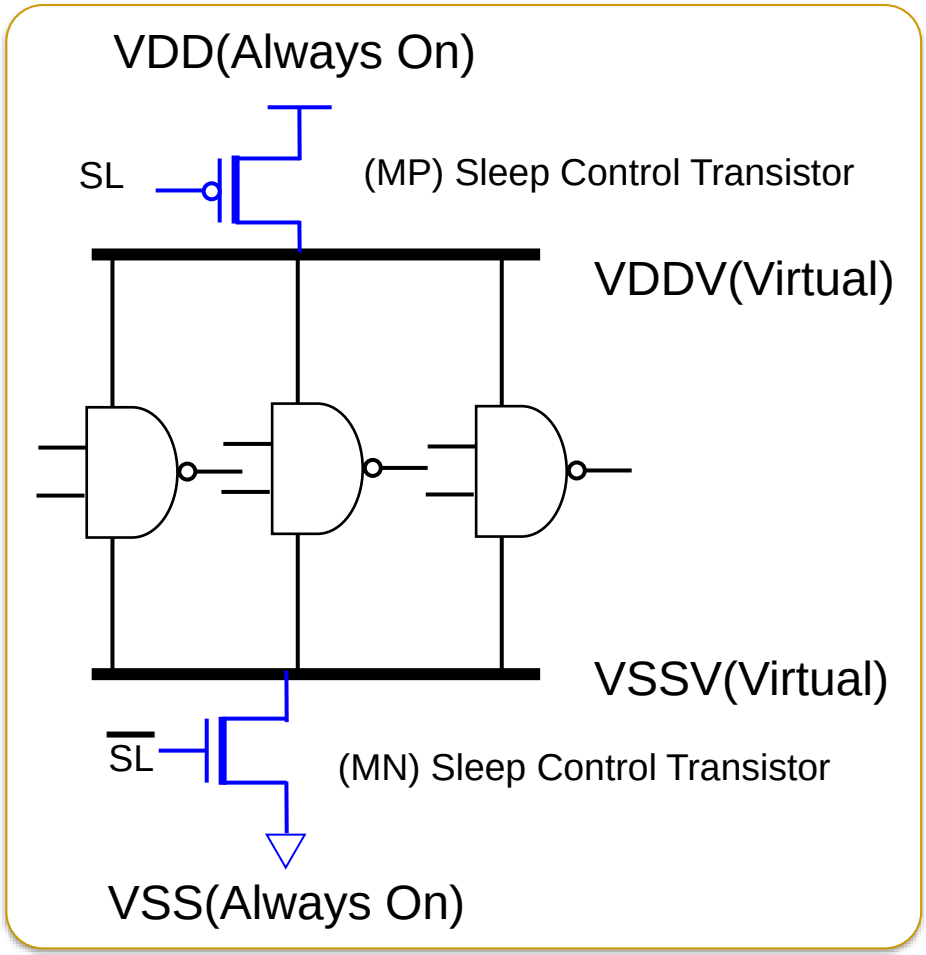


Power gating

SYNOPSYS®

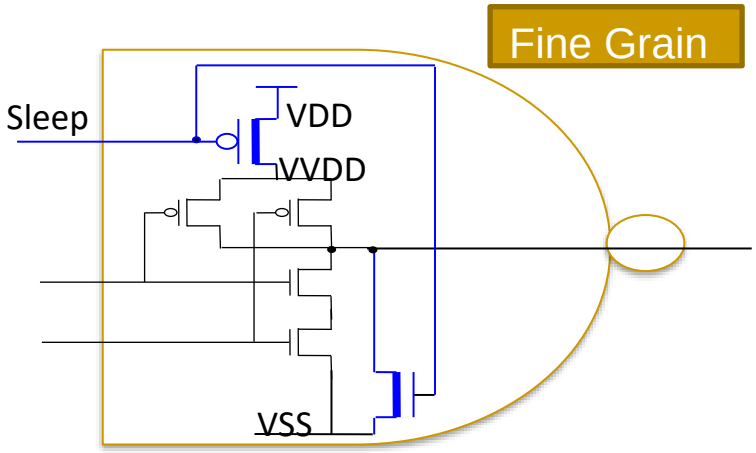
Power Gating Details

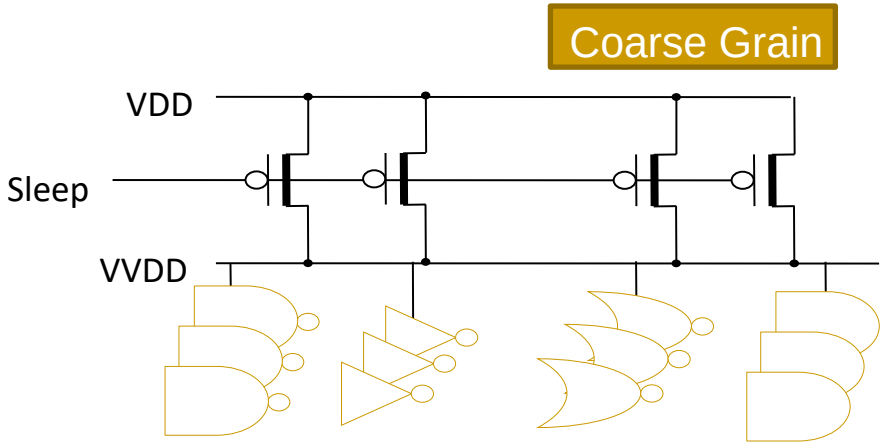
- In active mode: SL=0, MP and MN are “on”
  - VDDV and VSSV almost function as VDD and VSS.
- In standby mode: SL=1, MP and MN are “off” and leakage is suppressed.



SYNOPSYS

Fine Grain vs. Coarse Grain

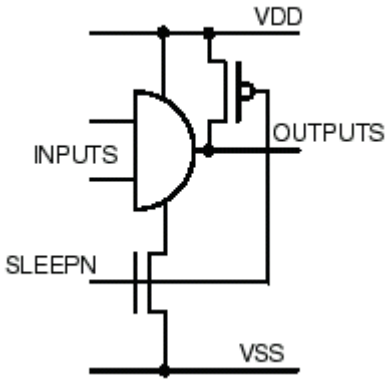




Parameter	Fine Grain	Coarse Grain
Reduce leakage	10x	50x
Wakeup time	Fast	Slow
Wakeup power	Small	Large
Library requirements	New cell library	New footer or header cell
Sensitive to PVT variation	Large	Small
Can be implemented by usual physical synthesis	Yes	No
IR-drop variations	Large	Small
Chip Area	Large	Small

Fine Grain vs. Coarse Grain

- A critical decision in power gating is how to switch power. In general, there are two approaches: fine grain power gating and coarse grain power gating.
- In fine grain power gating the switch is placed locally inside each standard cell in the library. Since this switch must supply the worst case current required by the cell, it has to be quite large in order not to impact performance. The area overhead of each cell is significant (often 2x-4x the size of the original cell).



Fine Grain AND Gate with Pull-Up

SYNOPSYS®

Advantage of Fine Grain Power Gating

- The key advantage of fine grain power gating is that the timing impact of the IR drop across the switch and the behavior of the clamp are easy to characterize as they are contained within the cell. This means that it is still possible to use a traditional design flow to deploy fine grain power gating.

Coarse Grain Power Gating

- In coarse grain power gating, a block of gates has its power switched by a collection of switch cells. The sizing of a coarse grain switch network is more difficult than a fine grain switch as the exact switching activity of the logic it supplies is not known and can only be estimated. But coarse grain gating designs have significantly less area penalty than fine grain.
- Over the last few years, there has been a strong convergence towards coarse grain power gating as the preferred method. The area penalty for fine grain power gating has just not proven worth the savings in design effort.
- Today, virtually all power gated designs use coarse grain power gating.

Power Gating Challenge

- Managing the in-rush current when the power is reconnected
- Design of the power switching fabric
- Design of the power gating controller
- Selection and use of retention registers and isolation cells
- Minimizing the impact of power gating on timing and area.
- The functional control of clocks and resets
- Interface isolation
- Developing the correct constraints for implementation and analysis
- Performing state-dependent verification for each supported power state
- Performing power state transition verification to ensure all legal state entry and exit arcs are simulated and verified
- Developing a strategy for manufacturing and production test

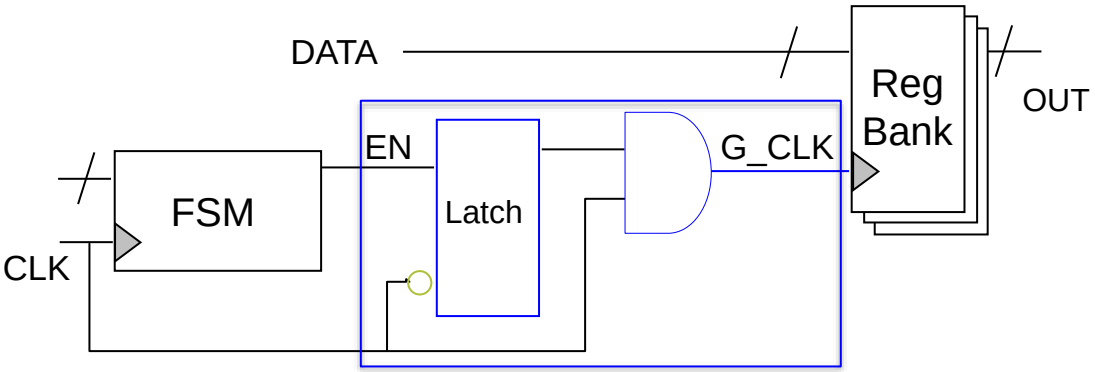
Impact of LPD on Digital Standard Cell Library

Library Requirements for LPD

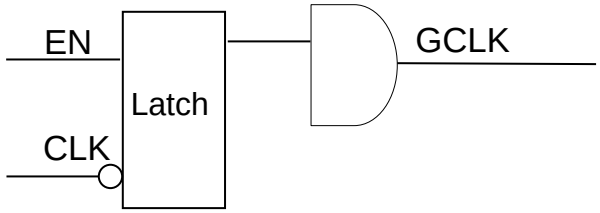
- Special cells
- Special versions of library
- Characterization in additional corners
- Additional files/attributes

Special Cell Requirements: Clock Gating

Circuit with Clock Gating

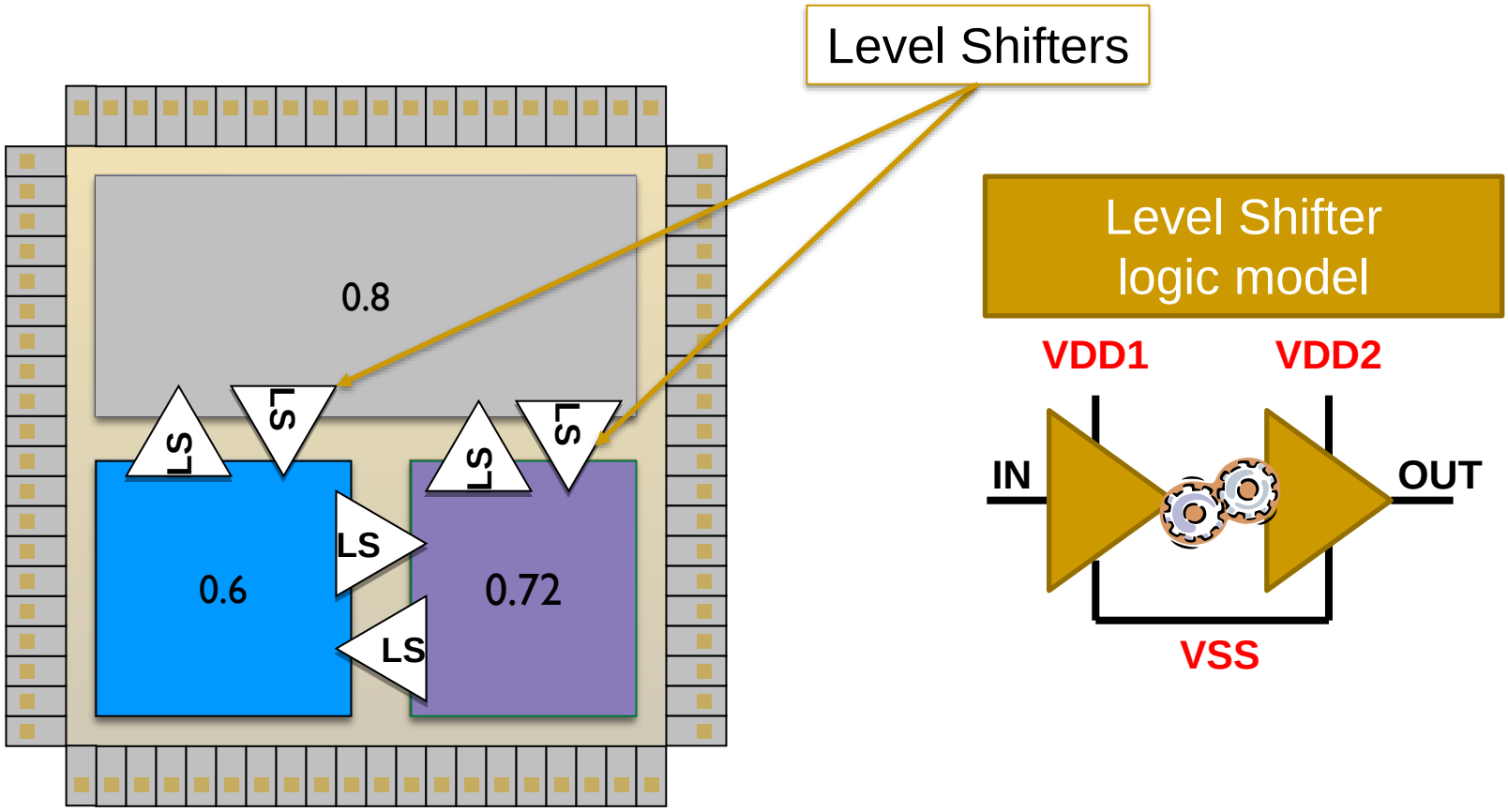


Clock Gating Cell



SYNOPSYS®

Special Cell Requirements: Multi Voltage



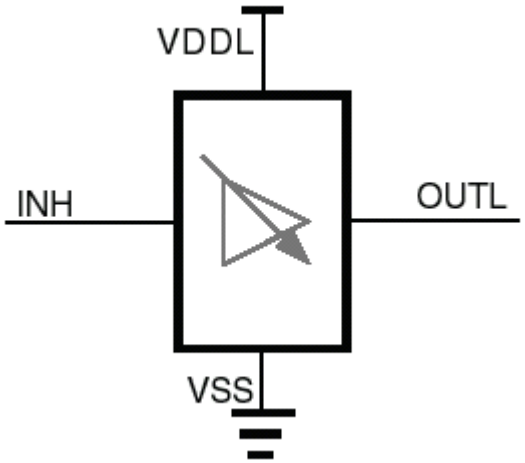
SYNOPSYS®

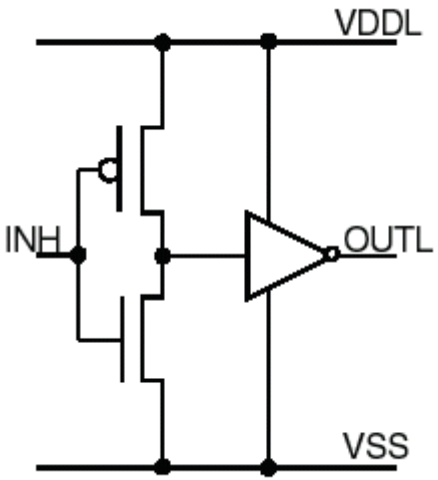
Level Shifters – High to Low

- Overdriving a CMOS input from an output buffer on a higher voltage rail does not appear to be a problem – there are no latch-up or breakdown issues, simply a “better”, faster edge compared to normal CMOS logic high or low level switching levels.
- However for safe timing closure one does need some specially identified “downshift” cells characterized specifically for this purpose.
- If specialized high-to-low level shifter cells were not provided in the library then the entire library would have to be re-characterized to allow accurate static timing analysis.
- Each gate would have to be characterized for an arbitrary input voltage swing.

Level Shifters – High to Low

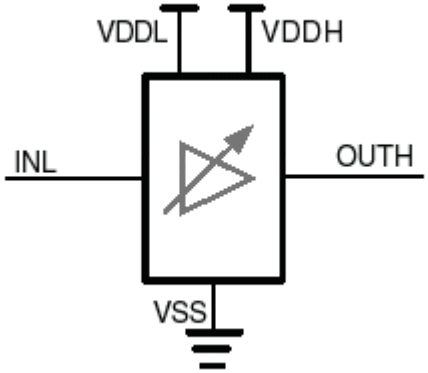
- High to low level shifters can be essentially two inverters in series. Note only one single power rail is required, which is the one from the lower or destination power domain.
- As implied by the drawing, a high-to-low level shifter only introduces a buffer delay, so its impact on timing is small.

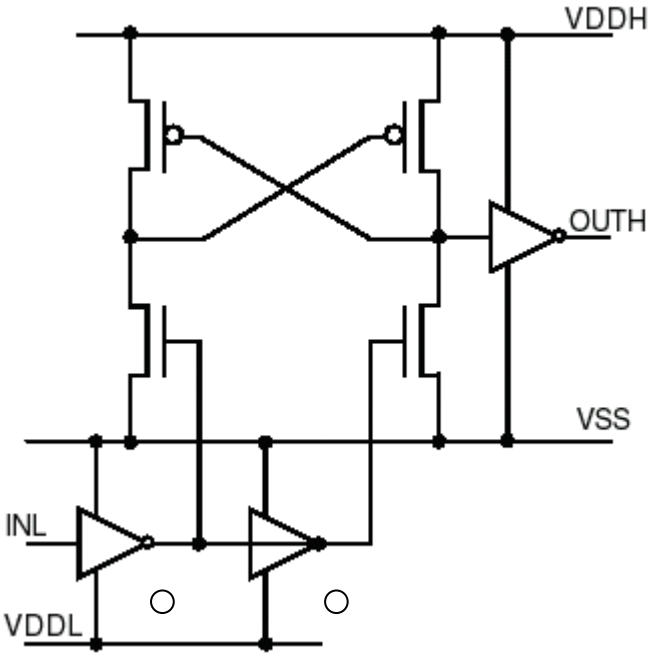




Level Shifters – Low-to-High

- There are a number of design techniques – but a simple straight-forward design is shown below.
- This design takes a buffered and an inverted form of the lower voltage signal and uses this to drive a cross-coupled transistor structure running at the higher voltage.





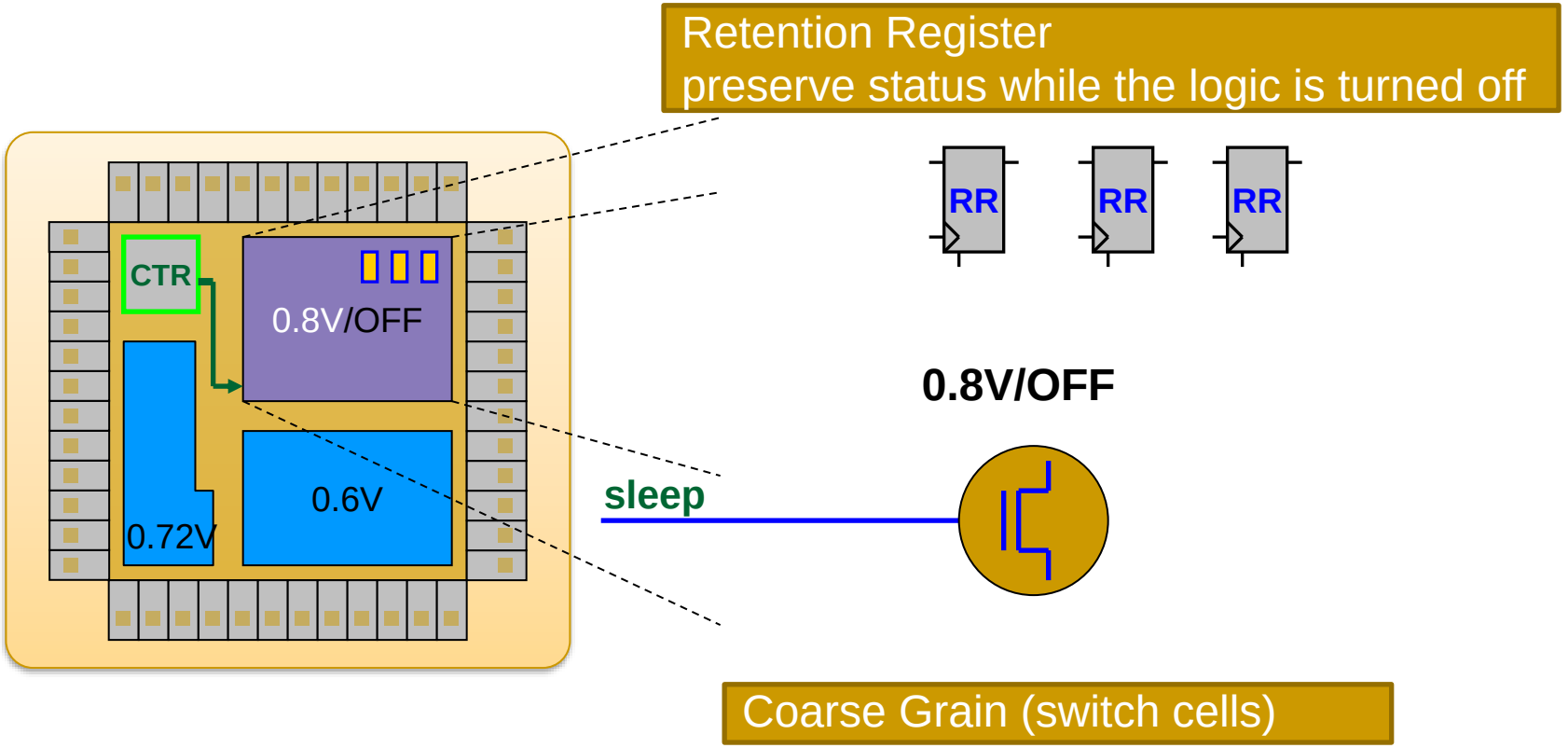
SYNOPSYS®

Level Shifter Recommendations and Pitfalls

Recommendations:

- Place the level shifters in the receiving domain – in the lower domain for High-to-Low shifters, in the higher domain for Low-to-High shifters.
- Low-to-High level shifters have significant delays that need to be understood and thoughtfully factored into RTL design partitioning for timing critical blocks.
- Ensure there is a defined relationship between different voltage domains such that the operating conditions make it clear whether an up- or down-shifter is required.

Special Cell Requirements: Power Gating

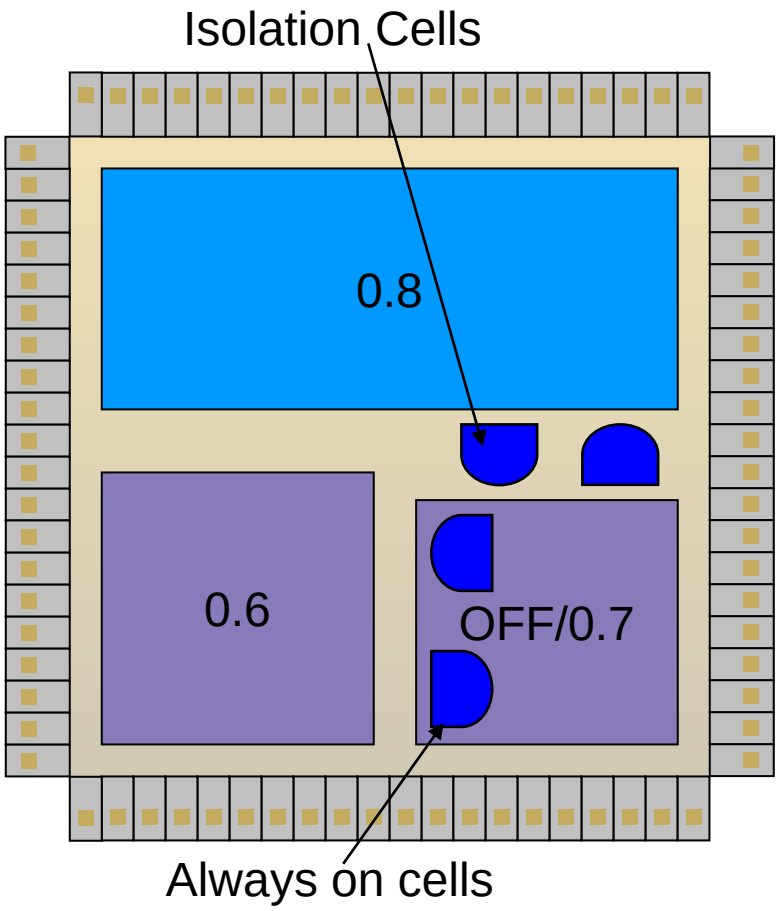


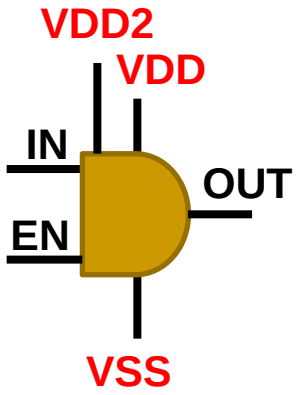
SYNOPSYS

Retaining the State

- For some power-gated blocks, it is highly desirable to retain the internal state of the block during power down, and to restore this state during power up. Such a retention strategy can save significant amounts of time and power during power up.
- One way of implementing such a retention strategy is to use retention registers in place of ordinary flip-flops.
- Retention registers typically have an auxiliary or shadow register that is slower than the main register but which has much less leakage current. The shadow register is always powered up, and stores the contents of the main register during power gating.
- These retention registers need to be told when to store the current contents of the main register into the shadow register and when to restore the value back to the main register.
- This control is provided by the power gating controller.

Special Cell Requirements: Power Gating



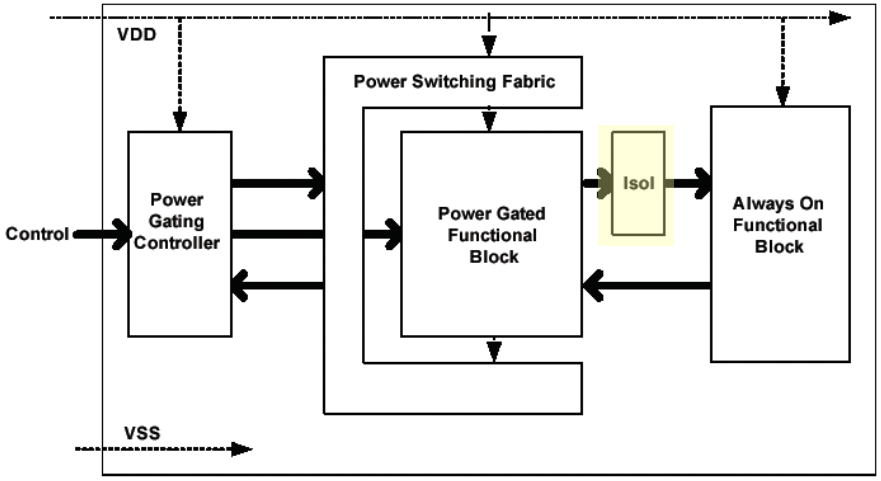


simplified logic model

SYNOPSYS®

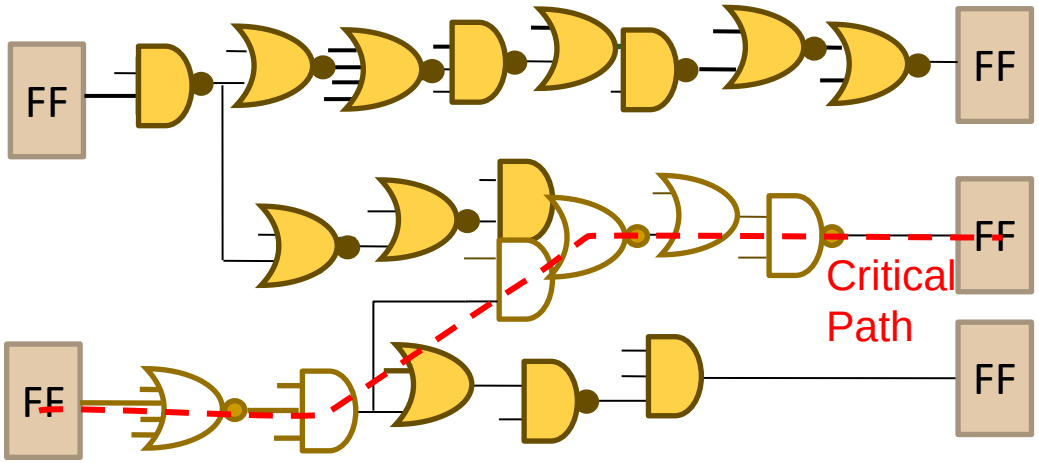
Isolation Cells

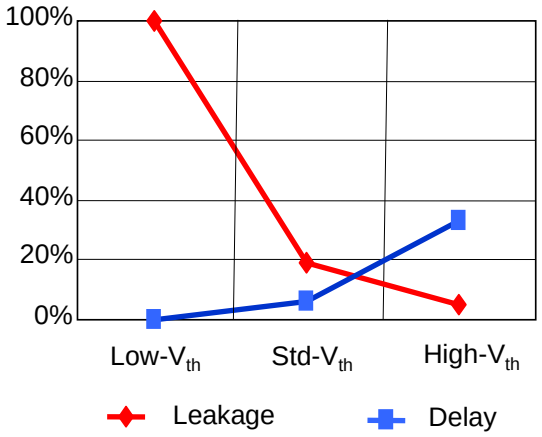
- One challenge for power gating designs is that the outputs of the power gated block may ramp off very slowly. The result could be that these outputs spend a significant amount of time at threshold voltage, causing large crowbar currents in the always powered on block. To prevent these crowbar currents, isolation cells are placed between the outputs of the power gated block and the inputs of the always on block. These isolation cells are designed so that they do not experience crowbar current when one of the inputs is at threshold, as long as the control input is off. The power gating controller provides this isolation control signal.

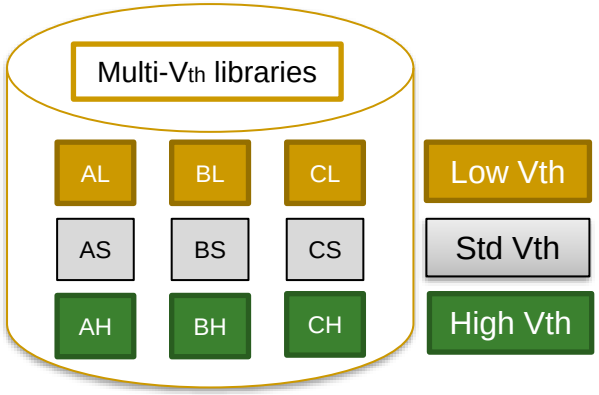


SYNOPSYS

Multi-Threshold Libraries







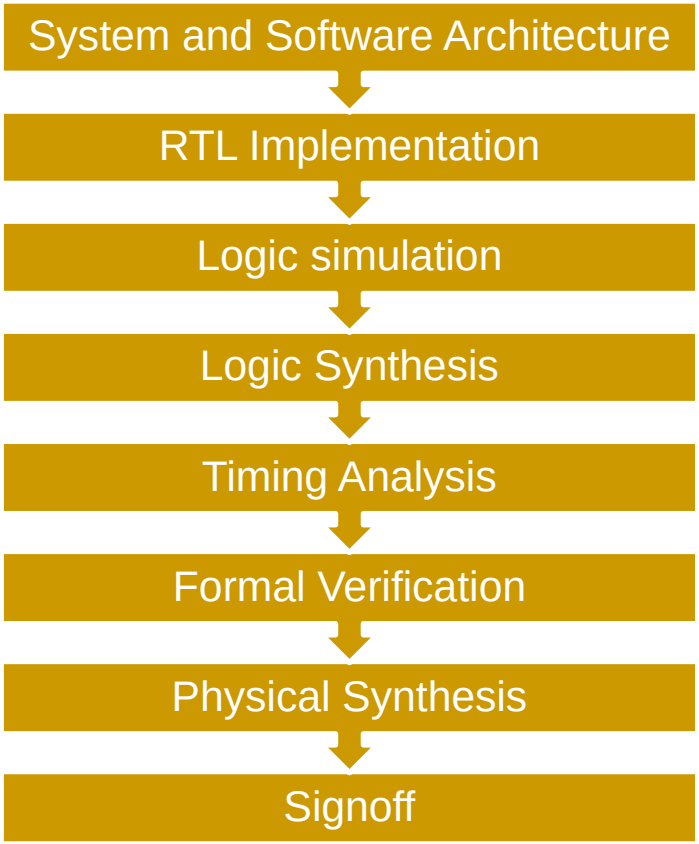
SYNOPSYS®

DSCL: Characterization Corners

#	Corner Name/ Library Name Suffix	Process (NMOS proc. – PMOS proc.)	Power Supply (V)	Temperature (^{0}C)
	Typical Characterization Corners			
1	tt0p8v25c	Typical - Typical	0.8	25
2	tt0p8v125c	Typical - Typical	0.8	125
3	tt0p8vm40c	Typical - Typical	0.8	-40
4	ss0p72v25c	Slow - Slow	0.72	25
5	ss0p72v125c	Slow - Slow	0.72	125
6	ss0p72vm40c	Slow - Slow	0.72	-40
7	ff0p88v25c	Fast - Fast	0.88	25
8	ff0p88v125c	Fast - Fast	0.88	125
9	ff0p88vm40c	Fast - Fast	0.88	-40
	Low Voltage Operating Conditions			
10	tt0p6v25c	Typical - Typical	0.65	25
11	tt0p6v125c	Typical - Typical	0.65	125
12	tt0p6vm40c	Typical - Typical	0.65	-40
13	ss0p6v25c	Slow - Slow	0.6	25
14	ss0p6v125c	Slow - Slow	0.6	125
15	ss0p6vm40c	Slow - Slow	0.6	-40
16	ff0p7v25c	Fast - Fast	0.7	25
17	ff0p7v125c	Fast - Fast	0.7	125
18	ff0p7vm40c	Fast - Fast	0.7	-40

Impact of LPD on Digital Design Flow

Conventional Design Flow

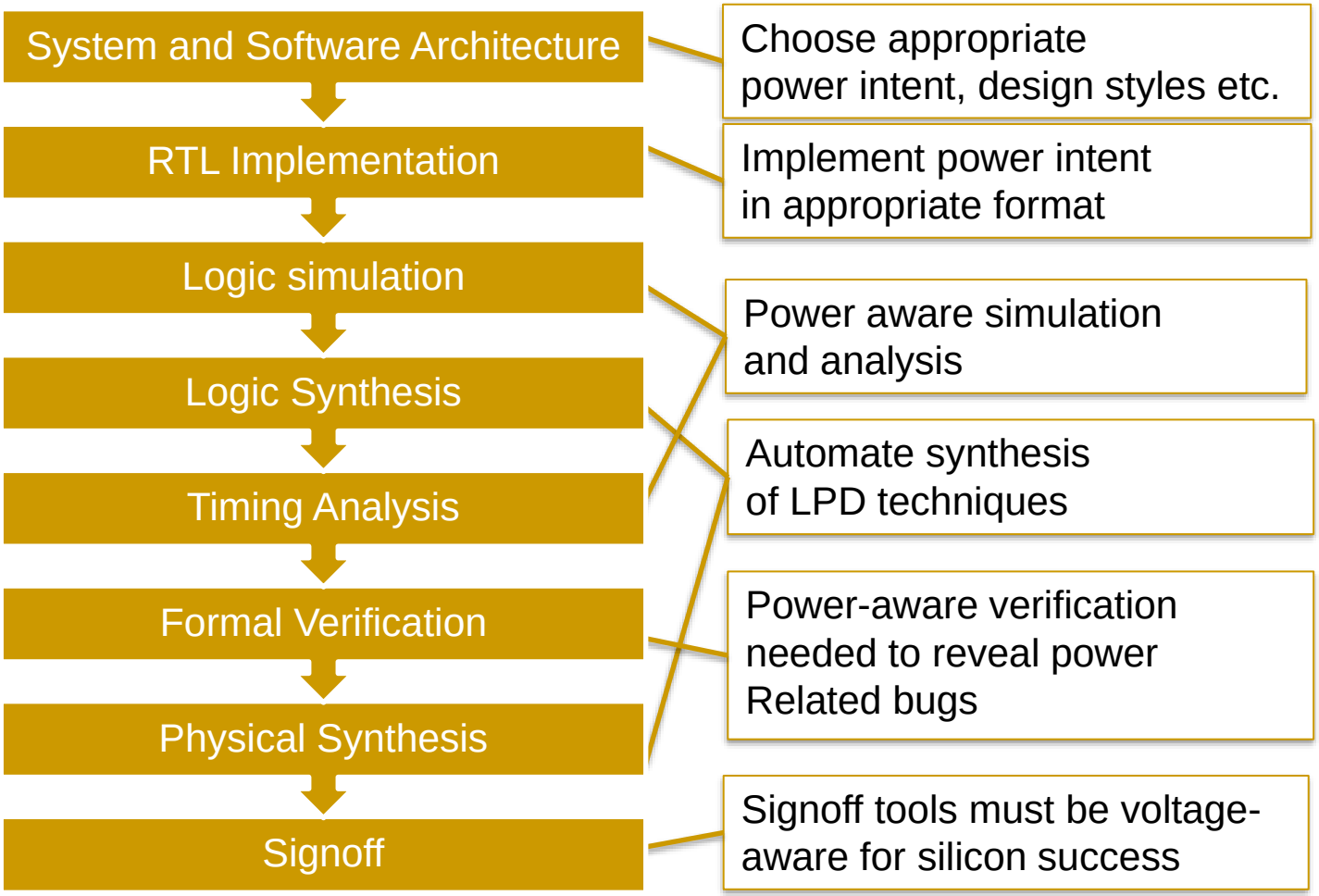


Power Management should be taken into account at the earliest design stages

Almost every step of design flow need to be modified for LPD

SYNOPSYS®

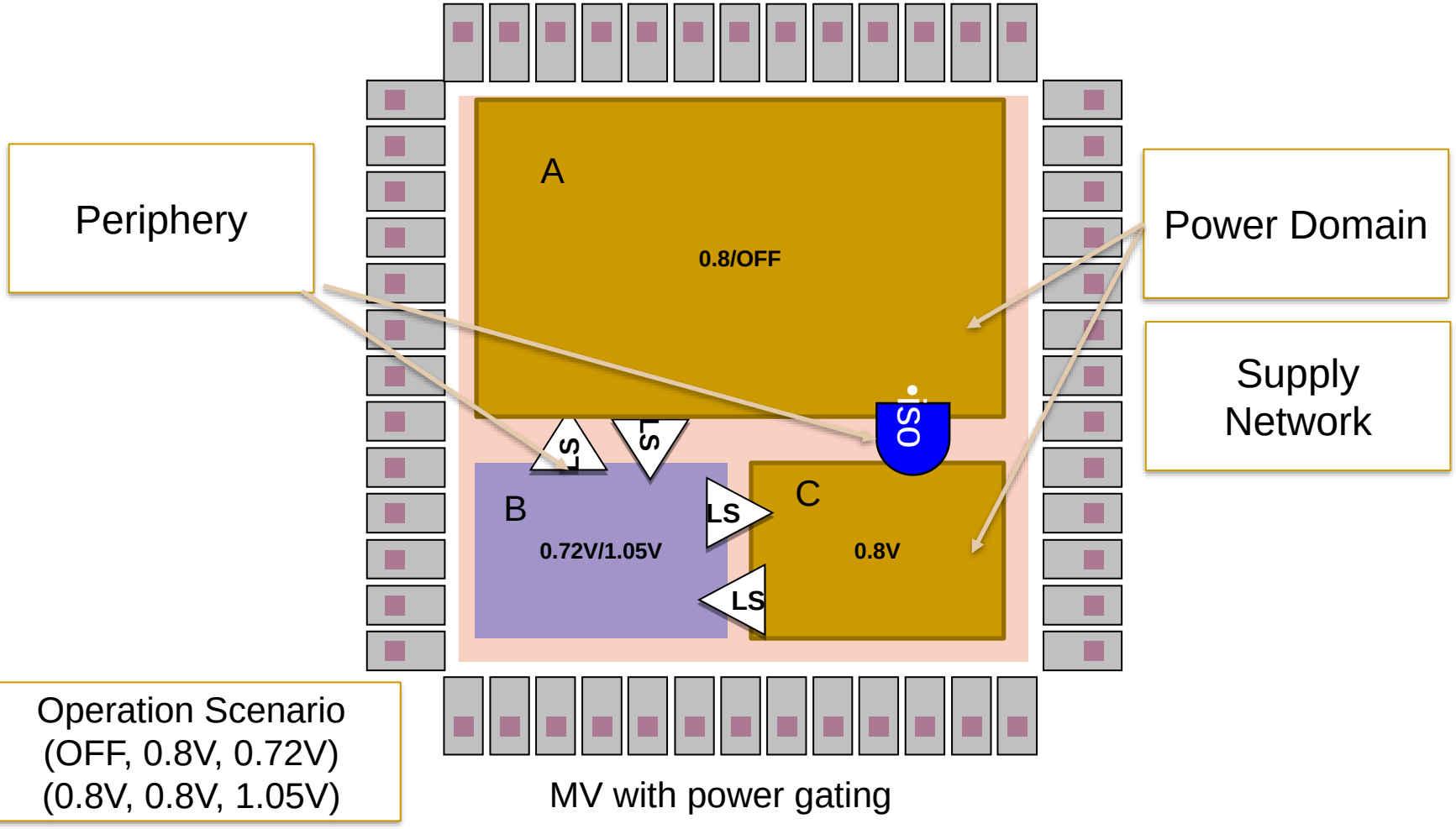
Power-Aware Design Flow



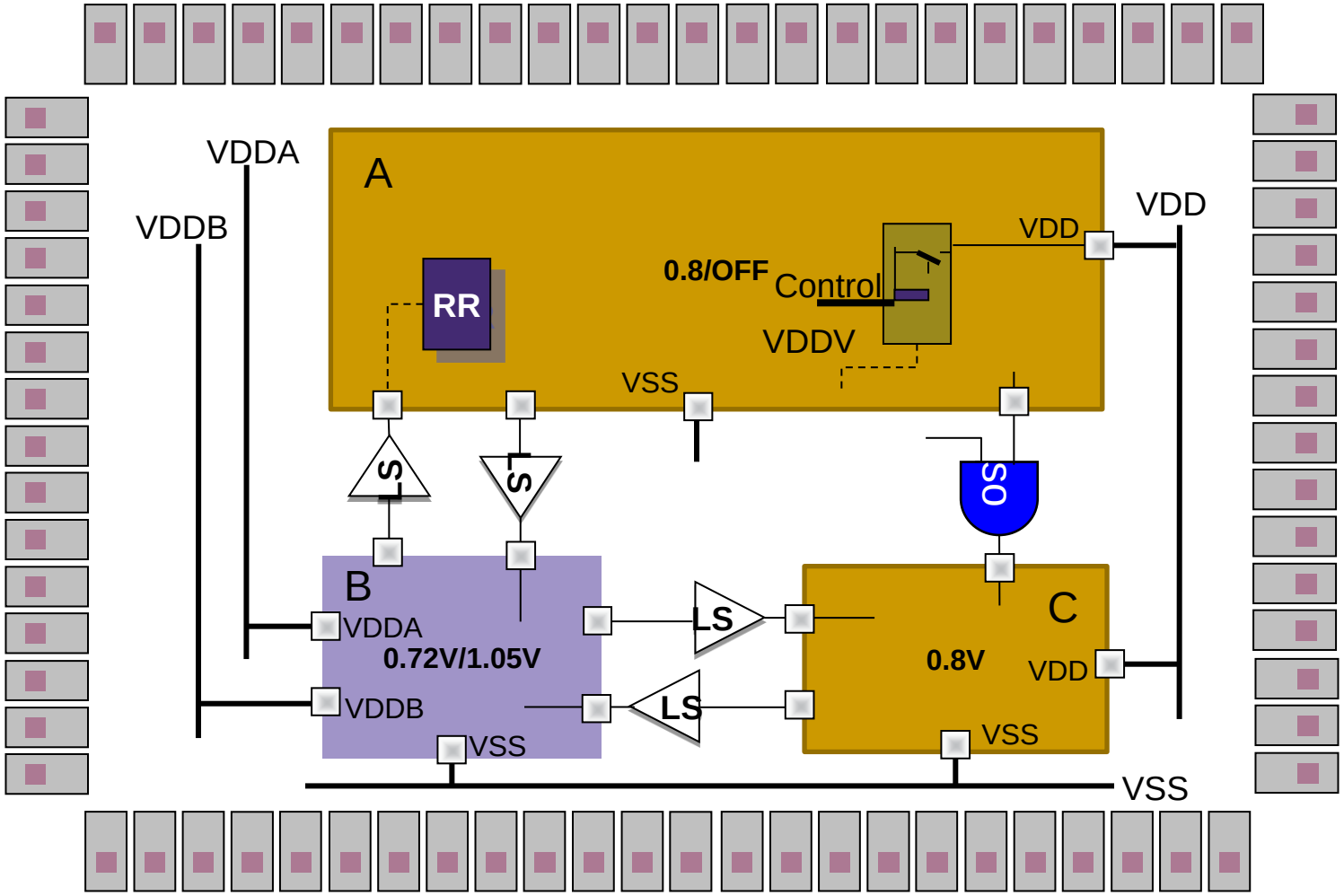
Unified Power Format (UPF): Necessity

Language	Specification of power intent	Interoperable among EDA tools	Can be freely used (open standard)
Hardware Description Languages (Verilog, VHDL, etc.)	-	+	+
Vendor –Specific Formats	+	-	-
UPF	+	+	+

Specifying Power Intent

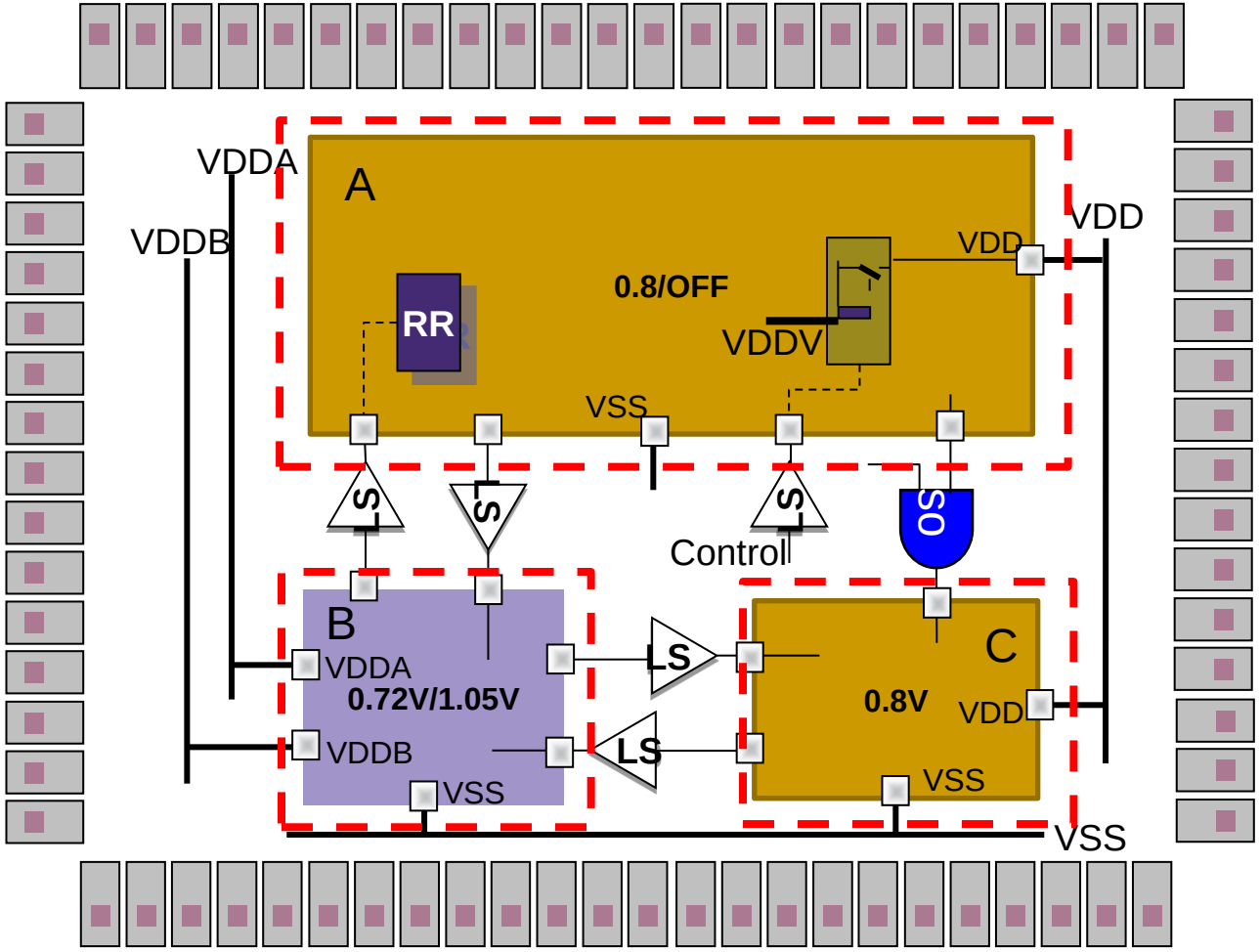


MV with Power Gating Example



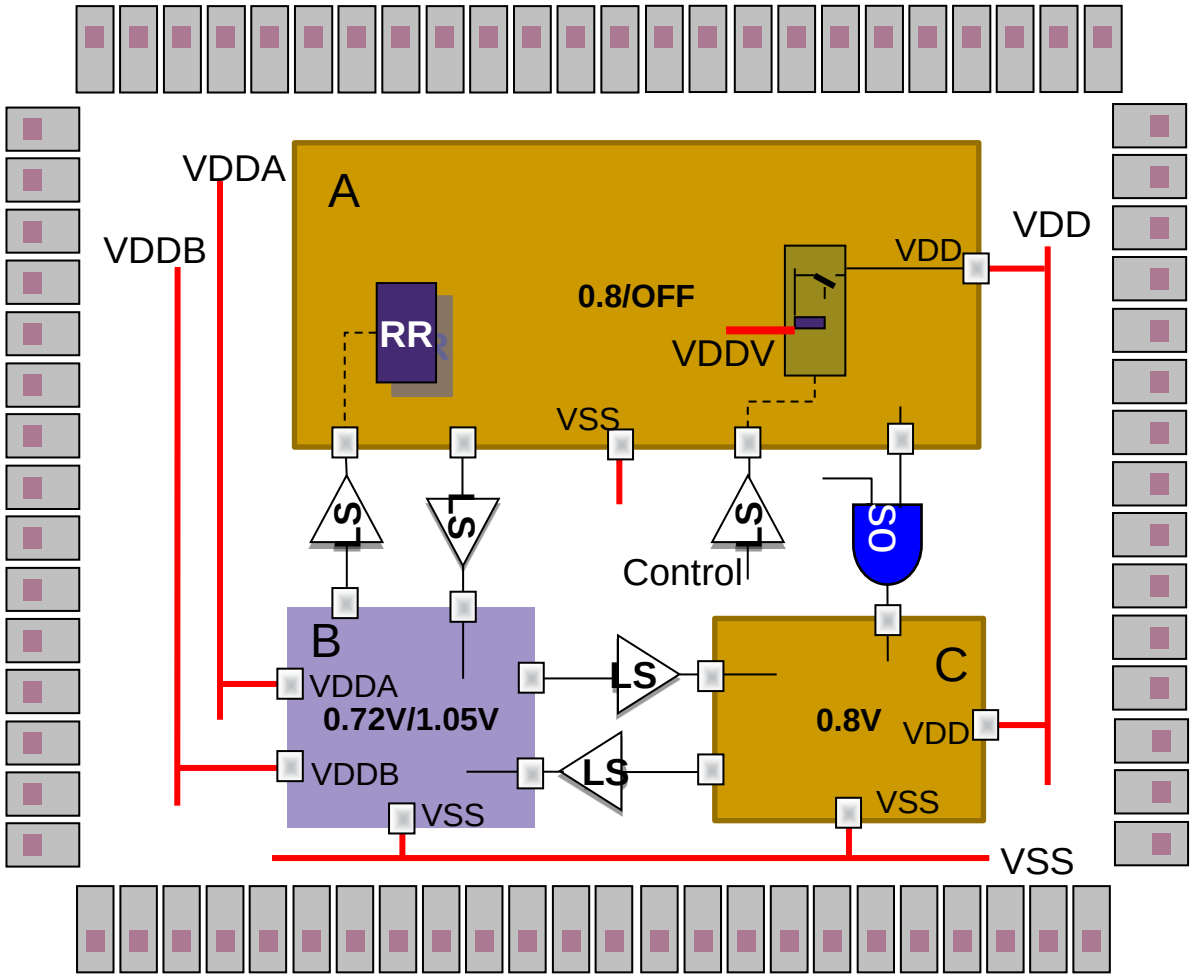
SYNOPSYS

UPF: Power Domains



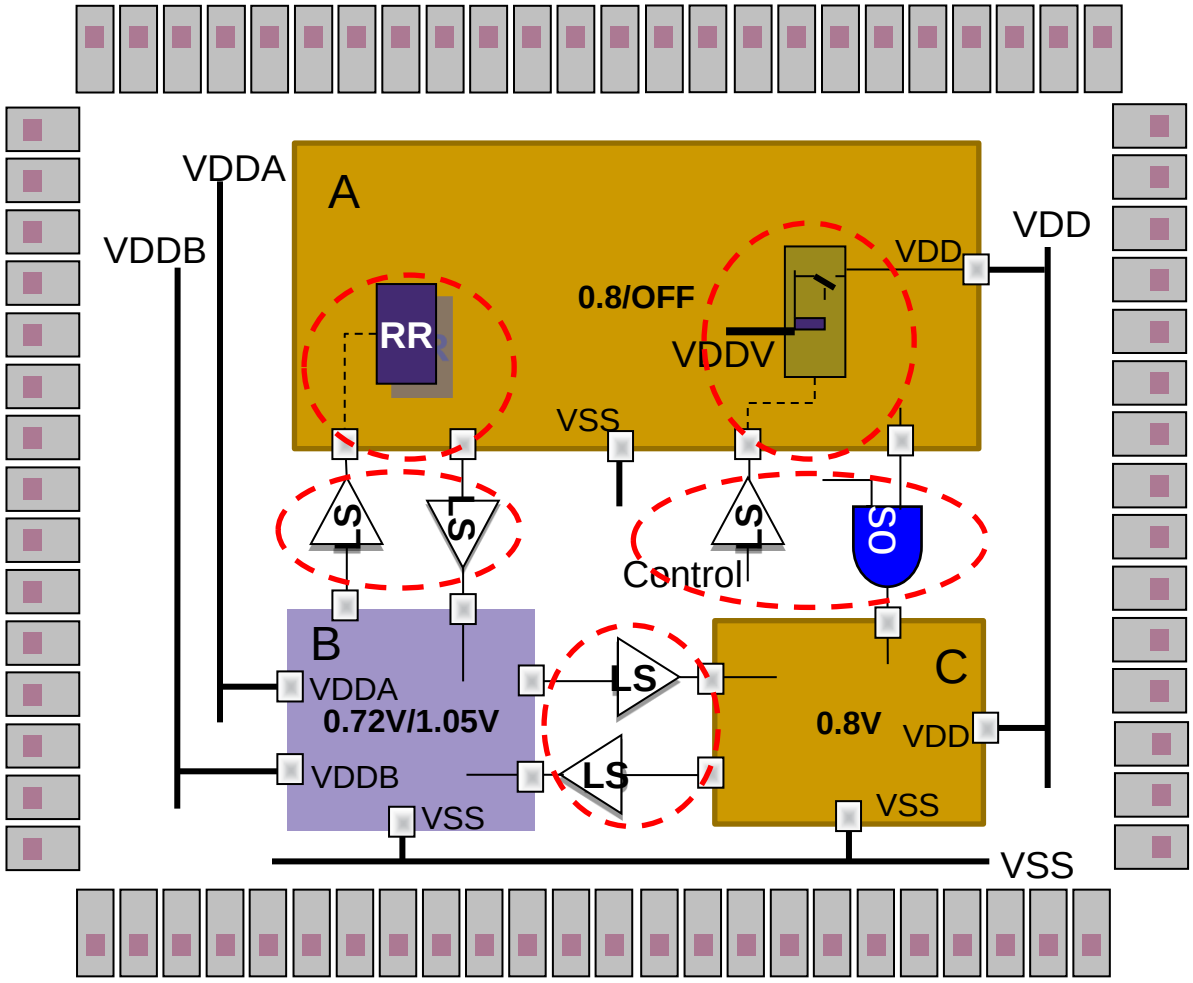
UPF	
Power Domain	**Power State**
A	0.8/OFF
B	0.72/1.05
C	0.8

UPF: Supply Network



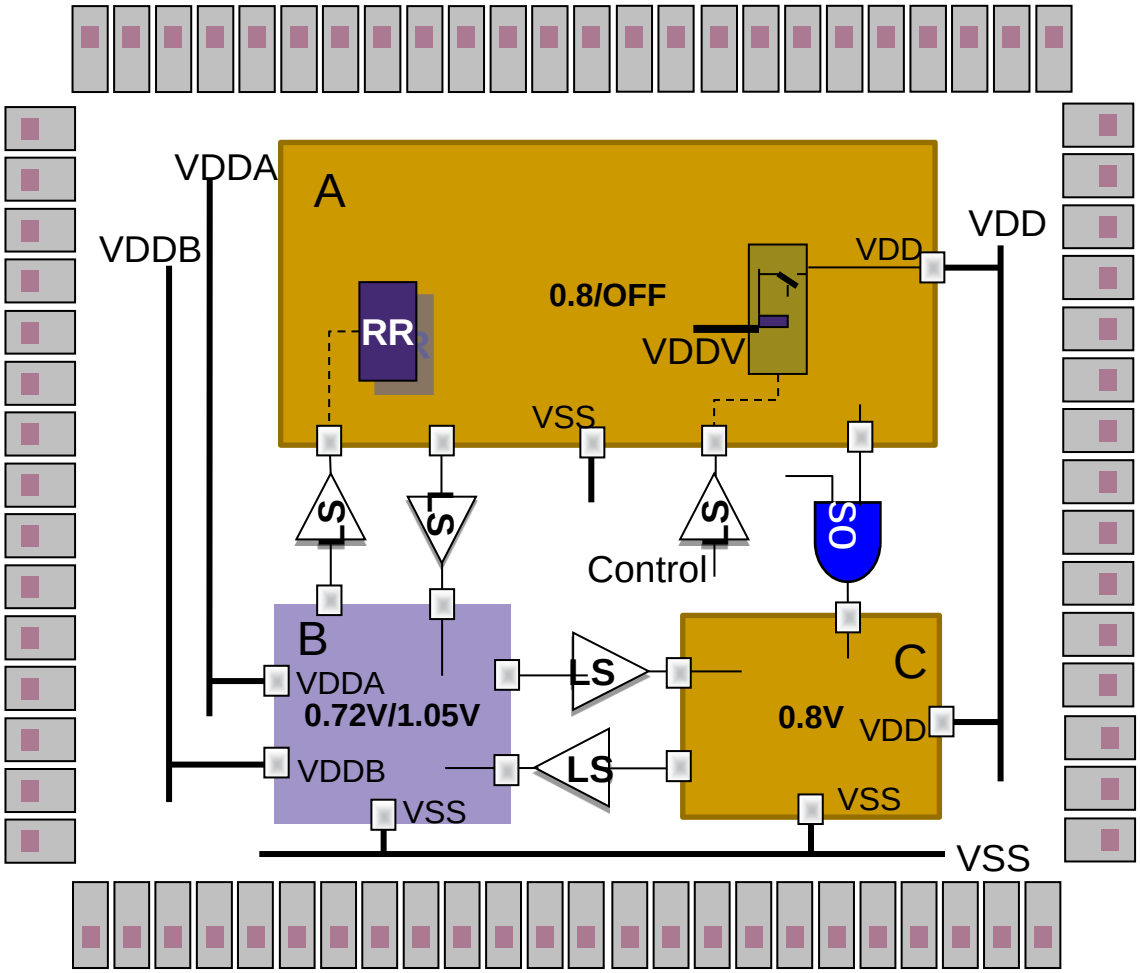
UPF		
Supply Net	**Voltage Level (V)**	**Power Domain**
VDD	0.8	C
VDDA	0.72	B
VDDB	1.05	B
VDDV	Virtual 0.8	A
VSS	Common Ground	A/B/C

UPF: Periphery



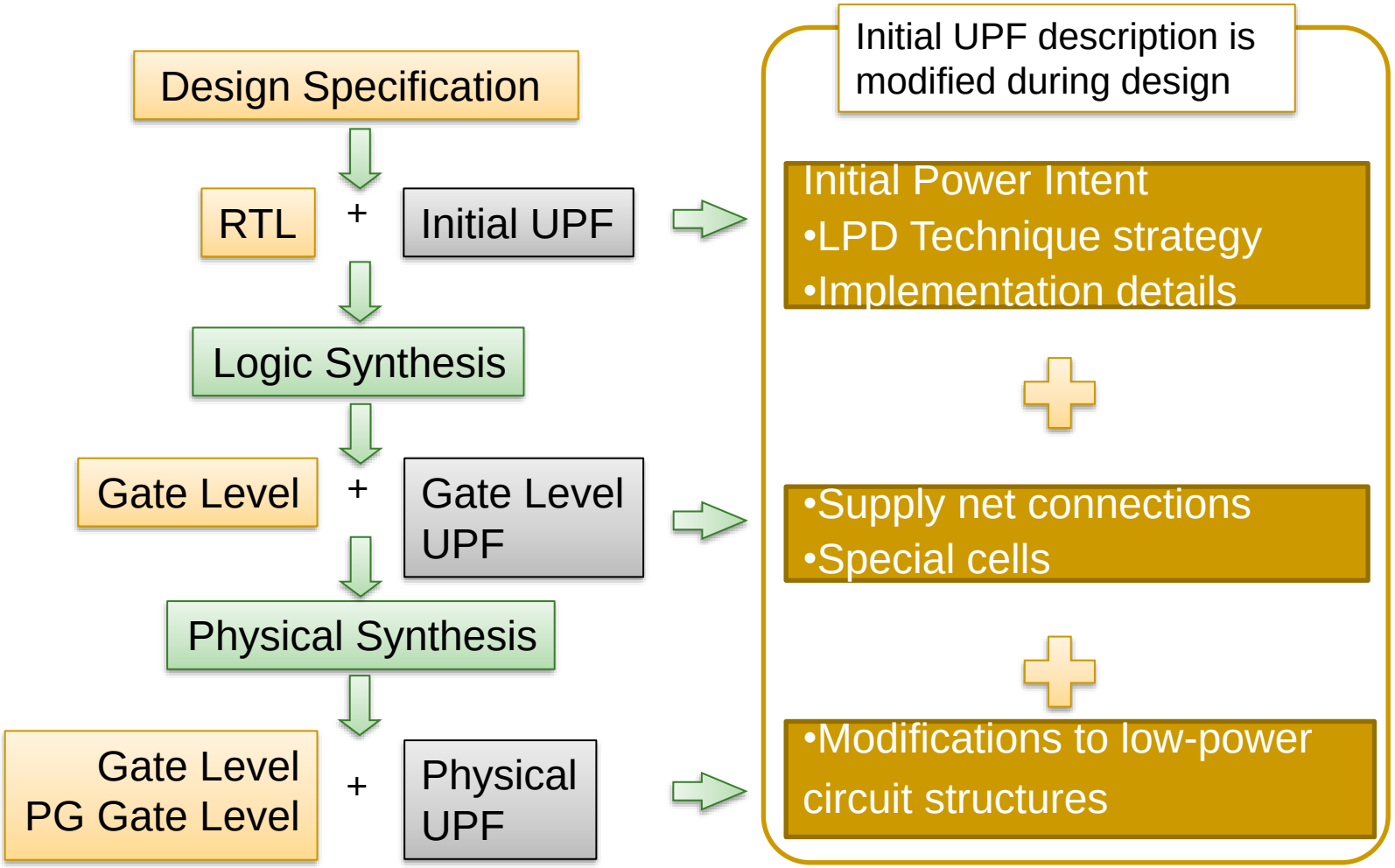
UPF
Required Periphery
Level Shifters between A and B
Level Shifters between B and C
Isolation between C and A
Retention Cell inside A
Control Block inside A

UPF: State Scenario



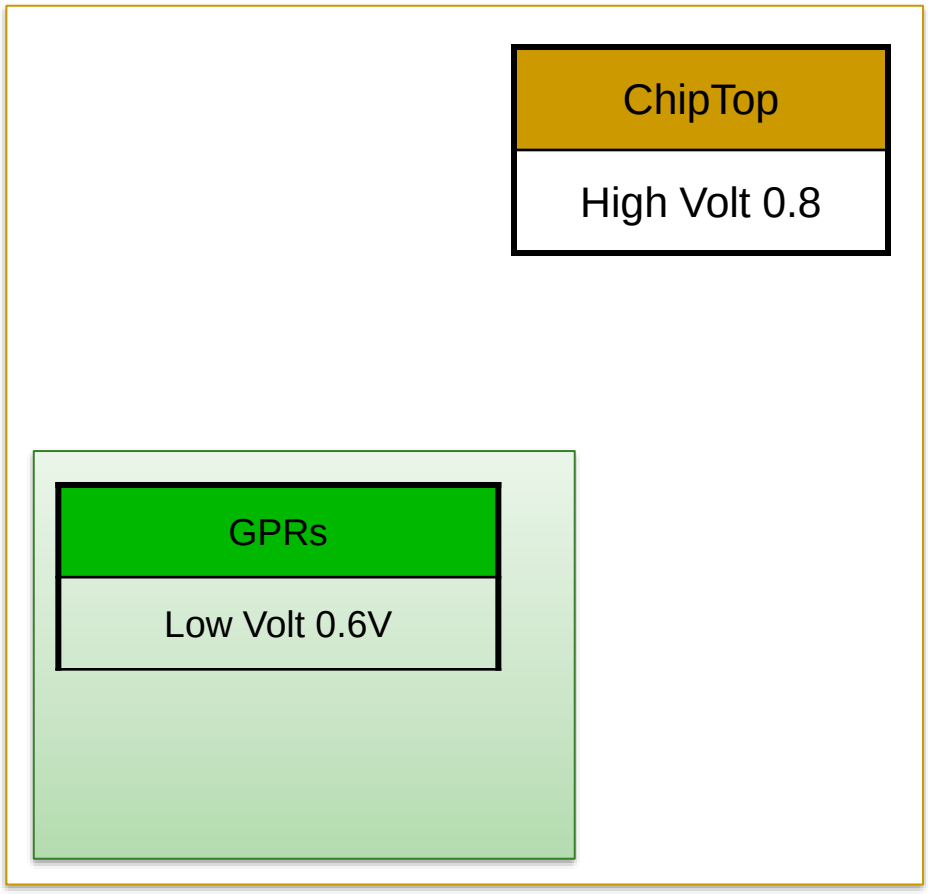
UPF			
A	B	C	Scenario
0.8	0.72	0.8	Allowed
0.8	1.05	0.8	Allowed
OFF	0.72	0.8	Allowed
OFF	1.05	0.8	Not Allowed

Design Flow Modification with UPF



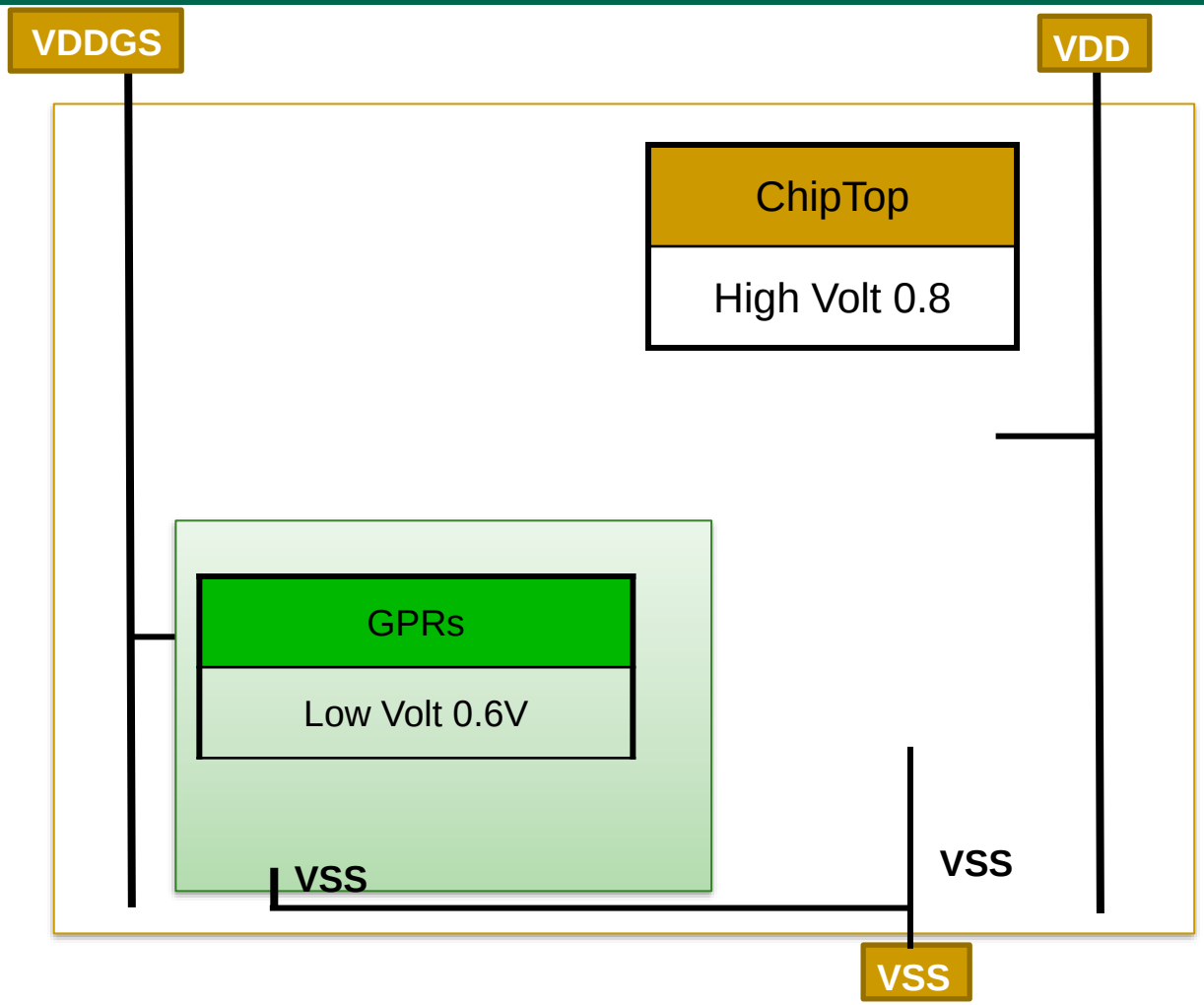
SYNOPSYS®

ChipTop : Operating Voltages



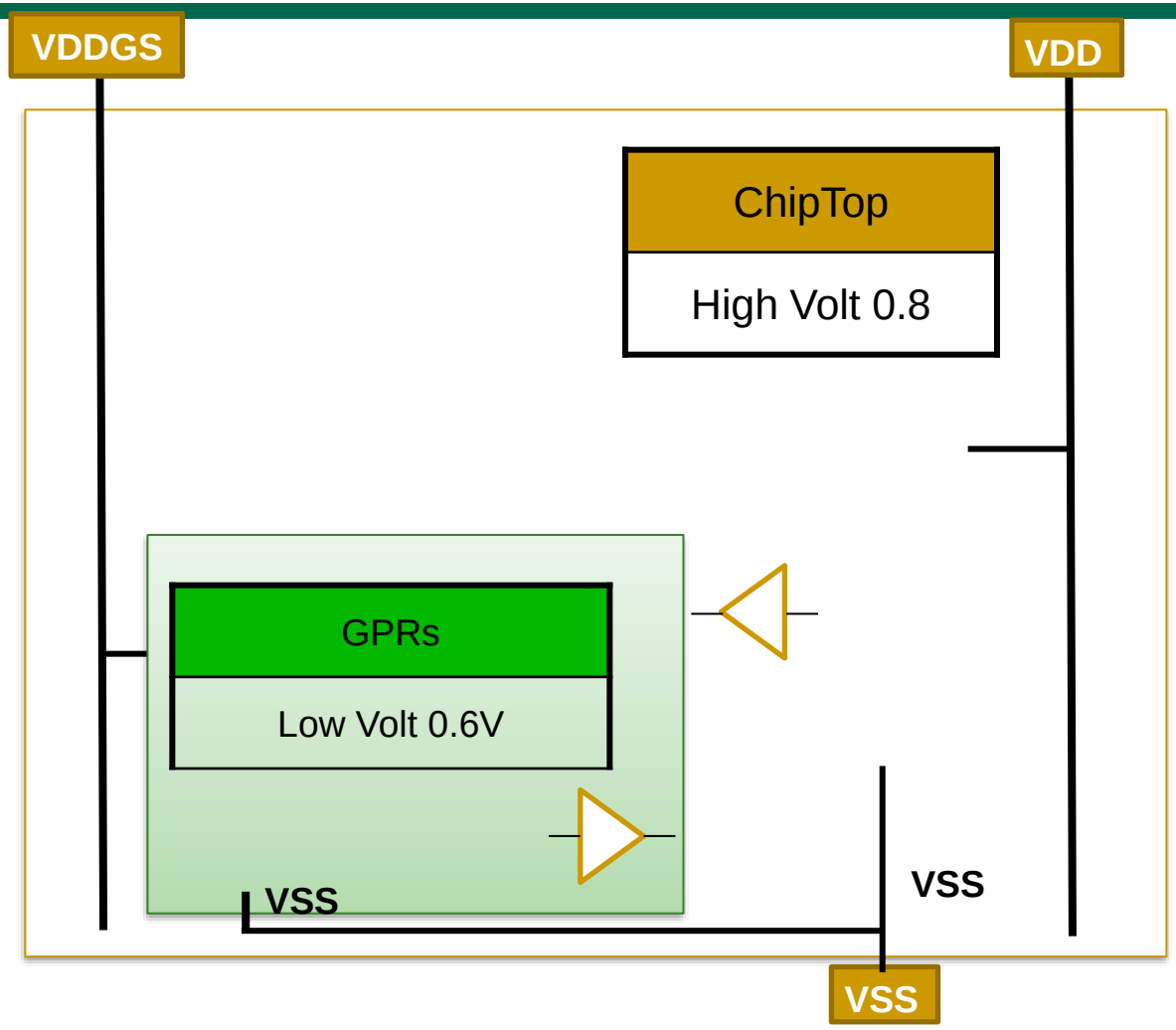
SYNOPSYS®

ChipTop : Supply Network



SYNOPSYS®

ChipTop : Level Shifters



SYNOPSYS®

UPF

```
create_power_domain TOP
create_power_domain GPRS  -elements GPRs
# VDD
create_supply_port VDD
create_supply_net  VDD
create_supply_net  VDD    -domain GPRS -reuse
connect_supply_net VDD    -ports VDD
# VSS
create_supply_port VSS
create_supply_net  VSS
create_supply_net  VSS    -domain GPRS -reuse
connect_supply_net VSS    -ports VSS
# VDDG
create_supply_port VDDGS
create_supply_net  VDDGS
create_supply_net  VDDGS    -domain GPRS -reuse
connect_supply_net VDDGS    -ports VDDGS
```

Creating power domains

Creating supply nets

UPF (2)

```
set_domain_supply_net TOP  \
       -primary_power_net VDD   -primary_ground_net VSS
set_domain_supply_net GPRS \
       -primary_power_net VDDGS -primary_ground_net VSS

add_port_state VDD     -state {HV  0.95}
add_port_state VDDGS   -state {LV  0.7}
add_port_state VSS     -state {GND  0}
```

Setting primary power/ground nets

Creating Port State Table

Characterization Corners

Corner Name	Process (NMOS proc. – PMOS proc.)	Temperature (C)	Power Supply 1 (V)	Power Supply 2 (V)	Notes
...	...	...	...	...	...
WORST	Slow - Slow	125	0.8	-	
...	...	...	...	...	...
WORST	Slow - Slow	125	0.72	-	
...	...	...	...	...	...
WORST	Slow - Slow	125	0.8	0.72	
...	...	...	...	...	...