

Low Voltage Design and Production in 28HPM of Bitcoin Mining ASICs

Authors:

Assaf Gilboa (Spondoolies)

Amnon Parnass (Verisense)

Michael Chen (GUC)

Igor Elkanovich (Verisense) - presenter

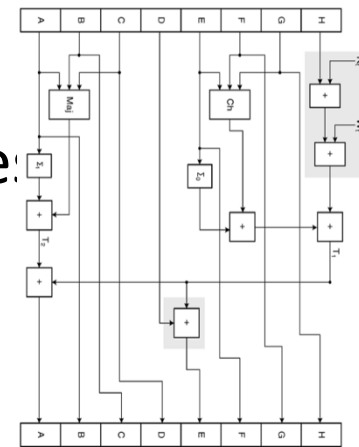
Spondoolies, Verisense & GUC

- Spondoolies:
 - Develops power-efficient Bitcoin mining equipment, Kiryat-Gat
- Verisense:
 - ASIC and FPGA development services, Jerusalem
- GUC:
 - IC implementation and manufacturing services, rich IP portfolio, Hsinchu, Taiwan
- Bitcoin ASICs responsibility:
 - Spondoolies: system, SW, ASIC design
 - Verisense: ASIC design, verification, synthesis, STA, netlist handoff
 - GUC: IPs, libraries, backend, package, test, production

Bitcoin Mining Application

- Architecture:
 - Bitcoin calculation is based on double SHA256
 - Many 128-stage pipelined engines, each generate result every clock
 - Random data: high toggle rate
- Optimization: system cost/performance
 - Chip **cost/performance**: mostly silicon area
 - **Power/performance**: power affects system cost
 - Dynamic power is dominant
 - Performance: GigaHash/sec
- Short lifetime: a new generation every 6 months

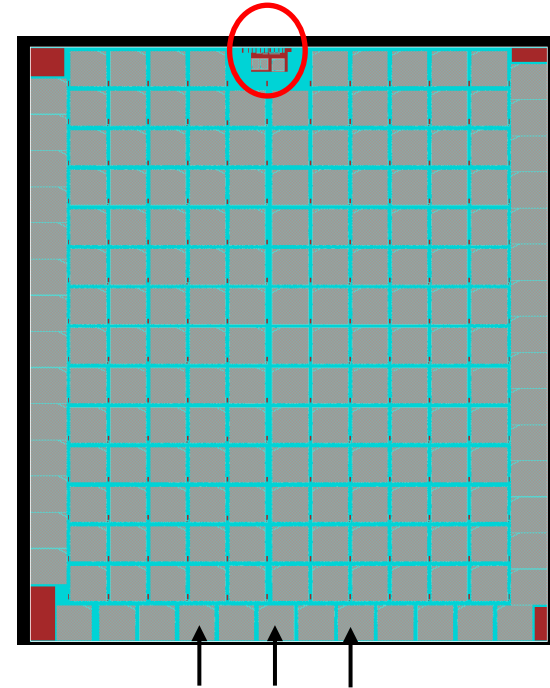
One Pipeline stage
768 bits width



2nd gen Mining Chip: RockerBox

- Process: TSMC 28HPM
- 246 M gates, no SRAMs
- Power: 80W (typical, 0.63V)
- Voltage range: 0.55V...0.8V
- Die: 116 mm²
- Package: FCBGA 19mmx19mm
- High volume production since July/2014

I/Os, PLL, Temperature Sensor
management logic



193 Double SHA-256 Engines

No I/Os on sides

ESDs are spread through the die

Key Development

- Optimization of a whole system of 30 chips
- Cost efficiency:
 - Logic redundancy for high yield
 - Proprietary logic BIST instead of Scan
 - Process shift for higher performance
- Power efficiency:
 - Operating voltage 30% below 28HPM nominal
 - Triple-loop Dynamic Voltage Frequency Scaling (DVFS)
 - Accurate dynamic power analysis and toggle rate spreading

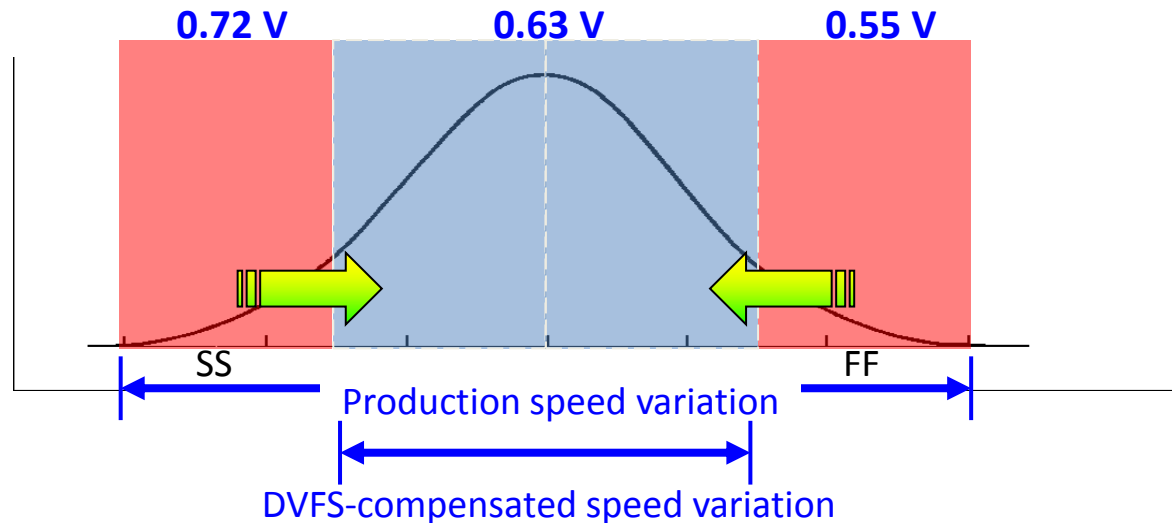


Logic Redundancy for High Yield

- System is tolerant to faulty SHA-256 engines
- Proprietary logic BISTs to identify faulty engines
 - The BIST uses SHA-256 pipeline itself
 - LSFR-based BIST for a strict test
 - Vector-based BIST for statistical system test
- Scan wasn't inserted to reduce area/power overhead
 - Fault coverage tool was developed by Ilia Greenblat to check BIST coverage
- Final product yield: 99%
 - Natural yield is about 90% (Die: 116 mm²)

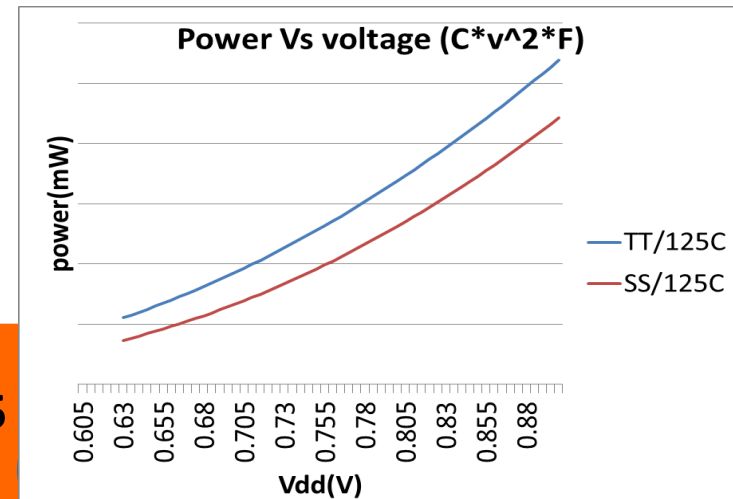
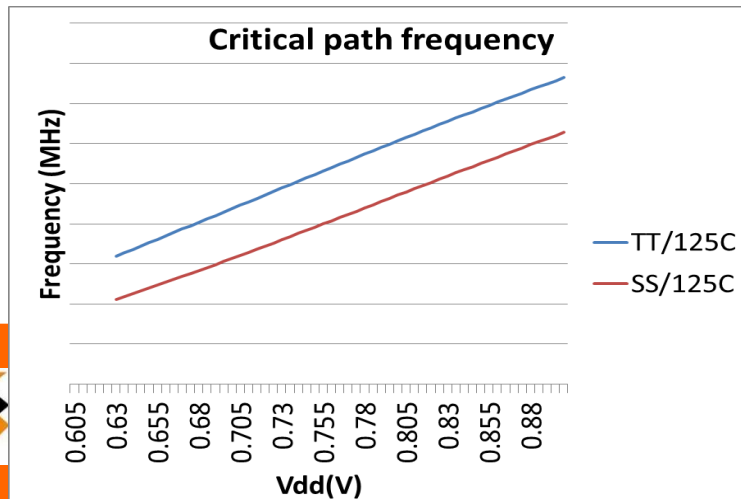
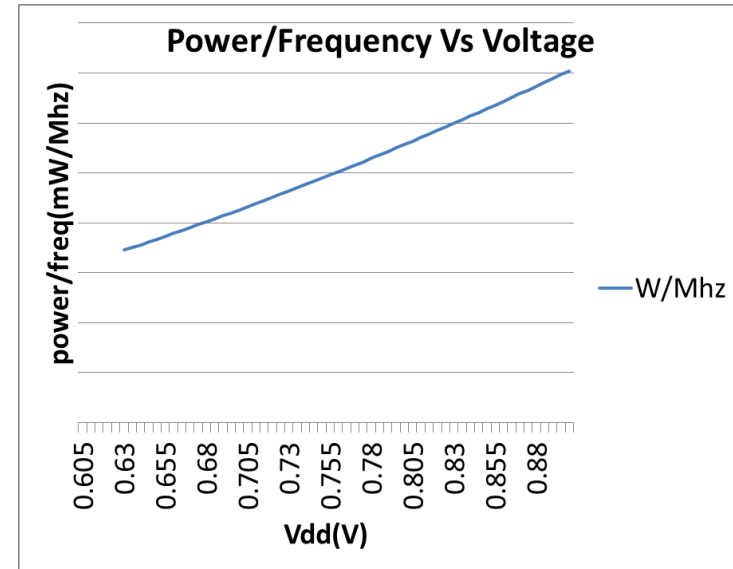
Dynamic Voltage Frequency Scaling (DVFS)

- Voltage regulator (DC2DC) per ASIC
- Slow and Fast corners are compensated by voltage adjustment



DVFS Voltage Target Definition

- Trends:
 - Frequency vs. Voltage → linear
 - Power vs. Voltage → V^2
 - Power/frequency vs. Voltage → linear
 - Conclusion: use lowest possible voltage
- Linearity range low limit:
 - At around $V_{tl N} + V_{tl P}$
- Selected DVFS target at TT/125C: 0.63V



Triple Loop DVFS

- DVFS loops
 - Frequency loop per chip: searching for max frequency
 - Temperature loop per chip: at 125°C voltage is reduced
 - Total system power loop: increase/decrease chips voltages to meet total system power budget
- DVFS performance
 - Speed sensor correlation vs. critical path is a key
 - Full correlation is achieved by using logic BIST (pipeline itself)
 - Voltage granularity: 1 mV, frequency granularity: 10 MHz
 - Hysteresis at every action point

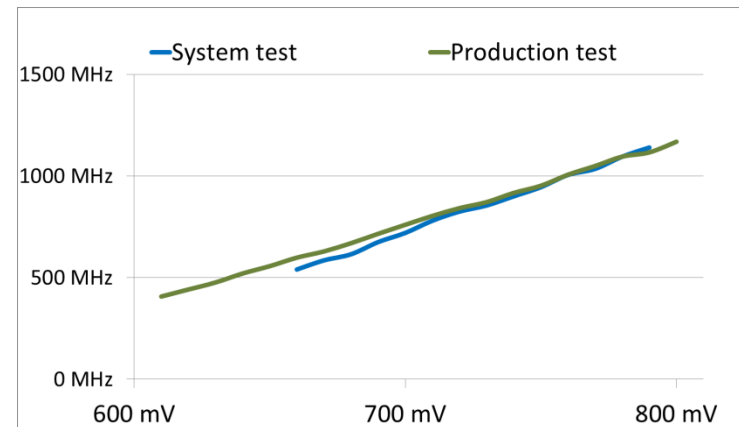
DVFS Operation in System

Achieved robust and stable DVFS system operation

Every chip and its DC2DC report:
voltage, frequency, power, temperature

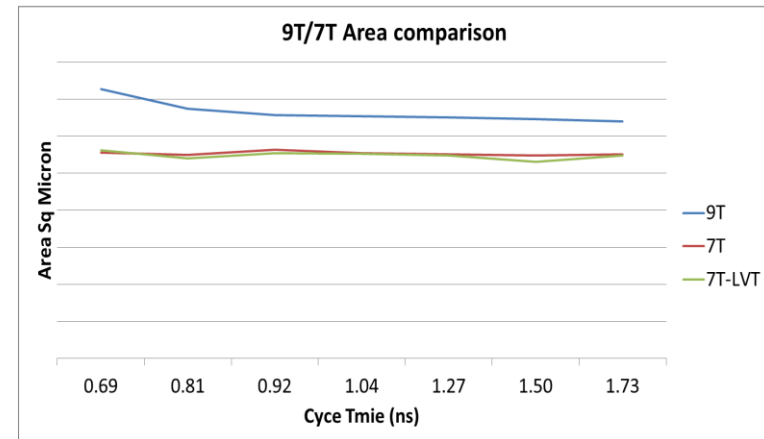
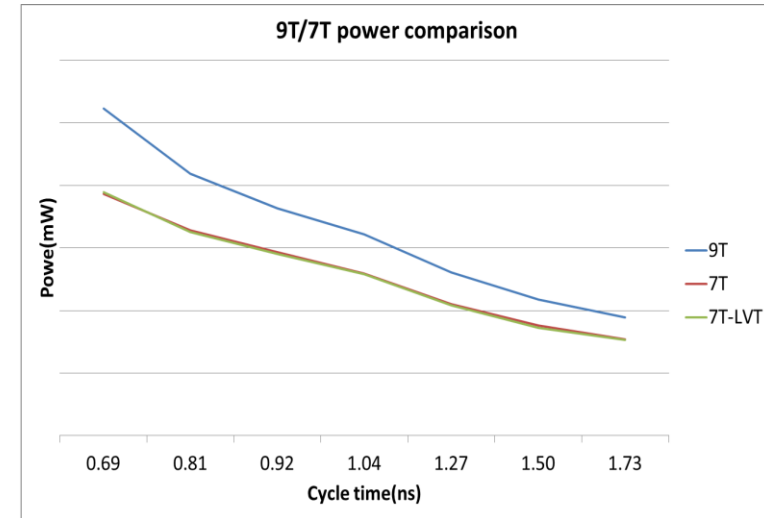
```
15: DC2DC/-1/: [vlt1:693 vlt2:698 (DC1:794 T1:794 U1:727) 21W 31A 33c] ASIC: [ 80c 705hz (
16: DC2DC/-1/: [vlt1:693 vlt2:698 (DC1:794 T1:794 U1:727) 25W 36A 37c] ASIC: [ 80c 810hz (
17: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 21W 31A 36c] ASIC: [ 80c 705hz (
LOOP[6] ON :)
18: DC2DC/-1/: [vlt1:693 vlt2:698 (DC1:794 T1:794 U1:727) 22W 31A 32c] ASIC: [ 80c 705hz (
19: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 23W 33A 35c] ASIC: [ 80c 795hz (
20: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 25W 36A 37c] ASIC: [ 80c 810hz (
LOOP[7] ON :)
21: DC2DC/-1/: [vlt1:697 vlt2:698 (DC1:794 T1:794 U1:727) 23W 33A 32c] ASIC: [ 80c 720hz (
22: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 18W 27A 34c] ASIC: [ 80c 690hz (
23: DC2DC/-1/: [vlt1:691 vlt2:698 (DC1:794 T1:794 U1:727) 24W 34A 34c] ASIC: [ 80c 735hz (
LOOP[8] ON :)
24: DC2DC/-1/: [vlt1:697 vlt2:698 (DC1:794 T1:794 U1:727) 22W 31A 31c] ASIC: [ 80c 705hz (
25: DC2DC/-1/: [vlt1:693 vlt2:698 (DC1:794 T1:794 U1:727) 22W 32A 29c] ASIC: [ 80c 750hz (
26: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 23W 34A 32c] ASIC: [ 80c 735hz (
LOOP[9] ON :)
27: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 19W 28A 27c] ASIC: [ 80c 660hz (
28: DC2DC/-1/: [vlt1:697 vlt2:698 (DC1:794 T1:794 U1:727) 26W 37A 26c] ASIC: [ 80c 795hz (
29: DC2DC/-1/: [vlt1:695 vlt2:698 (DC1:794 T1:794 U1:727) 17W 25A 30c] ASIC: [ 80c 675hz (
```

Correlation: production test vs.
system test



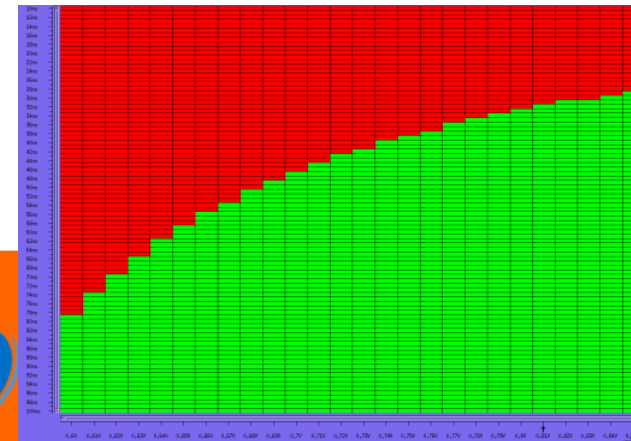
Library Selection for Low Voltage

- 7T Libraries were selected
 - 20% area/power reduction
 - Negligible performance impact
- Dynamic vs. leakage vs. performance trade-off:
 - SVT, 35 nm: 85% (Synthesis)
 - LVT, 40 nm: 14% (Synthesis)
 - LVT, 35 nm: 1% (Timing closure)
- Only 18% pre-layout to post-layout area growth
 - 18%: Clock tree, hold, set up, transitions fix



Timing Closure

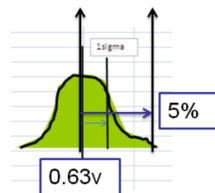
- P&R optimization corner: TT, 0.63V, 125C, Cmax
- Set up corners: SS, **0.72V**, 0C/125C, Cmax/RCmax
 - 5 corners
- Hold time corners: full matrix 0.63V-0.88V
 - 13 corners
- OCV and uncertainty: defined for every corner by Monte-Carlo spice simulations
- All used libraries were re-characterized for all defined corners
- Production tests were defined according to timing closure corners



Low Voltage Methodology

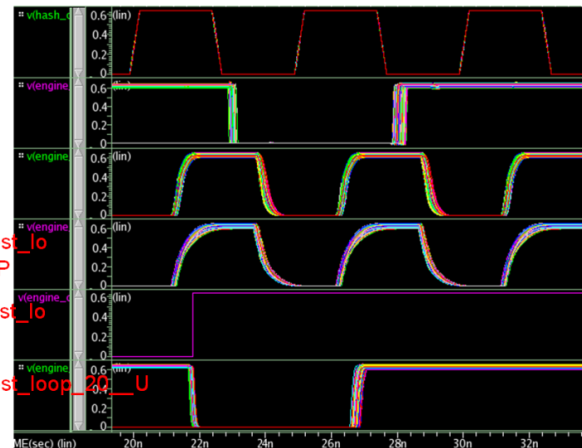
- 4-3 transistors in series cells were excluded from libraries
- Max Xtalk glitch and max transition parameters were tightened
- Extracted LO spice simulations:
 - All clock trees to check transitions
 - Critical path to check correlation vs. STA
 - Libraries' FFs were simulated to check metastability convergence
- Separate 0.9V power domain for PLL, TS and I/Os

MC Clocks simulation



sha256_transform_hash1_digest_lo
op_20__U0_tx_w_reg_468_/CP
sha256_transform_hash1_digest_lo
op_20__U0_tx_w_reg_468_/D
sha256_transform_hash1_digest_lo
op_20__U0_tx_w_reg_468_/Q

clk
D
CP



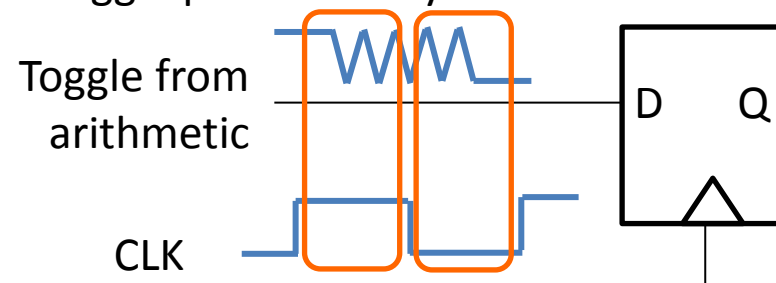
Dynamic Power Analysis

- Accurate dynamic power estimation flow was developed
 - 10% accuracy vs. post-silicon measurements
- For power analysis accuracy:
 - Representative activity from simulation
 - GL simulation resolution = gate delay (20-30 ps)
 - State dependent SAIF
- Allows accurate comparison of arithmetic architectures

Stages:
Netlist
Extracted RC
SDF generation
GL simulation
SAIF generation
Power reporting

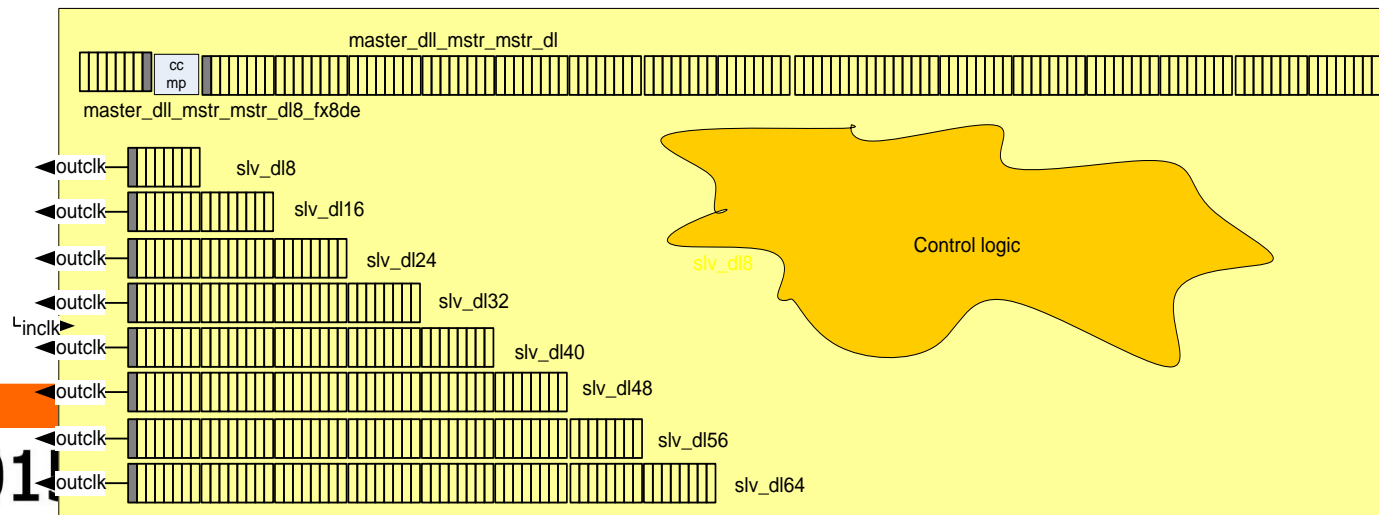
State dependent SAIF example:

D toggle power is very different at CLK high and low



Current Peak Challenge

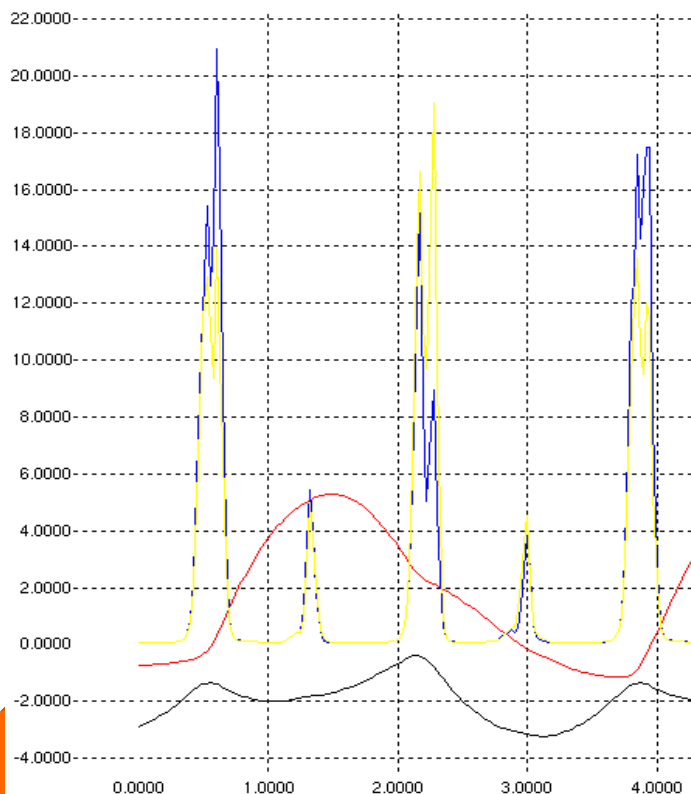
- Original toggle rate: FFs 50%, arithmetics 200%-300%
 - Random data flows through arithmetic pipeline
- New architecture:
 - Reduced FFs toggle to 34% (Spondoolies patent)
 - Divided toggle rate to 4 clock phases
- Master-slave DLL was developed to spread clock edges



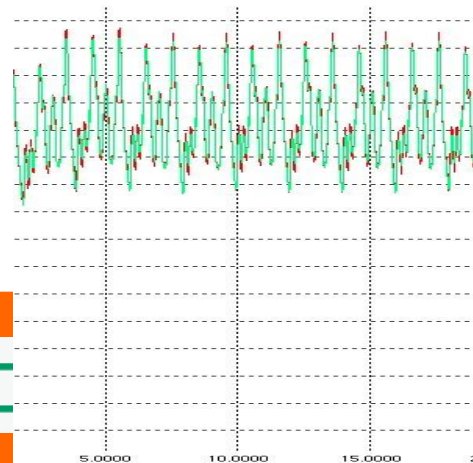
Spreading Current Peaks

Current peaks were reduced to acceptable level

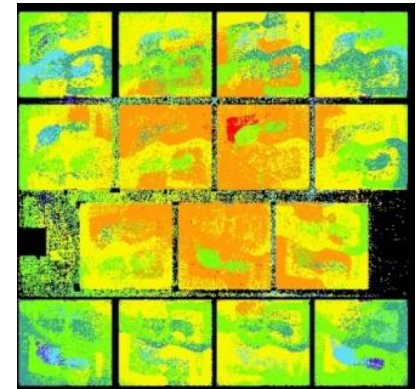
Original current peaks



Final current peaks

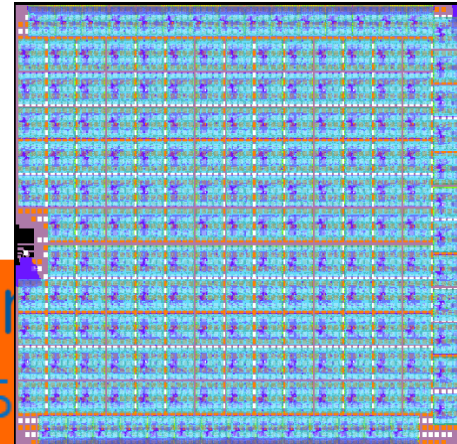


Dynamic IRdrop simulation



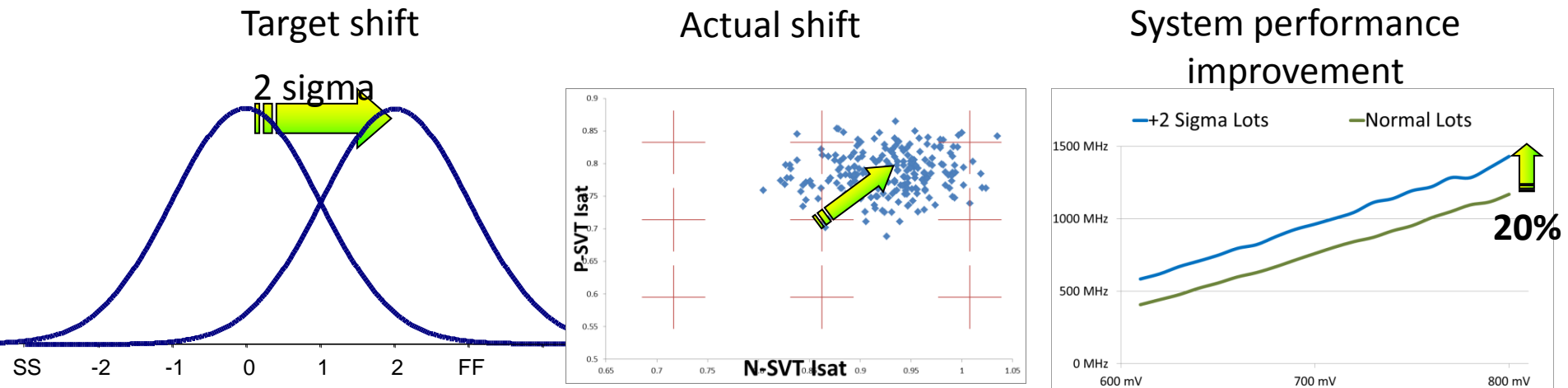
Metal Stack for Low IRdrop

- At low voltage and high supply current (130A) low IRdrop is critical
- Traditional power grid metal stack:
 - X direction: Z layer (8.5 KÅ copper)
 - Y direction: AP layer (14 KÅ aluminum)
- We added U layer (4x lower resistance than Z layer):
 - X direction: U layer (35 KÅ copper)
 - Y direction: Z layer (8.5 KÅ copper) + UT-AP layer (28 KÅ Al)
 - TSMC provided tech files for 5x1z1u1UT-AP stack
- Disadvantage: U layer metal density is limited to 50%
 - Z and AP layer densities are up to 70%
- Achieved static IRdrop 2%, dynamic IRdrop 5%



Process Shift

- 28HPM was shifted by 2 sigma to fast corner
 - 20% performance increase
- 98% yield due to redundancy and hold time margins
- More than 300 Ku were produced in 6 months



Summary

- Optimization for entire multi-chip system
- For cost and power efficiency:
 - Redundancy, logic BIST, triple-loop DVFS, process shift
- 28HPM process was used at low voltage and wide DVFS range
 - Methodology was proven in high volume production