Linzhi Semiconductors

Linzhi Working Papers
No 18

History of Bitcoin Mining Hardware.

by Chen Min

November 2019

Keywords: LWP18, History, Bitcoin, SHA-256, ASIC

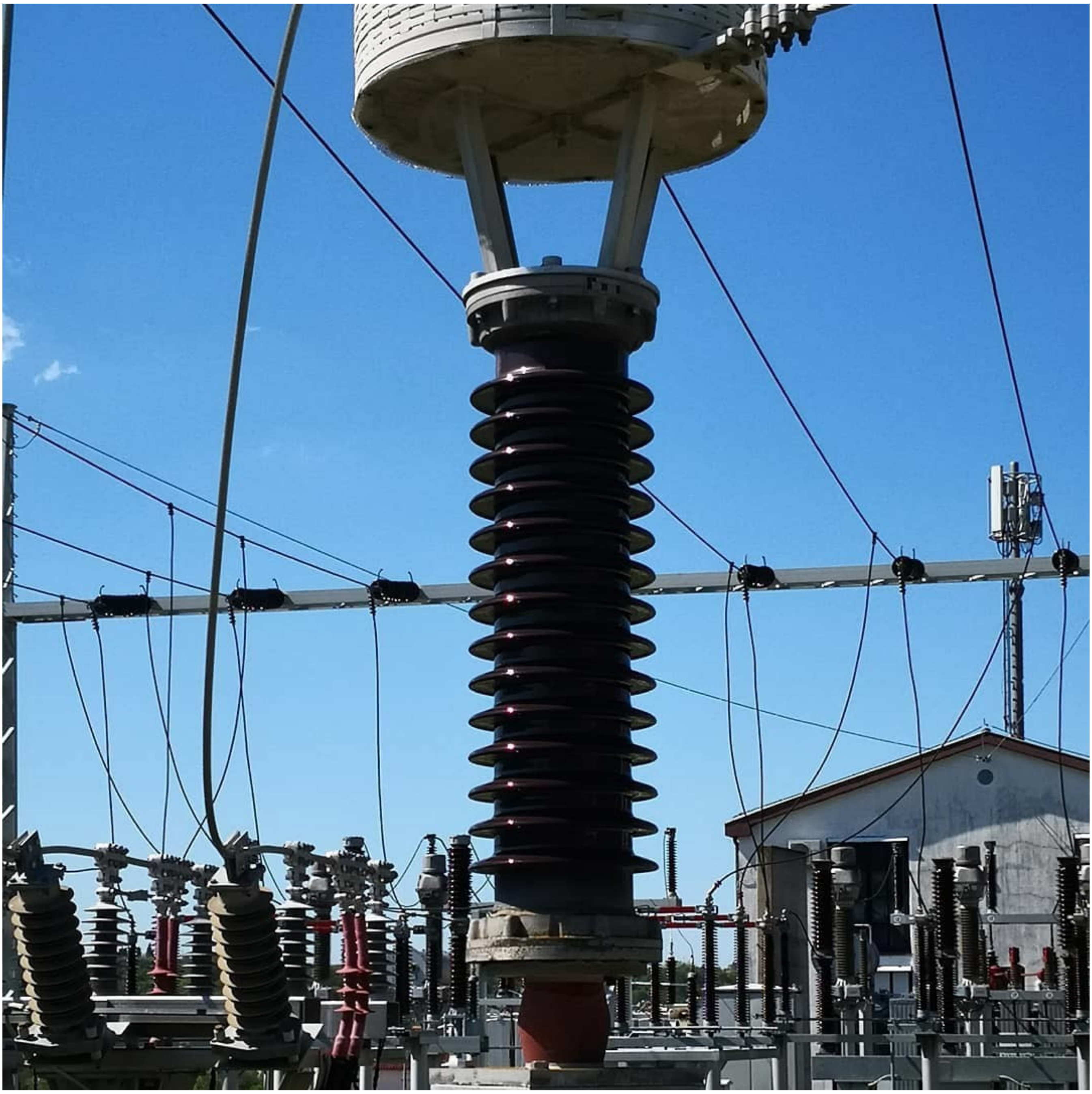
This publication is available at https://linzhi.io
Telegram discussion https://t.me/LinzhiCorp

All original rights for text and media are released into the public domain, attribution welcome. Rights of quoted or translated sources remain with their respective owners.

History of Bitcoin Mining Hardware

Linzhi ASICs
Nov 14 · 7 min read

Optimizing SHA-256



· · ·

A hardware maker optimizes one or more of: cost, performance, efficiency, time to market or reliability.

In this article we walk the interested reader through this process for SHA-256, with emphasis on ASICs, and finally leading to the supply chain. In parentheses you find the rough improvement. C = cost, E = efficiency, P = performance. Note that C+5% means cost was improved (=reduced) by 5%. The article is rather dry without drawings. With sufficient interest we might add more detail or drawings. From FPGA on it's a firsthand account.

For now it's good as it is. Enjoy.

. . .

Phase I — CPU era

Bitcoin mining started in 2009 on CPUs (~5000 J/GH).

. . .

Phase II — GPU era

In 2010 and 2011, GPUs increased performance 10x-100x (~500 J/GH)

. . .

Phase III — FPGA era

The first FPGAs didn't provide better performance per chip, but better efficiency and slightly better cost (~50 J/GH). The biggest optimization during the FPGA era was a **pipeline architecture**.

The performance bottleneck of SHA-256 was its deep algebra chain. Adding one more pipeline stage in the middle of the algebra chain doubled performance (P+100%, from 100 MHz to 200 MHz), for example on the Icarus FPGA miner. Further distribution of the algebra to different pipeline stages increased performance to 300 MHz (P+50%), for example on the Lancelot FPGA miner or Bitfury's FPGA miner.

Other optimizations during the FPGA era were:

1. **Pre-calculation**: Remove as much as logic as possible from the hashing loop (C+5%, E+5%).
2. **Toggle rate reduction**: Reduce average gate toggle rate (E+1%).
3. **Dynamic frequency control**: Let every hash core operate on it's best frequency (P+20%).
4. **Hand place and route**: Makes space for more hash cores in the FPGA (P+15%). FPGA toolchain limitations made this a very hard effort.
5. **Over Voltage**: The power distribution system of FPGAs is not designed for high power operation, so it drops a lot of voltage. Even when you power 1.2V at the PCB, on silicon you will have less than 1.0V due to package drop, which is a big performance hit. We couldn't change the chip design (yet), so the only way to counter this was to supply the chip with higher voltage (P+30%).
6. **Miner management bus and protocol:** Design for managing large amounts of miners, using minimum wire count. That saved cost for large scale operation (C+5%), and continued into the ASIC era.

· · ·

Phase IV — Function ASICs

In 2012 many teams started designing mining ASICs, such as Butterfly Labs, Avalon Miner, Bitfury, Friedcat, etc. At that time the design quality was not so important, time to market was paramount. Dozens of teams rushed into the market, hashrate increase reached 40% a week.

By **removing programmability**, even an old process ASIC cost much less than an FPGA (C+1000%).

The first ASICs fixed FPGA weaknesses in power distribution and thermal resistance. With only 1 clock pin, 1 reset pin and 4 signal pins, **all other pins were used for power and ground**. **QFN packages** provided good thermal performance at low cost (C+50%, E+50%, P+100%).

Good **bus design** also contributed, allowing for maximum pin count used to enhance power and thermal performance. Using a wide parallel bus wouldn't provide much package benefit. An improved **control protocol** enabled the connection of thousands of chips with low cost and low pin count.

Good thermal design was more important in the ASIC era because the heat density of an ASIC is much higher than CPU/GPU/FPGA and most other chips. It reached 100 W/cm2 or more. **High thermal packaging material** (P+30%) proved to be very useful.

In 2014 the Bitcoin price crashed, and the era of function ASICs came to an end. Hashrate was already high, low-tech ASICs didn't make sense anymore, and energy efficiency became the top priority.

. . .

Phase V — Efficiency ASICs

Many of the early ASIC teams had failed by now, but several big players started to optimize efficiency, which turned out to be a difficult and long path.

First was the semiconductor process. In the function ASIC era, people used mature processes, such as 180/153nm, and later 90/80nm or 65/55nm. Driven by performance, cost, maybe other reasons. Deeper processes provided better transistors, lower power and lower cost per transistor.

The arms race finally reached deeper process nodes such as 45/40nm, 32/28nm, someone even tried 20/16nm. In 2014, 28nm was already a state-of-the-art process, 20/16nm was just coming out, but it was expensive and quite unstable still.

Reaching 28nm meant you were at the leading semiconductor process at that time, there was not much room to go deeper, so optimization of the design on 28nm became the top priority.

The **28nm process itself** provided ∼5x better efficiency compared to 130nm (C+50%, E+500%, P+100%).

The efficiency formula of a CMOS circuit is

> Efficiency $\propto TCV^2$
> (T=toggle rate, C=capacitance, V^2=voltage squared)

The **toggle rate** was already optimized since the FPGA era, now it was further optimized to improve efficiency (E+10%). **Capacitance** depends on transistor and process, optimizing it meant a redesign of the entire circuit library, which turned out to be a lot of work but provided good results (E+15%, P+15%). Normally smaller transistors have better efficiency.

Voltage was the biggest deal in all optimizations, the only 2nd order (squared) effect in all of mining. Reducing the supply voltage from 0.9V (typical at 28nm) to 0.6V (most 28nm miners) provided an advantage (E+225%). However reaching such low voltage was hard. A single optimization may have enabled a reduction of only 0.01V, meaning that many optimizations had to be accumulated to reach low voltages. This played an even bigger role in FinFET processes such as 16/14/12/10/7/5nm.

16/14nm FinFET didn't provide much benefit over 28nm at the same supply voltage. However FinFET can run down to 0.30V, which itself provided a big boost (E+400%), but was difficult to design.

Dynamic circuits were the single biggest improvement. Normally ASICs will be designed using standard cells, static CMOS design. Due to Bitcoin not needing to store data longer than 1 clock cycle, we could DFF (D-flipflop) using a dynamic circuit. A typical non-scan DFF cell is 22 transistors (22T), with a well designed dynamic circuit we could reduce the transistor count to 10T, 8T and below, with latch design we could reduce it to 4T or less. It provided significant cost reduction and power savings (C+30%, E+30%, P+20%).

Flipchip packages became common at this time. Flipchip and good substrate provided much lower electrical resistance, to enable lower voltage operation. It also provided good thermal performance on both top and bottom side. Large silicon die designs

(>100mm2) were abandoned after 28nm, because the heat density became so high that it was impossible to maintain a reasonable temperature. A small die can dissipate heat over 6 sides (top, bottom, 4 edges), but a large die can only dissipate heat over one side (top). Good packaging brought substantial benefit (E+50%) to the chip.

Power cascading was another major cost optimization. When ASIC supply voltages became less and less, using DCDC supplies became uneconomic. Energy was lost on inductors, capacitors, wires, connectors, etc. A better choice was to cascade ASIC chips together, for example a 0.6V cascade of 20 stages to reach 12V, or 0.4V 30 stages, or 0.3V 40 stages. Power cascades turned out to be difficult. They were unstable, semiconductor processes have variations making every chip slightly different. Small differences caused big problems, especially during power-up sequences in cold weather. Stabilizing needed several iterations of trial and error. IO and clocking, reset, software, temperature distribution within the machine — all became important. Balance loss meant to lose a whole chain of chips. In the end power cascading led to reduced current, increased voltage, saved cost and increased efficiency (C+100%, E+50%).

The famous **ASIC Boost** (consider it a math optimization) brought some advantage (C+15%, E+15%), but in the context of all other improvements wasn't outstanding or worthy of being considered an attack vector in our opinion. It didn't provide more advantage than voltage or dynamic circuits. The efficiency improvement of ASIC Boost was just as good as an improved **fan speed adjustment** algorithm (E+15%), or keeping **each chip at the optimum temperature** for that chip (E+15%).

Approximate adders (C+10%, E+10%, P+10%): Mining chips can tolerate an error rate of 1–2%. Making a circuit less accurate will reduce cost and provide better efficiency. Lack of toolchain support made this harder to implement than expected.

Process adjustments (E+15%, P+15%): Semiconductor processes have many parameters to adjust, such as the transistor threshold voltage, annealing temperature, poly gate size. While such adjustments can provide a 30% benefit, they need a lot of sampling and testing resources, and then analysis of the results for efficiency and performance, a process which will take several months.

. . .

Phase VI — Supply Chain

When 16nm chips became stable, the process arms race stopped. There was no better process for a long time. Volume became the top priority. Big ASIC companies rushed to book materials. From foundry to packaging house, substrate vendor, thermal material vendor and other key components. **Pre-ordering 80% of key supplier capacity** or signing **exclusive supply agreements** became important tools to compete.

By now mining occupied a significant percentage of the semiconductor industry, every supplier was aware and had to **prioritize their large customers**. Large ASIC mining companies now dominated the supply chain.

Small companies could design a chip, but couldn’t order enough capacity from a foundry. Small companies had to prepay 100% for wafer orders, where large companies could prepay as little as 0% or 10%, giving them a big cashflow advantage. The best packaging materials were sold exclusively to big companies, small companies wouldn’t even get a sample. Same for most other key components.

. . .

Epilogue

What does this mean? What’s next?
Who knows, you tell us! Come to our Telegram and let’s hash out ideas.

Chen Min
Linzhi, Shenzhen
Telegram: https://t.me/LinzhiCorp

Bitcoin Asic Mining Semiconductors Sha 256 History