艺

yì

https://scale.qihardware.org

2019 . Week 12 . Mar 22 - Mar 29

This page left intentionally blank
to power your imagination
of what interesting art, ads,
sponsorship, standard
frontmatter or blank space
should be included in
future editions of scale.

艺 yì

艺 (yì) means art, skills and craft.
They are all the same thing in Chinese

艺术 (yishu) art, 工艺 (gongyi) manual work, 手艺 (shouyi) craftmanship are all part of the same group of ideas in Chinese, expressed with 艺 (yi). In its traditional form, yi (藝) is a compound from grass (艹艸 cǎo), clouds (云 yún) and 埶 (yi) - an ancient word representing a kneeling person using its hands (丮) to model earth and wood (坴). More than just art, yi represents the talent, that is the ability to do something that should be cultivated to become a culture (文艺 wenyi).

Linzhi ASICs
Mar 29 · 4 min read

EIP 1057 (ProgPoW): Open Chip Design for 1% cost/power increase
□□□**1%**□□□□□□□□□□□□

Adding a pseudo-random program to the PoW algorithm is a key technical idea described in EIP 1057 (ProgPoW). The theory goes that the compute area of a GPU is underutilized compared to memory bandwidth. ProgPoW is then designed such that it “saturates both compute and memory bandwidth at once”.

https://github.com/ethereum/EIPs/blob/master/EIPS/eip-1057.md

> *Since the GPU is almost fully utilized, there’s little opportunity for specialized ASICs to gain efficiency.*
> *There should be little opportunity for efficiency gains compared to a commodity GPU.*
> *While a custom ASIC is still possible, the efficiency gains available are minimal.*
> *These would result in minimal, roughly 1.1x-1.2x, efficiency gains.*

Bitmain and Innosilicon both currently have Ethash systems, a good starting point for ProgPoW. What would they need to add for the new EIP 1057 compute logic?

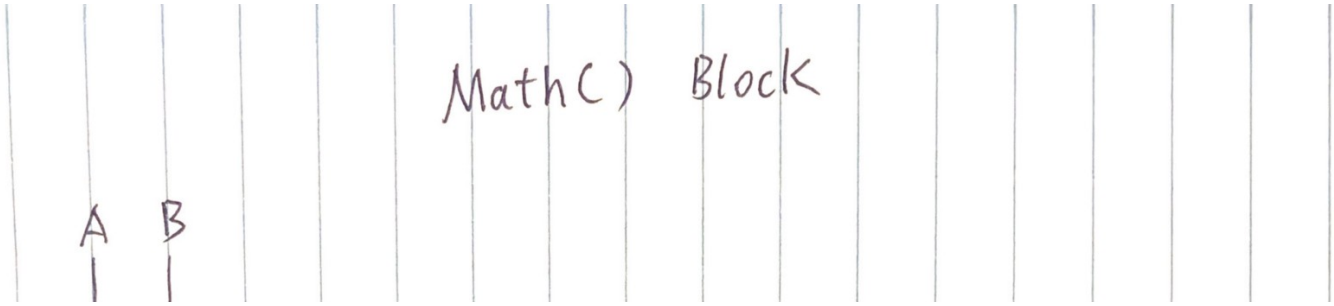
Math Block

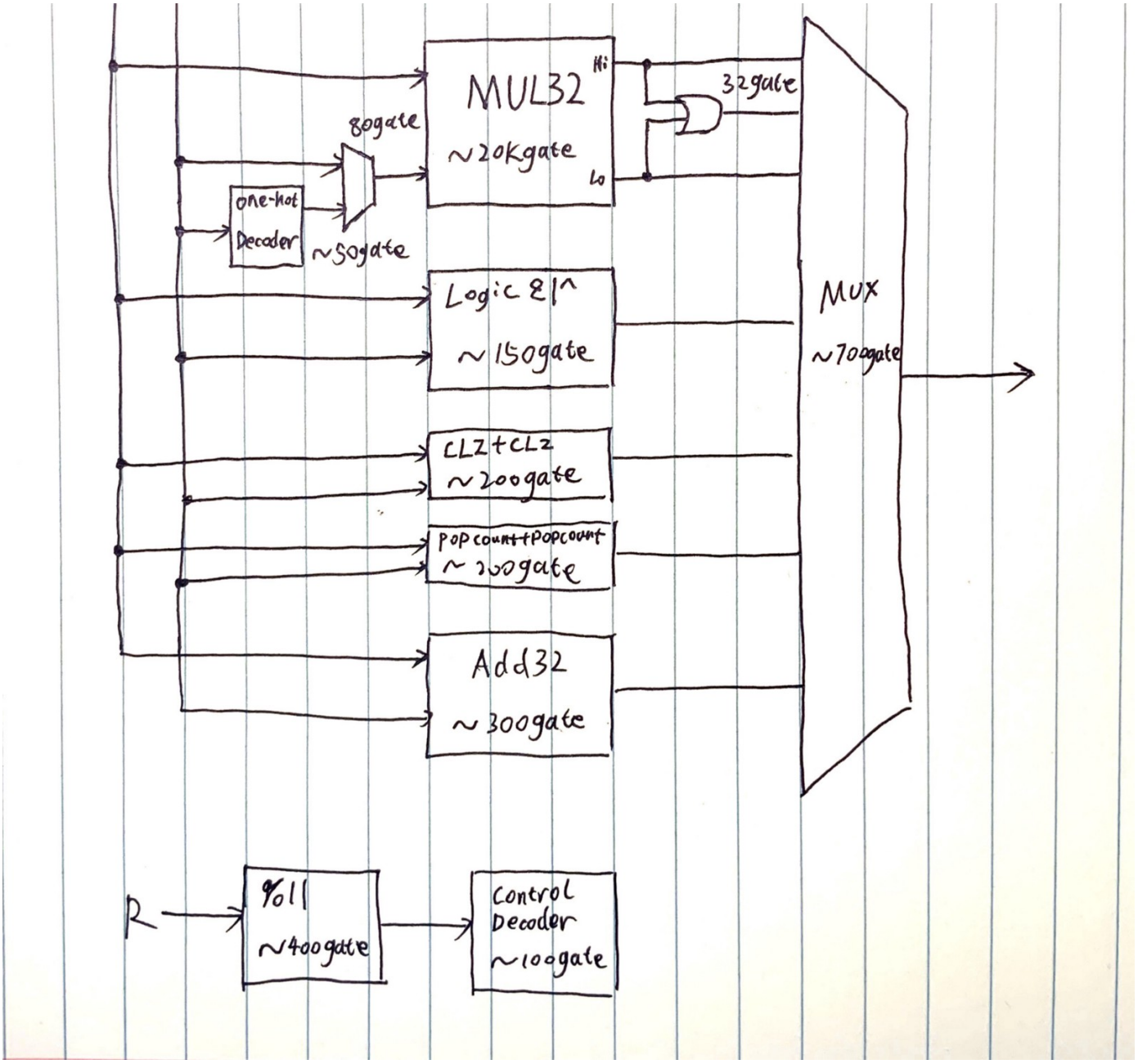
First let’s look at Math() which includes 11 different instructions

```
// Random math between two input values
uint32_t Math(uint32_t a, uint32_t b, uint32_t r)
{
    switch (r % 11)
    {
        case 0: return a + b;
        case 1: return a * b;
        case 2: return mul_hi(a, b);
        case 3: return min(a, b);
```

```
        case 4: return ROTL32(a, b);
        case 5: return ROTR32(a, b);
        case 6: return a & b;
        case 7: return a | b;
        case 8: return a ^ b;
        case 9: return clz(a) + clz(b);
        case 10: return popcount(a) + popcount(b);
    }
}
```

- modulo 11 operation, this is quite small logic, ~400gate, 1ck latency, can be a parallel process during mix read so we can hide this latency.
- 32-bit Add, simple logic, ~300gate for a fast one.
- 32-bit Multiplier, mature IP, ~20Kgate for a fast one, since multiplier only have ~4/11 activity rate, we can use a two cycle multiplier to half the area, small possibility to increase delay.
- Rotation operation can easily map to a multiplier, for example I want to calculate ROTL(0x12345678, 8), I can do 0x12345678 * 0x00000100 = 0x0000001234567800, then we just need to OR higher word and lower word together to get 0x34567812. so just cost ~160gate extra logic
- logic operation, A&B only cost 32 gate, A|B 32 gate, A^B 96 gate, it looks like three different instructions but actually extremely small on silicon (<30um²)
- clz and popcount are also very small
- We only need a multiplexer to select output.
- Total size of Math() is about 0.0015mm² on a TSMC16ULP process.
- Merge() is similar but even smaller, only shifter, adder, and tiny logic (no multipliers because constant multiply can be mapped into adder).
- Size of Merge() is roughly ~0.0005mm².

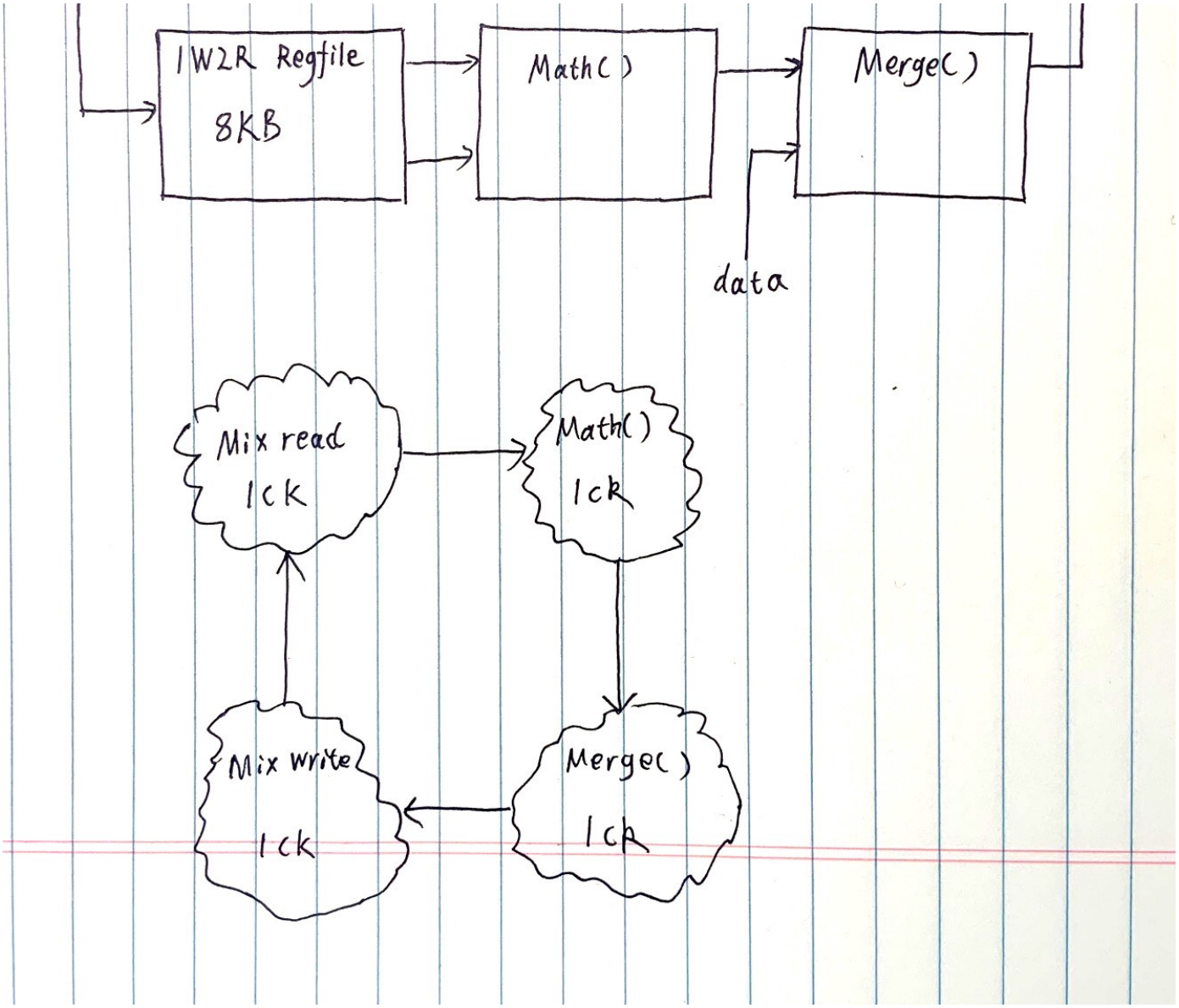




Pipeline

Now we can build our pipeline, 4-stage:

CK1: Mix regfile read, two operators in parallel (calculate %11 at same time)
CK2: Math()
CK3: Merge()
CK4: Mix write back to regfile



Task latency is 4 cycles, but we can have 4 independent threads so we can make this pipeline fully loaded.

Mix Regfile we use a 1W2R (1 write port, 2 read port) Regfile, mature IP, 8KB (for 4 threads), single cycle read/write, 1GHz operation.

If you want load/unload mix data on the fly without disturbing the pipeline, we can upgrade to a 12KB 2W3R Regfile. We then have extra read/write ports for load/unload, and 12KB is enough for 6 tasks (4 running, 1 loading, 1 unloading).

An ASIC can implement 10K sets of that block:

- running at 1GHz frequency
- 0.55V voltage (typical voltage for TSMC16ULP)
- generating ~10T Math() + Merge() throughput per second

- Power estimate roughly 3mW each pipeline, 30W in total.
- No customized circuit/layout
- All standard cells, auto placement route
- Use mature IP only
- No aggressive overclocking
- No aggressive under voltage
- ~8.4K Merge() and ~4.8K Math() per hash
- pipeline includes 1 Merge() and 1 Math()
- pipeline is only the logic part, not the memory part
- the xGB memory are not included
- we have 10T throughput on single chip asic, divided by 8.4K Merge() per hash, means 1.2GHash

Performance

The design (logic part only) can provide 1.2 GHash ProgPoW performance at 30W power.
Nvidia GTX1070Ti can provide 15.7 MHash at 115W, but including memory.

Summary

The random instruction of EIP 1057 increases die cost/power by about 1%, and causes a die increase of <1mm². The proposed open design is demonstrating a logic-only performance of 1.2 GHash at 30W and could be deployed by Bitmain or Innosilicon, resulting in a machine with about half the hashrate of their predecessors, similar to the best GPUs.

If you are an independent chip designer and want to take the ProgPoW design idea further, please get in touch. We hereby release the design into the public domain.

Hopefully the open design process can also inspire some developers to learn more about the world of chip design.

Linzhi Team, Shenzhen
Telegram: https://t.me/LinzhiCorp
email: sonia@linzhi.io

Thanks

Peter Salanki provided the initial idea for this writeup and encouraged us to do it. Alexey Akhunov always encourages everyone to think for themselves.

References

https://github.com/ethereum/EIPs/blob/master/EIPS/eip-1057.md
https://medium.com/@ifdefelse/understanding-progpow-performance-and-tuning-d72713898db3
https://medium.com/@infantry1337/comprehensive-progpow-benchmark-715126798476
https://etcsummit.com/2018-etc-summit/
Launch of Linzhi Ethash chip at ETC Summit: https://www.youtube.com/watch?v=LMofyroBfio
Linzhi Website: https://linzhi.io/

Ethereum | Progpow | Asics | Chip Design | Proof Of Work

Before you continue, an update from us

Tumblr is now part of the Oath family. Please review Tumblr's updated Terms of Service and European Privacy Policy. Due to EU data protection laws, we (Oath), our vendors and our partners need your consent to set cookies on your device to use your search, location and browsing data to understand your interests and personalise and measure ads on our products. Oath will also provide personalised ads to you on our partners' products. Scroll down to review some privacy updates and set your preferences.

Tip: Log in to your account to avoid repeating this across your devices.

How data brings you better experiences

We want to provide you with the best experience on our products. Sometimes, we show you personalised ads by making educated guesses about your interests based on your activity on our sites and products. Learn more about how Oath uses this data.

Get personalised content and ads from our trusted partners

This doesn't mean more ads, it means personalised ones. When you let our partners use cookies to collect similar to what we collect on our sites, they can provide ads that they think match your interests, and measure, report and analyse your interactions with them. Learn more about how our partners use this data, and select 'Manage options' to set your data sharing choices with our partners.

Bloggers and your data

When you visit a blog in the Tumblr network, that blog may collect more information than we do, and may provide information to third parties that we have no relationship with, including to advertisers. We aren't responsible for the information collection and use practices of our individual blogs and bloggers.

Select Accept to continue to Tumblr. Otherwise, you will not be able to access our sites and apps. Select 'Manage options' to set your data sharing choices with our partners. For more information and settings, go to the Privacy Dashboard.

Manage options Accept

My stance on progpow

Below are my reasons for supporting Progpow. This post may change over time, if I change my opinion or add more thoughts on the matter.

Ethash successor

Ethash was pretty well designed. After a couple of years, the first ASICs started coming, from two manufacturers, Innosilicon and Bitmain. These were not *dramatic* improvements, compared to other ASIC-adaptations of POW mechanisms.

- The second generation EThash asics are in the works. Rumored to be quite a lot more powerful (7x improvements are rumored). We can't really know if those are already mining somewhere.
- There exist FPGA accelerators for ETHash, which 'exploit' the fact that offloading keccak can speed up the calculation.

So at this point, while we know that ETHash was 'pretty good', we also know that it suffers from a few flaws. These are fixed by progpow.

Progpow modifies ETHash to be even more GPU-friendly. This means that the already pretty slim margins for creating an ETHash-asic become extremely slim indeed,

- which has a significant impact on the ROI for developing and manufacturing a progpow asic.
- Additionally, if Ethereum proves a willingness to switch PoW, this further disincentivizes future Ethereum asic development.

My personal belief is that no progpow asic will ever be produced – and if it *is* produced, the scale will be kept low enough to fly under the radar – since large scale production, and wide-spread use, creates the risk of the hardware becoming bricked again.

A progpow asic producer is not only in a race against Ethereum PoW developments, but also in an arms-race against Nvidia and AMD. Their next-gen asic also needs to beat the next-gen GPU.

The path already taken

Ethereum has historically been aimed to be ASIC-resistant.

We should *not* consider original intent as sacred – and please don't think that's the argument I'm making here.

However, to me, this means that the 'conservative' route forward is to keep this model. And conversely, if we are to change this model, then the burden of proof for *why* we should change this model is on the progpow opponents.

I think we should keep this aim unless we have very good reasons not to do so. A change in that policy should be an explicit decision based on rational discussion.

Decentralization

If we remain with ETHash, a miner / mining farm have the following option:

- Buy asic(s) from one of three manufacturers (Innosilicon, Bitmain or Linzhi). In practice, if Linzhi

develops and releases a next generation asic miner, that means the miner needs to get their latest asic to stay in the game.

That is an incredible monopoly to hold –

- Can they possibly supply the entire mining ecosystem?
- Would they even want to - since that would undercut their own prices?
- Could you import those into any country, with import/export regulations and taxes?

If you are on their blacklist, they might not even sell to you.

Today, mining farms exist geographically spread – north america, iceland, china etc.

If we stay on ETHash,

- mining will be centralized to a group of chinese miners who have the right ties to get hold of the latest hardware.

If we switch to Progpow,

- existing mining farms can continue to operate where it makes sense for them to operate; depending on factors such as cost of electricity, cooling, space etc.
- GPUs can be acquired from anywhere in the world, and although there are two manufacturers (Nvidia/AMD), Ethereum miners are only a small part of the global market of GPU cards.
- There can be global mining competition, instead of an oligopoly.

Proof of stake transition

An ASIC ecosystem of miners means that there is nothing else for that piece of hardware to do, other than mine that particular coin. There are two sides to that coin:

1. Since the asic miners are more vested in the coin, they are incentivized to not mount attacks against that coin, since it lowers the value of their hardware. This is one argument in favour of ASIC.
2. Since the asic miners are more vested in the coin PoW, they are incentivized to mount any and all resistance possible against switching to PoS. This is an argument against ASIC.

These arguments are both good, in my opinion. However, the first argument means that ASICs would be less incentivized to attack Ethereum than GPU miners. During the lifetime of Ethereum, this is not something we have had problems with. So while I concur that it's a theoretically valid point, I would not choose the benefit of (1) at the expense of (2).

The GPU ecosystem, on the other hands, can find other uses for their hardware when Ethereum transitions to PoS.

Why we should adopt progpow

So, above are the reasons about why I support progpow. There are a couple of more reasons why I also think that Ethereum should go forward with the switch to progpow.

- Community. As far as I've seen, most signals point to 'the community' wanting to go ahead with progpow. This has been shown both via coinvote and a running mining signalling.
  - The carbon vote at the time of writing this, has `94%` yes. That is almost *3 million* ether voted yes, against `184K` ether voting no.
  - The miner vote at the time of writing, shows that `77%` of all miners voted, and all votes were pro progpow.
- Known "evils". If we switch to progpow, we will keep the 'same' ecosystem of miners around. These miners represent a large base of active Ethereum users, many who have been Ethereum enthusiasts

from early on.
- Small overhead. I have taken part of the behind-the-scenes work on all hardforks since TheDao fork. I set up Hive to perform blackbox consensus-testing across clients, performed differential fuzzing on EVMs for the last couple of hardforks, and personally discovered a number of consensus vulnerabilities in mainnet clients. In my view, a hardfork which updates ethash-hashimoto to ethash-progpow will be the simplest fork we have ever done. The reason for this is that it touches nothing on the complicated parts: The execution engine (EVM) and the state transition function (rewards/refunds/post-process-cleanups). It is a simple matter of verifying a different block envelope.

I'm not married to the progpow proposal – this is not a hill I'm willing to die on. But based on technical merits and community signals, I think it is the right thing to do, and I think it's worth doing,

---

Edits: I originally wrote Intel when I meant Nvidia

2019-03-28

tweets

Tweets by @mhswende

favorites

Favorites

Contact

- Twitter : @mhswende
- Email : martin [at] swende.se

About the site

This site was created using Jekyll, Bootstrap, Retriever and assorted tools.

Obelisk GRN1 Chip Details

Mining

Taek 2019-03-20 22:02:37 UTC #1

I wanted to come forward and share some of the details about the GRN1 chip that we are making. Obelisk is here to move the space forward, and a part of that is increased transparency around the ASIC design process so that the community can better understand what goes into an ASIC project and make more informed decisions around proof of work. The chip was co-architected between Obelisk and ePIC Blockchain, and ePIC Blockchain (**www.epicblockchain.io**) led the implementation of the chip.

The Obelisk GRN1 specs have been updated to 150 graphs per second at 800 watts. The first units are on track to begin shipping early October, and the final units are on track to ship in late October. Our specifications are based on a pessimistic interpretation of simulations that are generally +/-10% for speed and +/-30% for power.

Because of NDAs, I can't share exact information about our design choices. But I can link to public information and speak broadly about the implications. The GRN1 chip is a single-die cuckatoo31 miner. The chip we made has a full 512 MiB of memory on board. Our technology partners have verified that we are within the margins of test, packaging and yield boundaries, and that the final product will be fully viable. Beyond substantially increased speed and efficiency, using a single chip also reduces manufacturing complexity and enables a more reliable final product.

We believe that the most efficient Cuckatoo32 miner is also a single-die chip using today's technology. And we also believe that the most efficient Cuckatoo33 and Cuckatoo34 miners are single-die chips using today's technology. We've done substantial investigation into the memory capabilities of modern foundries, but before I go further, I want to provide some basic statistics:

16 nm lithography process - WikiChip

The 16 nanometer (16 nm) lithography process is a full node semiconductor manufacturing process following the 20 nm process stopgap. Commercial integrated circuit manufacturing using 16 nm process began in 2014. The term '16 nm' is simply a...

7 nm lithography process - WikiChip

The 7 nanometer (7 nm) lithography process is a technology node semiconductor manufacturing process following the 10 nm process node. The term '7 nm' is simply a commercial name for a generation of a certain size and its technology and does not...

According to the above articles, the smallest SRAM cell at TSMC 16nm is 0.074 square micrometers. And at 7nm, the smallest SRAM cell at TSMC is 0.027 square micrometers. The largest chips have over 800 square millimeters of area. If you do the math, this means that a 16nm chip has a maximum theoretical memory size of about 1.3 GiB, and a 7nm chip has a maximum theoretical memory size of about 3.5 GiB.

In practice, you cannot create a 16nm chip with 1.3 GiB of memory. A chip is a lot more than just an array of bits, but in practice at 16nm there is more than enough room to do a full CC31 mining algorithm, including cycle finding. Beyond that, the foundry does have limits to how much memory they can support, and every piece of memory that you add impacts yields. Due to NDA's, I'm unable to share the maximum amount of memory we believe we could put on a single 16nm die, but I am comfortable saying that it's more than 512 MiB. Putting this much memory on a single die does impact yields, however the impact is small enough that our chip remains viable. As a process matures, yields improve substantially, and the TSMC 16nm process is quite mature at this point.

At 7nm, we believe that we could do all the way out to Cuckatoo33 without needing to make a 2x time-memory trade-off. To say that again, we are confident that you can make a single-die ASIC to do CC31, CC32, and CC33. We also believe that CC34 is possible, though at this point substantial time-memory trade-offs are required.

Our Cuckatoo31 chip has a very interesting property relative to typical single-die ASICs - the heat signature. A typical highly optimized Bitcoin mining chip produces between 0.3 and 0.5 watts per square millimeter. Because of this heat profile, a typical $1200 miner may cost $1200 per year to operate, primarily due to the cost of electricity. For an American mining on typical consumer electricity rates, that annual cost is more like $2400 per year, meaning that consumers really cannot afford to mine at home.

Our Cuckatoo31 chip has a heat signature that's less than 0.1 watts per square millimeter. This translates to much lower electricity costs. The same $1200 spent on a mining device results in electricity costs that are closer to $400 per year for a typical mining farm, and $800 per year for a typical consumer. The total cost of ownership gap between a consumer and a professional farm is substantially lower for cuckatoo miners. This is a key distinguisher for the Cuckatoo31 algorithm and a way that Grin stands out.

There's another very interesting aspect to Cuckatoo31 specifically. When optimizing over total cost of ownership, wafer pricing becomes a lot more important. The primary cost of the machine throughout its lifetime is not electricity, but silicon. 16nm silicon is cheaper and more accessible, which means that 16nm chips are more competitive, and depending on price, potentially even strictly superior.

This is fantastic for competition. The development and tooling costs for 7nm are far higher, and the 7nm technology is a lot more exclusive. These higher costs mean that there is much less room for competition. If 16nm chips are potentially superior, smaller companies can compete with lower initial investment and the competitive environment for Grin will be more vibrant.

At cuckatoo32, 16nm is no longer the ideal node, because the memory takes up too much space and doesn't leave enough room for all the other elements of a grin miner. Even switching to cuckatoo32 means that new competitors have to be at a more advanced node, which restricts the competitive environment.

The phase-out also has another adverse effect. Manufacturers are forced to choose an algorithm to target. This complicates the game theory. Manufacturers have to choose a specific level to target, and since miners that target lower levels are more competitive (this is just the nature of the hardware and the cost structure), being able to support a higher level means you will not be competitive at the lower levels. This also harms the overall competitive environment, you want the economics and profit models to be as simple and low risk as possible to encourage competition.

I know this information is a lot different than what many people were expecting. Most manufacturers are not up-front about the nature of their hardware, but we really would like to see Grin succeed, and we would like to do so by collaborating with the Grin community and letting them know what's going on before it happens. We're hoping to open a dialog between manufacturer and community, and we're hoping to give the community all the information it needs to make informed decisions about the future of the proof of work ecosystem that protects the consensus layer. At the end of the day, Obelisk wants to see Grin succeed and we believe that being open is the best way to empower the Grin community. We're looking forward to your thoughts and discussion.

---

tromp 2019-03-20 23:28:56 UTC #2

Thanks for sharing that, David!

Needless to say, this throws quite a spanner in the works of Grin's anti-single-chip-ASIC stance, as laid out in **Scheduled PoW upgrades proposal**, and as further discussed in **Cuckatoo32 feasibility**.

Your point about the very different heat density is well taken; that kind of invalidates my claim that single-chip cuckoo ASICs are not meaningfully different from SHA256 ones.

Perhaps I should admit defeat and accept the reality of single-chip ASICs. If what you say is true and none of the next few size upgrades forces a memory IO bottleneck, then we might be better off disabling future phaseouts from some point on.

This would invalidate our **commitment to ASIC manufacturers**, so any such change would ideally need to meet with their approval.

I welcome community discussion on the best way forward.

---

spartadata 2019-03-21 00:55:48 UTC #3

I believe you are leading and paving the road for other ASIC makers to adopt the clarity and transparency you share openly with the GRIN community and how you have been doing with SIA Asics as well. Not only does it build miner (customer) loyalty and trust but it builds credibility and transparency that we can all appreciate and hope to expect to see others do as well. Sadly I believe the transparency can spoil many of us that will aim to seek this sort of information from other asic makers and not know if that will happen or not. As such it falls on GRIN as a community and coin to establish good ethics policy and guidelines which would be a positive step forward to guiding other makers how they should disclose specifics of production quantities and overall platform within what's reasonable.

Cheers to Transparency and clear helpful Disclosure.

Here is the GRN1 with the updated specs realtime profitability stats:

Realtime ASIC miners profitability

Live profit estimation of all known ASIC miners, updated every minute.

---

timolson 2019-03-21 17:52:09 UTC #4

> Taek:
>
> The primary cost of the machine throughout its lifetime is not electricity, but silicon.

We found this to be true for CryptoNight as well, which also has a large ratio of memory-to-logic on the die. Many people repeat the mantra of "hash-per-watt" but if the miner's lifetime is short, 12 to 18 months, then capex is more than lifetime opex, even for a Bitcoin miner.

---

rodarmor 2019-03-21 19:24:13 UTC #5

For phase-out to accomplish its goals, it must prevent single-chip ASICs from being competitive, but not hurt network security by increasing graph search time of multi-chip ASICs.

Also, to avoid centralization, it should operate on autopilot, and not require continual tweaking via hard or soft fork.

As this post demonstrates, it is likely impossible to adequately anticipate future technical developments such that the phase-out parameter can be fixed in advance.

To add a little bit of color from my own experiences, one ASIC manufacturer I spoke with about the possibility of producing Grin ASICs demurred on the basis that they did not expect to be able to produce a CC31 ASIC significantly better than a GPU. Another was only comfortable producing a CC31 mean-miner.

Ideally, a large number of manufacturers would be able to produce CC ASICs. From my own experience, this is not the case, and phase out would likely exacerbate the situation.

The commitment to ASIC manufacturers permits delaying phase out of smaller graph sizes if better-than 1 GPS miners are not publicly available. It is stretching the language a bit, but I think that delaying phase out of smaller graph sizes indefinitely is not a gross violation of this commitment. Especially if no manufacturer has invested money to research or produce a CC32 ASIC. (Separate from investments in a CC31 ASIC that would be repurposed for an eventual CC32 ASIC.)

Additionally, my own experience was that forecasting mining profitability and investment was greatly complicated by the limited lifespan of Grin mining hardware under phase-out. CC31 ASICs, for example, only mine for something like 6 weight-adjusted months at best. This cuts off the long-tail of mining profits, and provides a very short period for earning a return on investment. Deterring investment thusly this works against the goal of securing the Grin network, and ensuring that GPU-powered 51% attacks are uneconomical.

---

tromp 2019-03-21 22:16:58 UTC #6

> rodarmor:
>
> not a gross violation of this commitment. Especially if no manufacturer has invested money to research or produce a CC32 ASIC

The goal of the commitment was to protect large CC ASIC investments. I believe that such investments have already been made for C32, based on our planned phase out of C31. In which case we shouldn't undo the latter.

But it's quite possible that no large investment has been made into C33+ ASIC design (we could ask manufacturers to speak up if they have). In absence of such investments, we can amend the commitment to indefinitely shelve the phase out of C32 and beyond. That would give us time to see how well the various current C31/C32 designs perform in practice, and whether freezing the PoW at C32+ would be preferable.

In theory, if we're happy to have single chip ASICs, we could even go back to C30+ in 2021.

The commitment has served its purpose in making ASIC manufacturers confident to invest in the current crop of C31/C32 designs. Now perhaps it's best to increase our options for the future health of the Grin mining market, by shelving the phaseout of C32 until further notice.

---

etocs138 2019-03-21 23:17:07 UTC #7

Circa 1492

Columbus "Hey, I think I am off on the bearing of the North Star…just by a couple of degrees"

His assistance "Oh, good one Chris, glad you spotted that…Instead of the Bahamas, we could have landed in Greenland with our beach ball and flip-flops"

Columbus "How about just a 1 degree adjustment? …that would take us to Florida"

His assistance "Nah, I hear it is crowded there during Spring Break".

Columbus "Ok, let's go make history, on we go with this one adjustment".

It takes a brave man to charter a course. It takes a braver man to stay committed. The bravest

Of all, are the ones that can internalize & course-correct.

---

asic_king 2019-03-21 23:36:21 UTC #8

First of all, thanks for sharing. We at Innosilicon also believe that a more open environment will help the ASIC companies gain trust from the blockchain community. In joining this effort, I want to give our view and progress on the development of Innosilicon Grin ASIC.

Overall, I believe the desperate move to pack a monstrous amount of SRAM (512MiB for Cuckatoo31 and for 1GiB for Cuckatoo 32) is the wrong approach against the evolution path envisioned by the core team. This will result in high risk and short-lived ASICs （about 4 months at most for Obelisk GRN1） which is why capex becomes more than opex during the lifetime of a miner. Innosilicon fully embraces the evolution path as planned and believes a multi-chips approach of one ASIC coupled with high speed DRAM results in less waste of ASICs and more stable and predictable mining.

Our Grin ASIC will be Cuckatoo31/32 compatible which gives the miner longevity and makes capex less than opex. More details about the chip will be announced in April. The multi-chips approach minimizes the investment on fixed purpose ASICs and allow the DRAM to be repurposed even after Cuckatoo 32 is phased out. This approach is not only more sustainable and greener but reduces risk capital investment by the miners. It also sets us apart from the other Crypto ASIC companies because Innosilicon has been working on high speed memory interface for over 12 years and holds the most advanced high-speed memory interface IPs.

To put 512MiB of SRAM in perspective, AMD's latest Zen based server processor has up to 64MiB of L3 Cache. Intel's most powerful Xeon E7-8891 server processor has 60MiB of Cache. Innosilicon has successfully put 160 MiB of SRAM on our ZCash A9 series using a process that is denser than 16nm. From our first-hand experience in embedding large SRAM we believe going for 512 MiB or over 1 GiB is a dead end. The yield loss will be exponentially higher. TSMC is touting the success of its 7nm process and cites yield of 32MiB (256 Mbit) SRAM at 76%. (**https://www.eetasia.com/news/article/tsmc-unveils-plans-for-7-12-22nm-nodes**) Imagine what kind of yield you will be looking at for 512MiB and beyond. The bad yield will force one to resort to higher operating voltages which helps to improve yield but power consumption will suffer. Thus I also believe 800W grossly underestimates the power consumption of the final product.

Again, I think it is healthy that we are discussing this topic in a open forum and look forward to more open discussion.

---

rodarmor 2019-03-22 01:06:42 UTC #9

@asic_king , thank you for your perspective! I think it's important to hear from a diversity of manufacturers.

> Overall, I believe the desperate move to pack a monstrous amount of SRAM (512MiB for Cuckatoo31 and for 1GiB for Cuckatoo 32)…

I don't think I would characterize is as a "desperate move". It is an equally valid point in the design space.

> …is the wrong approach against the evolution path envisioned by the core team. This will result in high risk and short-lived ASICs（about 4 months at most for Obelisk GRN1）

The surest way to reduce risk and increase ASIC lifespan would be to cancel phase-out entirely.

> which is why capex becomes more than opex during the lifetime of a miner.

I think that it can be debated whether a high cost of ASICs vs other capital expenditures is desirable, due to the way that it aligns incentives between miners and the community. ASICs are a form of capital that is coin-specific, i.e. in this case can only be used to mine Grin. Other forms of capital and operational expenditures, such as datacenter real estate and buildings, provisioned power capacity, and electricity, can be repurposed to mine any other coin. Thus, a miner who has invested a large amount in ASICs will not be inclined to do protect their investment and refrain from doing anything that might hurt the coin.

> Innosilicon fully embraces the evolution path as planned and believes a multi-chips approach of one ASIC coupled with high speed DRAM results in less waste of ASICs and more stable and predictable mining.

As I mentioned, the surest way to reduce ASIC waste and make mining more stable and predictable is to cancel phase-out.

> Our Grin ASIC will be Cuckatoo31/32 compatible which gives the miner longevity and makes capex less than opex.More details about the chip will be announced in April. The multi-chips approach minimizes the investment on fixed purpose ASICs and allow the DRAM to be repurposed even after Cuckatoo 32 is phased out.

Can you share the cost, power consumption, and graphs-per-second of your upcoming miner? An important point in this discussion is whether or not multi-chip ASICs will be profitable in the face of single-chip ASICs, and without the specs of some multi-chip ASICs, this is something that we cannot evaluate.

I think that given these plans to produce a Cuckatoo32 miner, it might not be fair to cancel phase out entirely. However, halting it at CC32 still might be reasonable.

> This approach is not only more sustainable and greener…

Again, cancelling phase-out is the most sustainable and green option.

> but reduces risk capital investment by the miners. It also sets us apart from the other Crypto ASIC companies because Innosilicon has been working on high speed memory interface for over 12 years and holds the most advanced high-speed memory interface IPs.

Individual manufacturers possessing intellectual property that enables them to produce more competitive ASICs by monopolizing specific technologies is a net negative.

> To put 512MiB of SRAM in perspective, AMD's latest Zen based server processor has up to 64MiB of L3 Cache. Intel's most powerful Xeon E7-8891 server processor has 60MiB of Cache. Innosilicon has successfully put 160 MiB of SRAM on our ZCash A9 series using a process that is denser than 16nm. From our first-hand experience in embedding large SRAM we believe going for 512 MiB or over 1 GiB is a dead end. The yield loss will be exponentially higher. TSMC is touting the success of its 7nm process and cites yield of 32MiB (256 Mbit) SRAM at 76%. (**https://www.eetasia.com/news/article/tsmc-unveils-plans-for-7-12-22nm-nodes**) Imagine what kind of yield you will be looking at for 512MiB and beyond. The bad yield will force one to resort to higher operating voltages which helps to improve yield but power consumption will suffer. Thus I also believe 800W grossly underestimates the power consumption of the final product.

I don't think that David has specified whether or not the GRN1 is using 16nm or 7nm, but it seems that 512MiB at 16nm is possible. For CC32, if the full 1024MiB is impossible to fit on a single chip, time-memory tradeoffs are possible to keep the design on a single chip with reasonable yields.

Edit: I also wanted to note that I'm not a current nor planned investor in Obelisk Grin ASICs.

---

asic_king 2019-03-22 07:57:47 UTC #10

I wanna highlight the importance of fair play and free market competition for all for anyone who truly want to support the Grin PoW. The Grin core team had formally outlined a very clear goal in its PoW design and a phase-out plan to encourage the multi-chip evolution path using external commodity memories with many months of serious discussion before the final decision for Grin. Whoever wants to supply the Grin mining equipment should adhere closely to the announced PoW rules instead of attempting to tailor the PoW for one's self interest. Whether using multi-chip strategy or single chips is vendor's own technical choice. Afterall, you make the chip to support Grin, not the other way around.

We think multi-chip using external memory can yield very competitive results for both CC31/32 and beyond. We will announce more specifics in April about our design. On the other hand, people who attempt the single ASIC chip approach should follow the same evolution rule to meet GRIN PoW requirement. I just want to point out the many hazards they will be facing. One can try the time-memory tradeoffs to limit the SRAM size beyond 512MiB to support CC32, but performance penalties are very steep according to our analysis. I think these risks are becoming apparent over time and single chip design won't be competitive comparing to multi-chip with external memory approach. You need to be responsible for your design approach and it's not right to lobby the core team into delaying the roll-out of the phase-out plan. Innosilicon is very good in single-chip design but we don't think single chip is the best approach for GRIN. We have come out to support the vision of the Grin team by providing more responsible and more robust solutions. We will demonstrate to the world that the multi-chip ASIC solution will work better and work longer. It fits well with the original vision of the team and will attract support from the mining community.

I also want to comment on the topic of memory interface monopoly since it is mentioned, which is funny. Commodity memory usage is nothing special nowadays. Innosilicon is a strong design company who masters high-speed memory IPs, but there are plenty of IP vendors in this space like Rambus, Synopsys and Cadence that offers similar IPs. Memory interface technology is widely accessible in the industry. Nothing prevents anyone from competing in the GRIN PoW mining field. Healthy competition on a level playing field encourages design excellence and innovation for Grin. All in all, if you are a GRIN supporter, you just need to play by the GRIN rule. I believe more ASIC companies trying to work on GRIN design will help it to GRIN to gain more popularity and Innosilicon is happy to be part of the GRIN family.

---

debilnypes 2019-03-22 17:10:21 UTC #11

This post was flagged by the community and is temporarily hidden.

Shobji 2019-03-22 18:01:13 UTC #12

Thanks for the detail and explaining every detail, as said by you this update will empower the Grin community. 👍

---

rodarmor 2019-03-22 20:32:28 UTC #13

> Whoever wants to supply the Grin mining equipment should adhere closely to the announced PoW rules instead of attempting to tailor the PoW for one's self interest.

I don't believe the issue is that anyone is trying to tailor the PoW to their own self interest. The issues, as I understand them are:

- Phase-out cannot prevent single-chip ASICs, which was its original intention
- Phase-out may result in periods where single-chip ASICs dominate, and periods where multi-chip ASICs dominate, which increases hashrate fluctuations
- Phase-out makes mining operations less profitable, due to limited lifespan of miners
- Phase-out makes planning investments unappealing, due to uncertainty of projecting lifespan of miners

> We think multi-chip using external memory can yield very competitive results for both CC31/32 and beyond. We will announce more specifics in April about our design.

I would love to be proven wrong, but my assumption, until specifics are available, is that multi-chip designs will not be competitive due to increased latency.

> I also want to comment on the topic of memory interface monopoly since it is mentioned, which is funny.

Earlier you said:

> It also sets us apart from the other Crypto ASIC companies because Innosilicon has been working on high speed memory interface for over 12 years and holds the most advanced high-speed memory interface IPs.

In other words, Innosilicon is exploiting patent and copyright-granted monopoly power to gain an advantage over other manufacturers. Many manufacturers do this, so I don't want to imply that Innosilicon is especially bad in this respect. However, it is not healthy for the coin if producing a competitive Grin ASIC requires acquiring large amounts of expensive IP in order to compete.

> Commodity memory usage is nothing special nowadays.

This contradicts your statement that Innosilicon has a strong advantage here.

Taek 2019-03-23 16:03:27 UTC #14

> asic_king:
>
> First of all, thanks for sharing. We at Innosilicon also believe that a more open environment will help the ASIC companies gain trust from the blockchain community. In joining this effort, I want to give our view and progress on the development of Innosilicon Grin ASIC.

Welcome to the community, and thanks for taking the time to comment! I have long wished that more manufacturers than Obelisk would get involved in the discussion and provide a more transparent view into the world of hardware, so that cryptocurrency communities can make more informed decisions, and also so that the historically adversarial relationship between hardware companies and cryptocurrency communities can become a more collaborative one.

> asic_king:
>
> From our first-hand experience in embedding large SRAM we believe going for 512 MiB or over 1 GiB is a dead end. The yield loss will be exponentially higher. TSMC is touting the success of its 7nm process and cites yield of 32MiB (256 Mbit) SRAM at 76%. (**https://www.eetasia.com/news/article/tsmc-unveils-plans-for-7-12-22nm-nodes**) Imagine what kind of yield you will be looking at for 512MiB and beyond.

This yield quote is on TSMC's 7nm process from more than a year before the 7nm process was cleared for mass production. During process bringup yields are usually incredibly low - part of process bringup is improving the yields, and I believe this statistic was chosen to be published because many people thought 7nm was very far away at the time, and yields that good already meant that 7nm was getting very close.

Yields today for TSMC's 7nm process are substantially better. But that's also not the right comparison point anyway, because Obelisk is using TSMC 16nm, one of the most mature and highest yielding processes in the industry. So not only is it an unfair and misleading statistic to asses 7nm, it's also unfair an misleading because it's comparing 7nm statistics to a 16nm chip.

Here is a newer article. Though it doesn't give specific yield statistics, it does mention AI chips that go up to 2 gigabits of SRAM, which is 256 MiB. Clearly, this volume of memory is not as unprecedented as Innosilicon is suggesting. **https://www.chipestimate.com/The-New-Deep-Learning-Memory-Architectures-You-Should-Know-About/eSilicon/Technical-Article/2018/10/16**

> asic_king:
>
> The multi-chips approach minimizes the investment on fixed purpose ASICs and allow the DRAM to be repurposed even after Cuckatoo 32 is phased out.

The Bitmain Ethereum miner has DRAMs on it. Those DRAMs cannot be repurposed by a consumer, if you were to repurpose those you would need a professional shop to do it, and you would also need special firmware, and probably a custom board to receive those DRAMs. Bitmain did it this way because it's cheaper and more efficient to use non-repurposeable DRAM for specialty applications. Doing the repurposing would almost certainly be more expensive than just buying brand new DRAMs - in the case of the Bitmain Ethereum miner, that DRAM is not practically repurposable and nobody is going to try.

I had assumed that Innosilicon would be using a similar strategy to save on cost, power, and manufacturing complexity. Innosilicon: are you suggesting instead that your miner will have DRAM that is intended to be repurposed, and if so, how difficult will it be to repurpose and what application / hardware could it be repurposed for? Will consumers be able to repurpose the memory without professional help?

> asic_king:
>
> Innosilicon is a strong design company who masters high-speed memory IPs, but there are plenty of IP vendors in this space like Rambus, Synopsys and Cadence that offers similar IPs. Memory interface technology is widely accessible in the industry.

It may be widely accessible, but it's also expensive. Certain manufacturers, namely Innosilicon, have this IP in-house and do not need to pay for it. For everyone else, it is an expense that increases the barrier to entry and makes competition more difficult. This goes against Grin's goals of having a competitive ASIC ecosystem with many manufacturers.

> asic_king:
>
> I wanna highlight the importance of fair play and free market competition for all for anyone who truly want to support the Grin PoW. The Grin core team had formally outlined a very clear goal in its PoW design and a phase-out plan to encourage the multi-chip evolution path using external commodity memories with many months of serious discussion before the final decision for Grin. Whoever wants to supply the Grin mining equipment should adhere closely to the announced PoW rules instead of attempting to tailor the PoW for one's self interest. Whether using multi-chip strategy or single chips is vendor's own technical choice. Afterall, you make the chip to support Grin, not the other way around.

Obelisk's current roadmap includes both a CC31 miner in October, and a single-die CC32 miner in April. Similar to the performance difference between the Antminer S7 and the Antminer S9 (both TSMC 16nm products), we expect computational speed and efficiency to roughly double. Because CC32 is more

computationally complex, these things cancel out. Our internal target for our CC32 miner is 200 graphs per second at 1000 watts.

The phase out creates a planned obsolescence for our customers buying the GRN1, and that harms the value of the GRN1. However, we believe that the total block reward for AF-CC31 still makes sense to move forward with production, provided that we are careful with total production volumes. At the current price for Grin, we would probably ramp back a bit from the originally planned 10,000 units, though we have not yet fully determined what is appropriate. It's also difficult for us to determine production volumes when our competitor - who claims to have an earlier shipping date (summer vs. fall) - has not yet released an indication of what the specs are or how many machines they intend to produce.

Obelisk currently believes that we will be able to capture the CC31 and CC32 markets both regardless of the phase-out, however that requires charging our customers more than twice as much for NRE over the next 18 months. CC32 also requires moving away from 16nm in order to avoid TMTO, which means tape-out costs will be higher and barrier to entry will be more difficult for competitors. This is actually good for Obelisk, as it means less competition and higher margins, but I don't think it's good for the Grin network.

The original goal of the phase-out was to ensure that single-die ASICs could not be created for the Cuckatoo algorithm. I had protested this at the time, though because I didn't have strong yield statistics and we hadn't done the full work on analysis yet, I was unable to confidently assert that single-die ASICs would be viable at Cuckatoo31 and beyond. Obelisk has since shown this, and we are confident asserting that we do not believe multi-die ASICs will be competitive at any Cuckatoo before Cuckatoo35 thanks to the sheer amount of memory you can fit on a single die, and thanks to the fact that yields are in fact not as bad as many fear.

For those still doubting that Obelisk's chip is viable, I have included a link to a third-party audit of our chip that we had done, both to assure ourselves that we were not making a mis-step, and to assure others that we are a world-class team that is staying within the limits of the technology that we are working with even when we are pushing the boundaries of what people thought was possible: **https://pixeldrain.com/l/o3hi6WtF#item=0**

> tromp:
>
> The goal of the commitment was to protect large CC ASIC investments. I believe that such investments have already been made for C32, based on our planned phase out of C31. In which case we shouldn't undo the latter.

I fully agree that Grin needs to protect large CC ASIC investments, and that changing the promise creates uncertainty and harms manufacturer willingness to participate. However, it is not clear to me that delaying the phase-out of Cuckatoo31 is actually damaging to anyone. Cuckatoo32 miners are already allowed on the Grin network today, and in fact get to mine at an advantage. That is, from a per-edge basis, Cuckatoo32 solutions have more weight than Cuckatoo31 miners.

Because this bonus weight exists, the primary advantage of the phase-out of Cuckatoo31 to another manufacturer is that competition is eliminated. And that as the sole advantage goes against the stated goals of the Grin ASIC-friendly PoW algorithm - to foster competition between ASIC manufacturers.

The phase-out, to the best of my knowledge, does not exist to make miners obsolete, but instead to prevent single-die ASICs from entering the marketplace. But as Obelisk has discovered, Cuckatoo32 and Cuckatoo33 are both feasible to do without a 2x TMTO on a single-die chip using today's technologies. Even where big investments have been made, I believe that the primary effect of eliminating the phase-out will be increasing competition overall.

I believe, based on several conversations I've had with people looking to buy Grin miners, Bitmain, Whatsminer, and Canaan all considered making a Grin ASIC, and all decided against it, with the primary reason being the fractured block reward. With the block reward being spread over 3 different algorithms (Cuckaroo29, Cuckatoo31, Cuckatoo32) over the typical expected lifetime of a miner, the incentive for manufacturers to enter is substantially reduced.

I believe that it would help the discussion a lot if Innosilicon published more information about their miner. As of now, we don't even know the full range of algorithms that it is capable of targeting. Innosilicon continues to claim that they will be able to ship by the end of the summer. In order for that to be possible, Innosilicon will need to have finalized their architecture already and be working on implementation. So these questions at this point in time should all be reasonable:

- Is the miner a mean miner or a lean miner?
- Does the chip itself have a significant amount of memory on it?
- Is the miner capable of targeting Cuckaroo29 in addition to Cuckatoo31 and Cuckatoo32?
- How many graphs per second are you expecting (+/- 25%) on Cuckatoo31? Cuckatoo32? Cuckaroo29? And what ballpark (+/- 50%) are you expecting for power consumption?

> asic_king:
>
> We will demonstrate to the world that the multi-chip ASIC solution will work better

It does not seem like Innosilicon is concerned about Obelisk's single chip designs. Based on their own confidence, it does not seem like they believe nixing the phase-out would change the dynamics between Obelisk and Innosilicon, and as best I can tell there is no other ASIC manufacturer seriously considering Grin at this time.

Grin Improvement Proposal 1: put later phase outs on hold and rephrase primary PoW commitment

---

etocs138 2019-03-22 23:55:03 UTC #15

In the semi industry there are a couple of acronyms used DTCO (Design Technology Co-Optimization) and PPAC (Power, Performance, Area, Cost). Time will tell who did their homework on this one. But the it feels like tailwinds are behind the single-die implementation. Innovation and out-of-box thinking will prevail.

Lolliedieb 2019-03-23 06:01:31 UTC #16

The thing is: here we got two companies now with different strategies. One that will have a single chip doing C31 very quick but C32 will need an other chip, the other with a multi purpose chip, compatible to the phase out / in but likely slower on C31 or not that efficient.

Would it be fair to change the rules now?
The multi chip will have an advantage over GPUs if its released first and then again one when the phase starts and the single chip will rule C31 from release until the end. Note that phase out is slow, if you buy first batch its likely you got a livespan of 8-9 month until its unusable.

In this picture I see times working well for any investor and so I think cancelling the phase out would invalidate too many strategic desisions already taken. C31+ was picked asic “friendly” and with predictable future evolution rules to guaranty enough predictability so ASIC development is worth the effort. It should be sticked to that.

---

tromp 2019-03-23 15:30:19 UTC #17

> Taek:
>
> Our internal target for our CC32 miner is 200 graphs per second at 1000 watts.

Can you reveal at what process?

> Taek:
>
> that requires charging our customers more than twice as much for NRE over the next 18 months

I had to look that up and found **https://en.wikipedia.org/wiki/Non-recurring_engineering**
I would expect that a nontrivial fraction of the design cost of a C31 chip carries over to a C32 chip though.

> Taek:
>
> CC32 also requires moving away from 16nm in order to avoid TMTO

Does that imply moving to 7nm, or does TSMC offer an intermediate node like 10nm or 12nm that would accommodate C32?
There’s also the option of combining 512 MB of SRAM with 512 MB of EDRAM on a single chip, with negligible performance penalty, although at considerable increase in design complexity. With EDRAM 's much higher density, that should still fit on 16nm.

> Taek:
>
> it is not clear to me that delaying the phase-out of Cuckatoo31 is actually damaging to anyone.

> Taek:
>
> the primary advantage of the phase-out of Cuckatoo31 to another manufacturer is that competition is eliminated.

These statements are somewhat contradictory. The phase out of C31 is damaging to Innosilicon because they judged that, based on this phase out, a multi-chip C31+C32 miner made more economic sense, and thus made investments that depend on the elimination of competing C31-only designs.

> Taek:
>
> With the block reward being spread over 3 different algorithms (Cuckaroo29, Cuckatoo31, Cuckatoo32)

The reward is split, in a predefined balance, between the primary Cuckatoo31+ PoW and the secondary cuckaroo29 PoW. It’s not split in the same sense between Cuckatoo31 and Cuckatoo32 because these are competing within the same slice. But you’re right that a manufacturer considering making a single chip C32 ASIC at 7nm is now incentivized to wait for C31 phase out, so that it needn’t compete with much lower cost single chip C31 ASICs at 16nm.

I still think that the idea of a PoW focussing on the memory IO bottleneck is a worthwhile one, but I now doubt that it’s a good idea for Cuckoo Cycle. It makes more sense for PoWs like Equihash or ethash, that have a natural affinity for DRAM.

Cuckoo Cycle is unique in having a natural affinity for SRAM and SRAM in turn has a natural affinity for use on-chip. Originally I thought the phase out schedule would lead to implementations comprising a central Cuckoo controller surrounded by many SRAM chips, with some semblance to a random-access optimized

general purpose computer, which seemed appealing to me. But it may as easily end up as a humongous die in spirit that is forced to be spread in pieces across an interposer to get around die size and yield limitations. Those extra complexities may not be as beneficial for computing in general as I had imagined. And they may well be detrimental to competition, as @rodarmar and **@Taek** have pointed out.

In hindsight I think single chip ASICs are a better fit for Cuckoo Cycle, and the planned phase outs a mistake.

I hereby propose to change our upgrade schedule at the next planned hardfork (currently set for mid july) to include only the C31 phase out, and to correspondingly restrict our immutability commitment to a lifetime of 18 months.
That is to say, any change to our primary PoW cannot take effect until at least 18 months into the future, unless agreed upon by all affected parties.

If the multi-chip miners turn out to perform better than the single-chip ones on C32, then we can reconsider the C32 phase out at a later date.

Taek 2019-03-23 21:57:11 UTC #18

> tromp:
>
> Can you reveal at what process?

The process choice is not finalized, but we are currently considering TSMC 10nm.

> tromp:
>
> I would expect that a nontrivial fraction of the design cost of a C31 chip carries over to a C32 chip though.

Two of our biggest expenses are physical design and mask creation. Physical design is process specific, which means we lose all of our work when switching from 16nm to 10nm. Over a million dollars, and several months of work. Mask creation though is the one that really kills you. At 16nm, masks are on the order of $5 million. 10nm mask sets are even more expensive than 16nm masks, and those expenses do not translate at all - you have to pay the full price at each tape-out.

The mask prices are the main barrier to entry for competing in the semiconductor industry.

All of the design and architecture work we've done (6 months of design work and logical implementation) would translate very well. What took us six months and several million dollars the first time around would take us 6 weeks and several hundred thousand the second time around. It's the masks that kills us on price and the physical layout that kills us on time.

> tromp:
>
> Does that imply moving to 7nm, or does TSMC offer an intermediate node like 10nm or 12nm that would accommodate C32?

TSMC offers 12nm, 10nm, and 7nm. 12nm is probably more fairly called "14.5nm", and likely isn't good enough for CC32, just not quite enough space for all the things we would want. 10nm is a lot better though, and also a fairly mature process with decent yields. Very likely our choice. It's price is close to that of 7nm, but it's not as exclusive or difficult to access as a small company.

A company with large-volume history (hundreds of millions in production) like Innosilicon or Bitmain would have a much easier time getting access to 7nm than a smaller company like Obelisk (tens of millions in production) would.

> tromp:
>
> These statements are somewhat contradictory. The phase out of C31 is damaging to Innosilicon because they judged that, based on this phase out, a multi-chip C31+C32 miner made more economic sense, and thus made investments that depend on the elimination of competing C31-only designs.

The way I currently understand it, which may not be correct, is that Innosilicon did not believe a single-die Cuckatoo31 chip was possible, and would have made the same miner regardless of the phase-out. I do not think that they compromised their CC31 performance in order to support CC32. But if they are not willing to reveal their architecture or decision making process, we will not know.

> tromp:
>
> The reward is split, in a predefined balance, between the primary Cuckatoo31+ PoW and the secondary cuckaroo29 PoW. It's not split in the same sense between Cuckatoo31 and Cuckatoo32 because these are competing within the same slice. But you're right that a manufacturer considering making a single chip C32 ASIC at 7nm is now incentivized to wait for C31 phase out, so that it needn't compete with much lower cost single chip C31 ASICs at 16nm.

It's definitely a much bigger issue for single-die chips than it is for multi-die chips. C31 is nice because it's viable at 16nm, and therefore has more accessibility and lower tape-out cost. Going to 10nm brings you into another tier of competitiveness and cost which will be more restrictive.

> tromp:

> I hereby propose to change our upgrade schedule at the next planned hardfork (currently set for mid july) to include only the C31 phase out, and to correspondingly restrict our immutability commitment to a lifetime of 18 months.
> That is to say, any change to our primary PoW cannot take effect until at least 18 months into the future, unless agreed upon by all affected parties.

If you stop at CC32 it would be good to just make that the permanent choice, barring some significant security issue. Any time you have a changeover, you have this game where manufacturers need to decide between chasing the closer one and chasing the further one. Changing to something like CC30 or another algorithm entirely will just re-awaken these issues.

> tromp:
>
> If the multi-chip miners turn out to perform better than the single-chip ones on C32, then we can reconsider the C32 phase out at a later date.

Based on how Innosilicon is talking, they believe that they will be superior because of yields, which is something we won't know for sure until chips are coming off of the line. Though, we've done our homework on the yields and have high confidence internally that it won't be an issue.

---

tromp 2019-03-23 18:32:51 UTC #19

> Taek:
>
> It's definitely a much bigger issue for single-die chips than it is for multi-die chips. C31 is nice because it's viable at 16nm, and therefore has more accessibility and lower tape-out cost. Going to 10nm brings you into another tier of competitiveness and cost which will be more restrictive.

In how many years (roughly) can we expect the 10nm process to be similarly mature and accessible compared to the current 16nm process?

---

asic_king 2019-03-23 19:03:47 UTC #20

> **Technology%20and%20Cost%20Trends%20at%20Advanced%20Nodes%20-%20Revised.pdf**
>
> 146.01 KB

Here is an aricle that might answer your question and gives general information about process trends.

Taek 2019-03-23 20:16:41 UTC #21

> tromp:
>
> In how many years (roughly) can we expect the 10nm process to be similarly mature and accessible compared to the current 16nm process?

10nm is meant to be an in-between node, kind of like TSMC 20nm. At some point in terms of accessibility and cost it'll be fully eclipsed by 7nm (similar to how TSMC 20nm got eclipsed by TSMC 16nm)

So the more relevant question would be how many years would it be until we can expect the 7nm process to be mature and accessible, and my guess is that it wouldn't happen until 3nm is in full production, which I think would reasonably happen in either 2023 or 2024.

---

tromp 2019-03-23 22:18:43 UTC #22

Thanks, asic_king. Slide 10 on MRAM opportunity looks interesting. According to **Wikipedia**, MRAM is denser than SRAM while not much slower. Would that be useful in fitting C32 into 16 nm? Or is this type of memory still too esoteric / immature?

Taek 2019-03-23 23:01:24 UTC #23

> tromp:
>
> Thanks, asic_king. Slide 10 on MRAM opportunity looks interesting. According to **Wikipedia** , MRAM is denser than SRAM while not much slower. Would that be useful in fitting C32 into 16 nm? Or is this type of memory still too esoteric / immature?

It looks like MRAM is currently only supported in 28nm or higher nodes, not at any 16nm node. Also, it looks like MRAM, while fast, is still several times slower than SRAM.

---

asic_king 2019-03-23 23:36:42 UTC #24

It is interesting that you picked out MRAM because there is a recent wave of announcement by Samsung and Intel about the availability of MRAM option on 22/28 nm nodes. (**https://www.eetimes.com/document.asp?doc_id=1334410**)

MRAM is being embraced by the AI community. One example is Gyrfalcon Technology Inc. who built two versions of neural accelerator chips, one with SRAM and one with MRAM using TSMC process. (**https://www.eenewsanalog.com/news/tsmc-embedded-mram-key-gyrfalcon-ai-chip**) A close friend of mine who is a memory expert worked on this project. One problem with MRAM is it has a slightly higher bit error rate due to things like stray magnetic fields. AI would be a good candidate for MRAM because CNN has a certain amount of tolerance for error. Here is a link that describes MRAM in more details for the hard core enthusiast
(**https://nepp.nasa.gov/files/24256/12_124_JPL_Heidecker_MRAM%20Technology%20Status%20jpl%20pub%2013_3%202_13%20rec%204_15_13.pdf**)

My view is that it is just a matter of time that MRAM will become available on 16nm node and beyond. If I have to guess it might take one year to two years.

---

tromp 2019-03-24 12:45:27 UTC #25

> asic_king:
>
> AI would be a good candidate for MRAM because CNN has a certain amount of tolerance for error.

So does Cuckatoo Cycle. Consider all possible bit flips during edge trimming, where N=2^32 is the number of edges in C32:

edge bit flips from alive to dead: results in loss of expected number of 42-cycles going through this edge; which is negligible at ~ 1/N.

edge bit flips from dead to alive: may result in finding fake 42-cycle, with probability 1/N, but this is trivially excluded on verification

node bit flips from from 0 to 1: might delay the trimming of a leaf edge by 2 rounds, which is totally harmless.

node bit flips from 1 to 0: might kill a connected edge, same effect as in first case of edge bit flipping to dead.

Even if the total fraction of bit flips during one graph is 1% of N (which is HUGE), the fidelity should still be over 79%.

Stuck bits (that are always 0 or always 1) are more of a problem, behaving like repeated bit flips at particular locations, but even something like 0.1% of stuck bits should be tolerable, unless the trimming is required to eliminate all but say, 0,01% of edges.

---

monkyyy 2019-03-24 17:17:47 UTC #26

> Even if the total fraction of bit flips during one graph is 1% of N (which is HUGE)

Is it though?

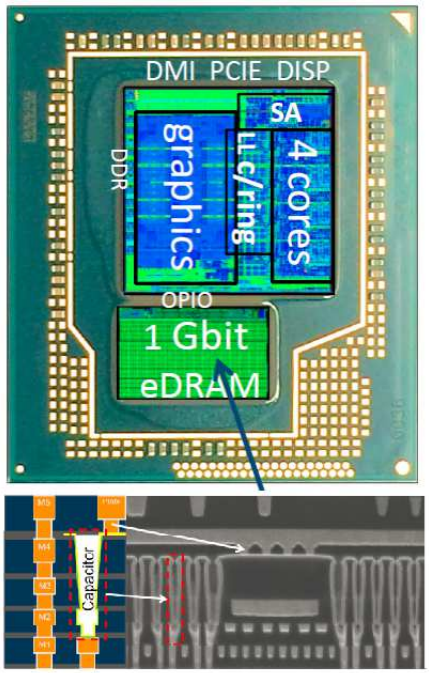
I can totally see an asic that does a few rounds of removing all nodes with connection counts under a 2^int; all super waterfalled with no control logic before passing off the ugly but much smaller graph to proper logic

---

p_semi 2019-03-25 00:42:46 UTC #27

EDRAM couldn't be made on 16nm/10nm chip directly. TSMC only provided EDRAM on 90nm~40nm.
https://www.tsmc.com/english/dedicatedFoundry/technology/edram.htm

A common way for integrating EDRAM with CPU/ASIC is MCP (multi-chip package).



---

etocs138 2019-03-25 22:28:06 UTC #29

P_semi, welcome to the board. We see that you joined about 24 hours ago. What happened to the post on the TSMC 7nm yields? I believe it was the N7 yields that you posted. Which is very out of date. If the Moderators removed it, thank-you for adhering to a code of conduct and setting expectations that this community does not want a reputation of posting confidential information. It can spoil it for every other honest manufacturer. It is also our understanding TSMC has already moved on to N7+. Attached is an article from almost 12 months ago, that mentions how the N7+ SRAM yields are performing better than the N7. This is public domain so it adheres to the governance that we abide by. **https://www.anandtech.com/show/12677/tsmc-kicks-off-volume-production-of-7nm-chips**

---

lehnberg 2019-03-27 11:32:33 UTC #30

Potentially relevant in the context of this discussion: **Slean mining - mining AT3x using low memory GPUs**

Slean mining - mining Cuckatoo3x using low memory GPUs

---

papacabeza 2019-03-27 12:49:22 UTC #31

Dear everyone —

I'm a lawyer/miner and I provided $20k advance payment to Obelisk for 10 DCR-1s in April 2018. Obelisk never delivered these devices (it's been a year!). I requested a refund within 2 weeks after paying and then I took them to small claims court. Obelisk hired Cooley (one of the most expensive firms in the states) to defend their contract of adhesion. Obelisk won.

It's been 11 months since I paid for the devices, 10.5 months since I requested a refund — not only do I have no equipment, but Obelisk won in small claims so I have no recourse.

I'm going to be taking this on as a significant consumer affairs issue. I want to expose David Vorick for his scammery and under no circumstances should a company like Obelisk be entrusted for any pre-sale. The company has stolen from me and has ripped off several others.

Here's **my Medium article** from last year. There's more public materials coming soon (e.g. a recording of what the attorneys said in court).

Obelisk is a problem. They are a bad actor. Stay away. I'm coming for them.

Taek 2019-03-27 15:37:35 UTC #32

Obelisk has shipped the majority of the DCR1 units, with the final units set to go out in the next two weeks. If you haven't received yours yet, it's because you are in the final group. We've put forward a compensation plan for all of our customers to make up for the fact that we shipped late. The compensation plan is that we will be giving all customers credit equivalent to the amount of cryptocurrency that they would have mined (assuming free electricity and no other expenses) had we shipped on time. We've done everything we can to make the situation right.

> papacabeza:

> I took them to small claims court. Obelisk hired Cooley (one of the most expensive firms in the states) to defend their contract of adhesion. Obelisk won.

You lost in small claims court because Obelisk has been fair to its customers in this situation. We're as conscious of our obligations to our customers as we are of our obligations to the communities that we build hardware for. Some unfortunate things happened with the later batch SC1s and DCR1s, and we've taken big steps to make things right. Cooley has been Obelisk's primary council since the inception of the company, they were not hired merely to handle your court case. They handle all of our legal, and they always have.

We haven't yet released the details to the public, but the major shipping hurdle that we hit was due to a big, unexpected mistake on the part of our chip design team (who is not involved with the GRN1 and will never work with Obelisk again). For the SC1 and DCR1, our chip team maintained full control of the relationship with TSMC. We were never allowed to interact with the chip manufacturer or review orders that were going out to the foundry, despite our best efforts to get involved and have more visibility into the manufacturing process.

About 6 months ago we received the final shipment of chips from TSMC. We had full confidence in the shipment because we had already received working parts from TSMC. We knew that the chips worked and that there was nothing to worry about.

Our chip design team had placed an order for the wrong parts. We got back broken chips because our chip team had put the wrong serial number down on the order form. We took steps to make it right, but there is nothing you can do to get around bad parts. We placed a new order as fast as we could, including taking direct control of the relationship with TSMC. But the fastest that we could get new parts was 6 months after receiving the bad parts.

We've already blogged about it, but following the SC1 and DCR1 we made massive improvements to our manufacturing process. We are now working with more transparent teams to design and build our rigs, we have direct relationships with all of our manufacturers and suppliers (including TSMC), and we have many more steps of diligence and double-checking that we have added to our processes to ensure high quality, low mistake, fast turn-around machines for our customers.

One major step we took this time around with our chip team is that we brought in auditors. I have already posted a link to the audit earlier in this thread, but we paid a third party auditing team to come in and verify that our current chip team (ePIC Blockchain) was using best practices and representing truthfully their capabilities. This is just one step of many that we have taken, but the audit was highly encouraging and demonstrated to us that our new process was working well.

---

papacabeza 2019-03-27 18:47:22 UTC #33

David, thank you for the response. I also received a "shipping notice" today from UPS that my DCR-1s are shipped and expected to come on April 1. The problem with that is this:

- The units don't perform as you said they would when they were pre-sold;
- You said you'd ship them by 9/30, and at April 1, that's 6 months late.

Additionally, your attorneys made representations in court about Obelisk as a crowdfunding platform, something that's completely wrong and misleading to consumers.

It's not too late to do the right thing. You can (and should) refund me and any other Obelisk consumer that's similarly situated. You should also adopt a practice for future purchasers (e.g., of GRN miners) to allow for refunds if you don't deliver what you promise.

You can't just ship these to me now and say it's ok. The units aren't as you described, they're late, and there's also the obvious point that I don't have a data center at the house that you shipped it to and I lost additional money from having reserved and paid for data center space in anticipation of September.

My issue with you and your company does not end with Small Claims. In fact, having lost in Small Claims I'm taking this to a much broader consumer complaint platform. We'll get to know each other pretty well.

For the community, here's my **court papers**. I also have recordings of the attorneys and am transcribing those for everyone's benefit.

I'm starting to speak publicly about my interest in Blockchain — one of your attorneys laughed in court when I said my interest was in governance. It's not a joke. [**here's recent vid** — pls forgive the family nature my daughter did this). In the future this issue is going to be at the center of my policy work in blockchain. The way you're mishandling and misleading consumers and beating them up with contract is a governance issue for blockchain and one that must be addressed.

You should really stop and figure this out. I am not going away and there are many other people that are similarly upset. There's a **class action**. There are these consumer **complaints**. Lots more to come. All follow the same theme of your company promising one thing and failing to deliver it.

The blockchain community needs the private sector to participate with us to develop solutions. You should be lauded for doing that but shamed for not living up to your promises. The good news is that you can change your policy.

---

papacabeza 2019-03-27 22:03:57 UTC #34

Grin community — this is the misrepresentation that worries me and why David Vorick should not be entrusted to make any statements about his future intentions with these rigs. Vorick makes a "mea culpa" admission here in the forum because he wants Grin's trust:

> Taek:

> You lost in small claims court because Obelisk has been fair to its customers in this situation. We're as conscious of our obligations to our customers as we are of our obligations to the communities that we build hardware for. [. . .] Our chip design team had placed an order for the wrong parts. We got back broken chips because our chip team had put the wrong serial number down on the order form. We took steps to make it right, but there is nothing you can do to get around bad parts.

Compare: the position above about refusing a refund and not shipping for six months. You say here you were fair. But you told us earlier, repeatedly, that you would give us refunds. By any objective standard your statements here are disingenuous and cannot be reconciled.

"The US has very strict consumer protection laws, if we are late in delivering the units, customers have the right to request a full refund, and the law will enforce their ability to receive that refund." *– David Vorick (in a forum chat), June 2017*

"You can legally demand a refund if we produce units that do not perform according to the specifications we promise in the presale, at least if you are a US citizen consumer protection laws are very real." -- *David Vorick (forum chat), November 2017*

"The FTC has very harsh rules. If you miss performance targets, you have to give everyone refunds" – David Vorick (forum chat) Jan 2018

then . . .

"It's no secret that Obelisk does not have enough money to refund all customers. We are not usre [sic] how many refund requests we will get, however we are quite confident it will be beyond our financial means to provide refunds to everyone who request. We will figure out how to provide refunds after we know the total number of refunds that must be issued and after we know how much money remains after all units have been built." *– David Vorick (forum chat) Aug 2018*

Source: screenshots of above **available here**.

In short, you stated repeatedly that we could order in confidence and that you would return the money if things didn't go well. Here you've now admitted the reason why things didn't go well, that those reasons were under your control — and you reversed the prior promise you made to provide refunds.

These are promises you made, and you, sir, are **not** a man of your word.

Taek 2019-03-28 00:58:11 UTC #35

> papacabeza:
>
> Additionally, your attorneys made representations in court about Obelisk as a crowdfunding platform, something that's completely wrong and misleading to consumers.

It was very clear during the fundraising for the SC1 and DCR1 that the money we collected was going to be used for development and NRE in addition to manufacturing. It was also clear in our terms of service that all specifications and shipping dates were estimates, and all purchasers were required to consent to the terms of service before they were allowed to place an order.

> papacabeza:
>
> You can't just ship these to me now and say it's ok. The units aren't as you described, they're late, and there's also the obvious point that I don't have a data center at the house that you shipped it to and I lost additional money from having reserved and paid for data center space in anticipation of September.

We provided a compensation plan which pays for the full revenue which the miners would have received had they been mining. You can't claim to have lost additional money for having reserved data center space because you have a compensation plan that is equal to all of the revenue that your miners would have made had that datacenter space been put to use.

> papacabeza:
>
> In short, you stated repeatedly that we could order in confidence and that you would return the money if things didn't go well. Here you've now admitted the reason why things didn't go well, that those reasons were under your control — and you reversed the prior promise you made to provide refunds.

That's not at all how the quotes you posted read. The quotes reference what my understanding was at that point of the law. Here's another quote from me from the same thread:

"Same thing would happen to us as happens to any company that runs out of money - we'd go bankrupt."

At no point did we assert that the money would be available to provide refunds.

> papacabeza:
>
> These are promises you made, and you, sir, are **not** a man of your word.

We've done our absolute best to make the situation right. Including providing full compensation for every bit of mining that you missed out on due to not having your unit. That is above and beyond what any other manufacturer in this space has done.

---

papacabeza 2019-03-28 07:43:48 UTC #36

David, I won't respond to your points here because I realize that you're just defending your current position. There's nothing that's changed in the law or in business from the time that you repeatedly reassured customers that you'd give a refund and the point now where you have refused. I'll let the community and the court of public opinion judge for themselves.

I must say that I am particularly grateful that you have engaged here because you've provided additional context, including an admission that your error caused a 6 month delay. You've shifted this to the consumers and whatever compensation plan you're referring to is not the **refund** that you had promised.

It's early days in our dispute, and even though I lost Small Claims I gained a full record, transcript, and now, additional information for this next phase. There are many more quotes and representations to share. We know what the Boston Municipal Court says, we'll now have a chance to see what consumer agencies think about your practice.

One more comment: It's ironic that we're having this discussion on a thread that includes reps from two major ASIC providers, one from the USA (Obelisk) and another from China (Innosilicon). One of the motivating reasons for my purchase of the Obelisk units was frustration with delays from Innosilicon (just of a few days) and complications with customer service. Even with those frustrations, I can definitively conclude, based on my own experience, that the customer service experience *and* the technology from Innosilicon are far and above what Obelisk is offering. In retrospect, I can hardly believe that my concern about a few days' delay in shipment from a Chinese manufacturer drove me to embrace a new, untested American company. That was a mistake that I won't make again.

[EDIT 3/28] Adding audio recording of small claims hearings (about 1 hr) **available here**.

---

bloomers 2019-03-28 04:08:15 UTC #37

Thank you for sharing this PSA. This topic is somewhat forked from the tech issue addressed in this thread, but it brings up a more practical point: track record on theory vs. execution. If GRN buyers end up like papacabeza, our community will be very unhappy.

Vorick has made beautiful, eloquent descriptions of technology and challenges, he's convincing to people. When he was selling pre-orders he reassured users about refunds during the time that he was locking down pre-orders (all those quotes and promises for refunds), but once the pre-orders were secured and the specs changed, he came out with all these new details on the changed specs, delays in shipments, and suddenly decided not to give refunds anymore. I don't even know what to say.

I was planning on ordering and won't now,. Currently, Obelisk is advertising GRN-1 to have a final design in March with delivery in October. Their projections are incredulous (if measured by past promises — all missed by months, some still unattended). The volume of litigation described could very well bring Obelisk down well before the GRNs are manufactured. If there were such a thing as a license for a presale, then Vorick's license for that should be revoked.

Anyway this information Vorick's technical proposal and his credibility for execution in new context.

Taek 2019-03-28 17:04:53 UTC #38

> bloomers:
>
> Vorick has made beautiful, eloquent descriptions of technology and challenges, he's convincing to people. When he was selling pre-orders he reassured users about refunds during the time that he was locking down pre-orders (all those quotes and promises for refunds), but once the pre-orders were secured and the specs changed, he came out with all these new details on the changed specs, delays in shipments, and suddenly decided not to give refunds anymore. I don't even know what to say.

We were quite up-front about everything. Had we gone the refund route, everyone would have received a fraction of what they paid, and Obelisk would have gone fully bankrupt. And this was apparent throughout the sale as well, it was well established that the money we were collecting would be used for research and development, and manufacturing, and it was also well established that we did not have separate money to fund the R&D and manufacturing in the event that something went wrong.

We were at that point ill prepared for the challenges associated with firmware development, with chip simulation and specification, and we had never built hardware before. That is all different this time around. You can see this blog post for more information about what we've changed:
https://medium.com/obelisk-blog/building-a-better-asic-company-8214c9c74b02

We are sorry that we did not make our original shipping date nor specification estimations. However, we provided a full compensation plan for the missed shipping date, which includes paying out IN FULL the revenue that you would have received had you received the machines on time. This is even better than receiving the machines on time, because it means that you don't even have to pay for electricity, maintenance, or other operation expenses.

And we feel that this is fully fair towards our customers. Obelisk does not and has never had the ability to give out refunds for the original sale, nor was it ever represented that we would have enough money to give out refunds, so we did the next best thing and put together a compensation plan which goes above and beyond what any other manufacturer has done to compensate for a missed shipping date.

---

papacabeza 2019-03-29 07:04:10 UTC #39

> Taek:
>
> However, we provided a full compensation plan for the missed shipping date, which includes paying out IN FULL the revenue that you would have received had you received the machines on time. This is even better than receiving the machines on time, because it means that you don't even have to pay for electricity, maintenance, or other operation expenses.

This is the equivalent of saying "give us your money now, we'll build something for you, but when we change our mind on specs, or with some third party screw up the order and deliver an inferior product, or don't deliver it all, since our contract says you take on all the risk, this situation is even better than receiving the machines on time because our low-quality, late product would never have worked right anyway."

The description you're provided, and all the coupons, upgrades, bait & switch and new technical descriptions (while teetering on bankruptcy), are characteristic of an unstable business model, to say the least. Here's my suggestion: instead of taking money from consumers, shafting them on product and telling them they should be happy for the shaft (that's how I read what you said), how about taking a pause from **any presales** at all until you catch up with your promises. This includes issuing a full refund to any Grin prepurchase customers that have coupons for the devices in this thread.

It's never too late to do the right thing.

Global Warming of 1.5°C

An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty

Headline Statements from the Summary for Policymakers*

Understanding Global Warming of 1.5°C

Human activities are estimated to have caused approximately 1.0°C of global warming above pre-industrial levels, with a likely range of 0.8°C to 1.2°C. Global warming is likely to reach 1.5°C between 2030 and 2052 if it continues to increase at the current rate. (*high confidence*)

Warming from anthropogenic emissions from the pre-industrial period to the present will persist for centuries to millennia and will continue to cause further long-term changes in the climate system, such as sea level rise, with associated impacts (*high confidence*), but these emissions alone are unlikely to cause global warming of 1.5°C (*medium confidence*).

Climate-related risks for natural and human systems are higher for global warming of 1.5°C than at present, but lower than at 2°C (*high confidence*). These risks depend on the magnitude and rate of warming, geographic location, levels of development and vulnerability, and on the choices and implementation of adaptation and mitigation options (*high confidence*).

Projected Climate Change, Potential Impacts and Associated Risks

Climate models project robust differences in regional climate characteristics between present-day and global warming of 1.5°C, and between 1.5°C and 2°C. These differences include increases in: mean temperature in most land and ocean regions (*high confidence*), hot extremes in most inhabited regions (*high confidence*), heavy precipitation in several regions (*medium confidence*), and the probability of drought and precipitation deficits in some regions (*medium confidence*).

By 2100, global mean sea level rise is projected to be around 0.1 metre lower with global warming of 1.5°C compared to 2°C (*medium confidence*). Sea level will continue to rise well beyond 2100 (*high confidence*), and the magnitude and rate of this rise depend on future emission pathways. A slower rate of sea level rise enables greater opportunities for adaptation in the human and ecological systems of small islands, low-lying coastal areas and deltas (*medium confidence*).

On land, impacts on biodiversity and ecosystems, including species loss and extinction, are projected to be lower at 1.5°C of global warming compared to 2°C. Limiting global warming to 1.5°C compared to 2°C is projected to lower the impacts on terrestrial, freshwater and coastal ecosystems and to retain more of their services to humans (*high confidence*).

Limiting global warming to 1.5°C compared to 2°C is projected to reduce increases in ocean temperature as well as associated increases in ocean acidity and decreases in ocean oxygen levels (*high confidence*). Consequently, limiting global warming to 1.5°C is projected to reduce risks to marine biodiversity, fisheries, and ecosystems, and their functions and services to humans, as illustrated by recent changes to Arctic sea ice and warm-water coral reef ecosystems (*high confidence*).

Climate-related risks to health, livelihoods, food security, water supply, human security, and economic growth are projected to increase with global warming of 1.5°C and increase further with 2°C.

Most adaptation needs will be lower for global warming of 1.5°C compared to 2°C (*high confidence*). There are a wide range of adaptation options that can reduce the risks of climate change (*high confidence*). There are limits to adaptation and adaptive capacity for some human and natural systems at global warming of 1.5°C, with associated losses (*medium confidence*). The number and availability of adaptation options vary by sector (*medium confidence*).

* Headline statements are the overarching conclusions of the approved Summary for Policymakers which, taken together, provide a concise narrative.

ipcc INTERGOVERNMENTAL PANEL ON climate change
WMO UNEP

Emission Pathways and System Transitions Consistent with 1.5°C Global Warming

In model pathways with no or limited overshoot of 1.5°C, global net anthropogenic CO_2 emissions decline by about 45% from 2010 levels by 2030 (40–60% interquartile range), reaching net zero around 2050 (2045–2055 interquartile range). For limiting global warming to below 2°C CO_2 emissions are projected to decline by about 25% by 2030 in most pathways (10–30% interquartile range) and reach net zero around 2070 (2065–2080 interquartile range). Non-CO_2 emissions in pathways that limit global warming to 1.5°C show deep reductions that are similar to those in pathways limiting warming to 2°C. (*high confidence*)

Pathways limiting global warming to 1.5°C with no or limited overshoot would require rapid and far-reaching transitions in energy, land, urban and infrastructure (including transport and buildings), and industrial systems (*high confidence*). These systems transitions are unprecedented in terms of scale, but not necessarily in terms of speed, and imply deep emissions reductions in all sectors, a wide portfolio of mitigation options and a significant upscaling of investments in those options (*medium confidence*).

All pathways that limit global warming to 1.5°C with limited or no overshoot project the use of carbon dioxide removal (CDR) on the order of 100–1000 $GtCO_2$ over the 21st century. CDR would be used to compensate for residual emissions and, in most cases, achieve net negative emissions to return global warming to 1.5°C following a peak (*high confidence*). CDR deployment of several hundreds of $GtCO_2$ is subject to multiple feasibility and sustainability constraints (*high confidence*). Significant near-term emissions reductions and measures to lower energy and land demand can limit CDR deployment to a few hundred $GtCO_2$ without reliance on bioenergy with carbon capture and storage (BECCS) (*high confidence*).

Strengthening the Global Response in the Context of Sustainable Development and Efforts to Eradicate Poverty

Estimates of the global emissions outcome of current nationally stated mitigation ambitions as submitted under the Paris Agreement would lead to global greenhouse gas emissions in 2030 of 52–58 $GtCO_2eq\ yr^{-1}$ (*medium confidence*). Pathways reflecting these ambitions would not limit global warming to 1.5°C, even if supplemented by very challenging increases in the scale and ambition of emissions reductions after 2030 (*high confidence*). Avoiding overshoot and reliance on future large-scale deployment of carbon dioxide removal (CDR) can only be achieved if global CO_2 emissions start to decline well before 2030 (*high confidence*).

The avoided climate change impacts on sustainable development, eradication of poverty and reducing inequalities would be greater if global warming were limited to 1.5°C rather than 2°C, if mitigation and adaptation synergies are maximized while trade-offs are minimized (*high confidence*).

Adaptation options specific to national contexts, if carefully selected together with enabling conditions, will have benefits for sustainable development and poverty reduction with global warming of 1.5°C, although trade-offs are possible (*high confidence*).

Mitigation options consistent with 1.5°C pathways are associated with multiple synergies and trade-offs across the Sustainable Development Goals (SDGs). While the total number of possible synergies exceeds the number of trade-offs, their net effect will depend on the pace and magnitude of changes, the composition of the mitigation portfolio and the management of the transition. (*high confidence*)

Limiting the risks from global warming of 1.5°C in the context of sustainable development and poverty eradication implies system transitions that can be enabled by an increase of adaptation and mitigation investments, policy instruments, the acceleration of technological innovation and behaviour changes (*high confidence*).

Sustainable development supports, and often enables, the fundamental societal and systems transitions and transformations that help limit global warming to 1.5°C. Such changes facilitate the pursuit of climate-resilient development pathways that achieve ambitious mitigation and adaptation in conjunction with poverty eradication and efforts to reduce inequalities (*high confidence*).

Strengthening the capacities for climate action of national and sub-national authorities, civil society, the private sector, indigenous peoples and local communities can support the implementation of ambitious actions implied by limiting global warming to 1.5°C (*high confidence*). International cooperation can provide an enabling environment for this to be achieved in all countries and for all people, in the context of sustainable development. International cooperation is a critical enabler for developing countries and vulnerable regions (*high confidence*).

Oui.Gallery Announces Cantocore Group Art Exhibition Featuring Matt Hope, Kingson Chan, Gianluca Crudele and Ann Johnson

Written on March 26th, 2019 by Rain Wong

PRESS RELEASE - FOR IMMEDIATE RELEASE

Hong Kong — March 27, 2019 - Oui.Gallery, Hong Kong Central's newest gallery 'emerging' artists with innovative art shows, announced a group show, Cantocore, opening Friday featuring Matt Hope's Zerowaste project showing results as "Entropic Bodies" cooling suits available in Hope Coins and Guided Drawings. Cantocore is the re-emergence of the hard core canto style "Cantocore". The show also features the work of Hong Kong native Kingson Chan Kin-Sing's stained glass panel diving into religious subjects and "Mente morbida" paintings from post-street-art turned contemporary artist Gianluca "Barlo" Crudele. Representing Saint Louis is artist technologist moving to the "New Bay Area" to get closer to technology innovation is Ann Johnson's LED performance cape, on-site.

"Oui.Gallery has arrived in Central Hong Kong," said Redelle Lee, Oui.Gallery Director and Co-founder. "With Tai Kwun Contemporary and the flow of new construction in central, Oui also are part of the creative solution bringing commercial art to your home & office. Cantocore is about the

Already all week artist Matt Hope and Fabricatorz have been gathering e-waste in Central Hong Kong. This e-waste has been turned into cooling suits and continues the rest of the week. The final results Matt Hope is showing as “Entropic Bodies” suits at Oui.Gallery Cantocore. Alongside Hope’s cooling suits are a set of drawings being collected internationally as plans for future Matt Hope artworks.

“Matt Hope’s new era artwork is #zerowaste,” said Jon Phillips, Oui.Gallery Co-founder and technologist. “From his guided drawings made in deep concentration considering the effects of human intervention on earth to the current zerowaste breakerspace happenings during Hong Kong Art Week, Matt Hope’s work is the leading cultural production and development ideology reshaping this century in the New Bay Area and world.”

Cantocore opens with private invite-only invites Friday, March 29 at 8 pm and runs until 10 pm. The following day the show opens to the public from 10 am until 3 pm. On Saturday, March 30, Oui.Gallery is holding Oui.Talk artist talk between Matt Hope, Gianluca “Barlo” Crudele, Kingson Chan and the public from 3 - 5 pm.

About Matt Matt Hope

Matt Hope (born 1976, London, U.K.) lives and works in Beijing, China. Hope received his M.F.A. from the University of California, San Diego in 2004. Selected recent solo exhibitions include Art Lights up Life: People’s Power Station – Lighting Up Project, Guangzhou, China, 2016; Sun Dragon Hardware, Ace Gallery, Los Angeles, 2015; and Spectrum Divide, Saamlung Gallery, Hong Kong, China, 2012. Selected recent group exhibitions include Desert Island - Epicenter Projects, Coachella Valley Art Center Indio, California, 2017; Shenzhen Biennale of Contemporary Art, Shenzhen, China, 2017; BI-City Biennale of Urbanism\Architecture, Shenzhen, China, 2016; BOOSTER: Art Sound Machine, MARTA Herford Contemporary Art Museum, Herford, Germany, 2016; The Solutions, International Design Exhibition, Chengdu Biennale, Chengdu, China, 2011; and What if, Beijing International Design Triennial, China National Museum, Beijing, China, 2011. https://matthope.org

About Kingson Chan Kin-Sing

Lives and works in Hong Kong. Kingson uses a number of approaches and uses a wide range of different media such as drawing, photography, video installation, performance, objects, also initiating projects that involve the public. His work is very idiosyncratic and is involved with a

Awarded British Council 60th Anniversary scholarship in 2009 to pursue study Master of Art, Fine Art, in Central Saint Martins College of Art & Design, United Kingdom. Chinese University of Hong Kong BA-Fine Arts (Hons) and Hong Kong Baptist University BBA-Marketing (Hons).

Worked as a member of production team for a cultural event “Detour” and in Ai Weiwei studio for art project management. He sets up his own studio for art creation in 2017.

About Gianluca “Luca” Crudele

Gianluca Crudele is an Italian painter and designer based in Hong Kong. “Luca” is also a well known street artist that goes under the name “Mr. Barlo”

About Ann Johnson

Ann Johnson is a St. Louis based artist using technology to create luminous sculptures with analog circuitry and electronics. Fascinated by the intersection of art and technology, she is constanty studying. Currently, she is practically studying electrical engineering at Washington University, in order to be able to create more emerging technology artworks and innovations. She is a member of Qi Hardware and her path towards the New Bay Area and Shenzhen, is unstoppable.

About Oui.Gallery

Oui.Gallery is an international contemporary art gallery featuring emerging artists and making innovative shows in Hong Kong and Saint Louis. Oui.Gallery Hong Kong is located at 10/fl 1009 Yu Yuet Lai Building, 43-55 Wyndham St, Central, Hong Kong https://oui.gallery

For more information

- RSVP for Private Show
- Press Release

Cantocore Graphics

- Cantocore Logotype: [SVG] [PNG] [JPG]

Matt Hope Entropic Bodies

- https://oui.gallery/entropic-bodies
- Project Posters: SVG

Sand, Death and Cryptocurrency: Life in a Decentralized Syria

Rachel Rose O'Leary

Mar 25, 2019 at 03:00 UTC • Updated Mar 25, 2019 at 03:06 UTC

I'm writing from the Democratic Federation of Northern Syria.

Known to sympathizers simply as Rojava – meaning West – the predominantly Kurdish region revolted against the Syrian regime in 2012 and achieved its de-facto autonomy as a result.

Since then, it has pioneered a new political model named democratic confederalism, which due to its stateless, decentralized nature, has a natural synergy with blockchain technologies – something that has been a point of research by technologists in the region.

That's partially why I am here.

I'm also here because, in December, U.S. President Donald Trump announced his retreat from the region, cited the impending defeat of ISIS, and denouncing Syria as the land of endless wars – of "sand and death," he called it.

The withdrawal has now effectively been reversed, but at the time, many believed Turkey, which shares a border with Northern Syria, would attack (the country has engaged a continuous offensive against the region since 2016).

The concern was that if Turkey seized control, Rojava's political system would crumble to the totalizing power of nation states. There would be no more resistance, a resistance I've come to care greatly about.

I had previously written about the potential of blockchain and cryptocurrency in Rojava. I felt that while the region lacks the basic security and resources offered by the West, it has something the West doesn't have – the opportunity for a new system of governance to be realized.

With this in mind, I spent a little over a month trying to get into the country to volunteer my skills, both in media and crypto, to a new network of technological academies being developed in the region.

On February 25, I arrived at my new home. Here, according to critics, in the process of implementing democratic confederalism, Rojava has succumbed to the pressures of the familiar, whereby the structures of capitalism and its hierarchies are being mimicked in local economies.

Erselan Serdem, the leader of Rojava's technological development program, would like to redeem this, creating structures that allow ecological, egalitarian economies to thrive – what proponents call "democratic modernity."

According to Serdem, with the right combination of philosophy and tech, this dream can be realized.

"We are talking about a new form of institution with a high level of technology, that can develop useful tools for society and achieve a good relationship to nature – that's our aspiration," Serdem said, adding:

> "Decentralized institutions can be supported by parallel, decentralized technologies."

Electricity cables in Qamishli, Northern Syria

War vets and social engineers

The academies Serdem is building will be used to train hackers in various decentralized technologies.

For instance, participants will research digital governance, cryptocurrency and blockchain solutions to fairly distribute natural resources. Serdem is still recruiting people into the academies, seeking out those with technical skills across Rojava and also training injured war veterans, starting with basic programming skills.

Currently, there are 30 war vet participants in the program.

Not only is Serdem recruiting throughout Northern Syria, but he's also enlisting what he calls "social engineers" – politically-oriented hackers and philosophers focused on reshaping technology.

Without these folks, Serdem said, "We have seen how history repeats itself. The current system will have the same destiny."

Hozan Mamo, a software developer and academy member, echoed what Serdem said, telling CoinDesk through a translator that the technological academies could solve problems that have emerged in civil society.

For instance, he continued, decentralized governance tools could help formalize decision making and keep power in check.

On the other hand, cryptocurrency could be useful as well, Mamo said, since there's no access to electronic transactions in Rojava. Instead, Rojava's inhabitants are dependant on cash that is issued by the Syria state – meaning that the region is still economically bound to the regime.

As a first step towards that, Mamo is looking into the feasibility of onboarding local merchants to accepting cryptocurrency.

Young men gather outside an internet shop, Qamishli, Northern Syria

The ethos of crypto

Still, there's a good amount of cynicism surrounding the project.

In Rojava, technology has mostly shown its face through social media, and a sudden proliferation of smartphones – mostly carrying Facebook, YouTube and Whatsapp – has had a tangible impact on the social sphere.

The obsessive use of smartphones has led a certain suspicion of technology to develop, which could have a negative impact on the adoption of blockchain and cryptocurrency technology.

In an effort to combat this, Serdem noted, he intends to use the academies to redefine technology, moving the narrative away from the corporate interest groups that have

monopolized social media, network infrastructure and even hardware in the region.

“There are different forms of technology,” he said. “[There is] technology by nation states and the companies, and then there is a resistance movement, that try to discover more ideas against the current system.”

Bitcoin and other decentralized technologies, for instance, qualify as this “resistance tech” – tools developed by oppressed people to take back power throughout history. Within the academies, Serdem would like to see some of this tech, these alternatives, built.

“We can use some of that kind of technology which is created by the resistance movement. Now we are in the beginning, but in the process we will see what forms of technology do we need to have for democratic modernity,” said Serdem.

Mamo believes that by focusing on usability and security, adoption could happen quickly, especially by younger generations. According to him, Rojava’s youth has no shortage of enthusiasm for technology, as well as a strong aptitude for it.

Prior to the revolution, he said, the Syrian regime deliberately held back the development of technology in the region, forbidding its teaching in universities and arresting people that tried to develop their skills.

But the revolution “opened the border” to technology, he said, leading the region to develop rapidly.

Poster of Abdullah Ocalan at International Woman’s Day, Qamishli, Northern Syria

Openness to technology isn't the only thing Serdem and others are trying to push. His academies also have a strong focus on philosophy, specifically the writings of Abdullah Ocalan, the incarcerated political philosopher whose writings inspired the Rojava revolution.

In his writings, Ocalan seeks to fundamentally restructure society by challenging the roots of hierarchy and domination that underpin it.

It's a philosophy that resonates strongly with the ideologies that many crypto advocates hold and even display in their interest and use of the open-source movement.

In this way, Serdem told CoinDesk, the academies will encourage much of the same.

"We are creating a commune of the technology, to solve technical problems, and at the same time to create the social engineer or the political person in the moral society," Serdem said, before concluding:

> "In Rojava, we are trying to achieve the philosophy of open source, of how to create a society informed by open source."

Note: *Due to security concerns, "Erselan Serdem" and "Hozan Mamo" are pseudonyms*

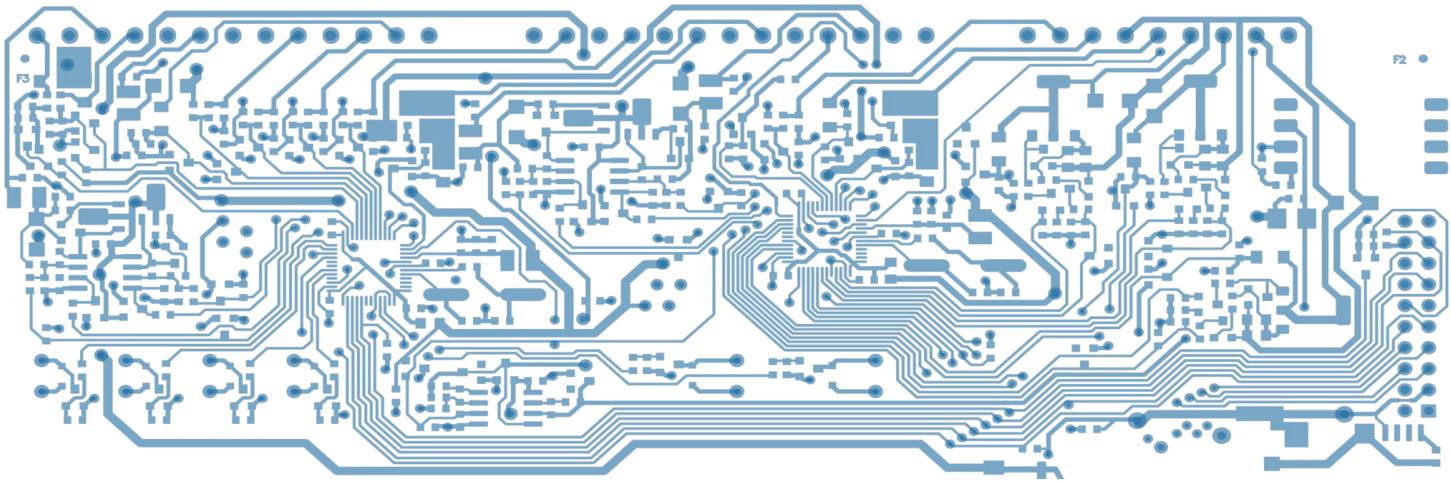
Photos by Rachel-Rose O'Leary for CoinDesk

Rojava | Erselan Serdem | crypto anarchy

November 14, 2018 POLITICS, REFORM & OPENING

From Windfalls to Pitfalls: Qualcomm's China Conundrum

Joy Dantong Ma



For many multinational businesses, the launch of China's Reform and Opening forty years ago seemed nothing short of an opportunity of the century. The potential to tap a market of nearly one billion customers, whose rising incomes invariably meant they would be able to afford more foreign products, was mind boggling. For KFC and other first movers like it, the bet on Chinese consumers paid off handsomely. Yet for other multinationals, the opportunity of the century turned out to have numerous strings attached, particularly in the realm of technology and market access.

That's because the opening up of China's market was never meant to be a free-for-all—it was a gambit always firmly anchored in the country's own economic and political priorities and interests. From Beijing's vantage point, it made perfect sense to gradually attach conditions to foreign direct investment (FDI) over time, because the underlying purpose of attracting FDI was to use it as an accelerant to catch up to advanced economies. And of course, the conditions attached to FDI changed as the Chinese economy, and its attendant priorities, evolved.

Today China is no longer a capital scarce country but a market that craves technology. This should hardly come as a surprise because it reflects dramatic changes in the Chinese economy. Of the three main inputs to economic growth—labor, capital, and technology—China has been deploying the first two to great effect. The country's tremendous growth over the past forty years has built up an enormous capital base, even as its labor market is tightening on the back of demographic shifts. So the dividends from capital and labor are diminishing, leaving China

little choice but to rely more and more on technology inputs to improve efficiency and economic productivity.

While capital and labor were comparative advantages for China, technology was, and continues to be, its notable deficiency. Recognizing its technological gap with industrialized economies, the Chinese government's attitude on FDI increasingly turned to demanding technology and intellectual property (IP) transfers to support the next stage of growth. Such a shift has been exemplified by the insistence on joint ventures, in which the foreign partner usually had a minority stake and was expected to transfer, at a minimum, know-how and expertise that could strengthen domestic industry.

Of course, the strategy of "you can continue to profit in our market if you help us gain a technological edge" has not sat well with multinationals from the United States to Europe. When boiled down, transferring critical technology is tantamount to creating your own future competitors, and no company would willingly do that. Needless to say, China's latest turn to focus on technology has become highly controversial and is at the heart of tensions between the United States and China.

In August 2017, the US Trade Representative Office launched a year-long investigation into China's practices in technology transfer and IP theft. What's more, the trade war the Trump administration launched against China is widely considered part of a strategy to get Beijing to modify its behavior on technology transfers.

These issues have seemingly come to a boil overnight. But in fact, they have percolated beneath the surface for years, if not decades. China's fixation on gaining technological leadership is hardly new and has always been a main purpose of Reform and Opening. In fact, as early as 1987, a concerned US congress had demanded extensive studies on China's technology transfer practices. What has changed are China's own capabilities and its goals for technology acquisition. Not only are Chinese companies now capable of developing their own leading technologies, they are also increasingly demanding the crown jewels of foreign technology firms.

On the flip side, these longstanding concerns didn't obscure the vast opportunities China's Reform and Opening brought to multinationals, including technology-intensive firms. For a certain set of technology companies in particular, it wasn't simply about the revenue potential of selling to a market that was a quarter of humanity. They also saw China as a unique opportunity to cement their technology standards to dominate global market share. These companies were playing a long game, with China being the focal point of the strategy.

One American technology giant that's emblematic of both the enormous windfalls and eventual pitfalls of operating in China is Qualcomm. In the early days, Qualcomm had pushed

its products, technologies, and standards into the China market, at times against the government's economic agenda. Its efforts yielded tremendous commercial success and allowed the company to gain dominance in global telecommunication standards for decades.

Qualcomm's very success, however, was also partly responsible for its own loss of momentum in the China market. It is tempting to blame Qualcomm's recent troubles—from fighting off a hostile takeover from Singapore-based Broadcom to scrapping its attempted acquisition of NXP because the Chinese authority blocked it—as simply collateral from the ongoing US-China trade war. But that would be overly simplistic and skirts the company's storied and complicated tenure in the China market.

Beijing's blocking of the NXP bid was bound to happen, irrespective of the trade war. At its core, this isn't about any single deal, but is a logical outcome of a brewing battle—between Qualcomm and China's rising technological ambitions—over the future of international telecom standards and market leadership. Indeed, Qualcomm's meteoric rise and gradual descent in China is emblematic of the country's transformation from a market that passively accepted Western companies' standards to a contender in the global technological race.

Qualcomm's 2G Windfall in China

Zhu Rongji, China's firebrand premier, wrote on the margins of a memo in March 1999, "Please have China Unicom consider adopting the CDMA standard and work with American companies." This marked a major victory for Qualcomm's seven-year push since 1992 to get a firm foothold in the China market, giving the company a significant edge in the global competition for second-generation (2G) cellular standards. To understand why that's important, a brief detour into the development of 2G standards is warranted.

CDMA vs. GSM

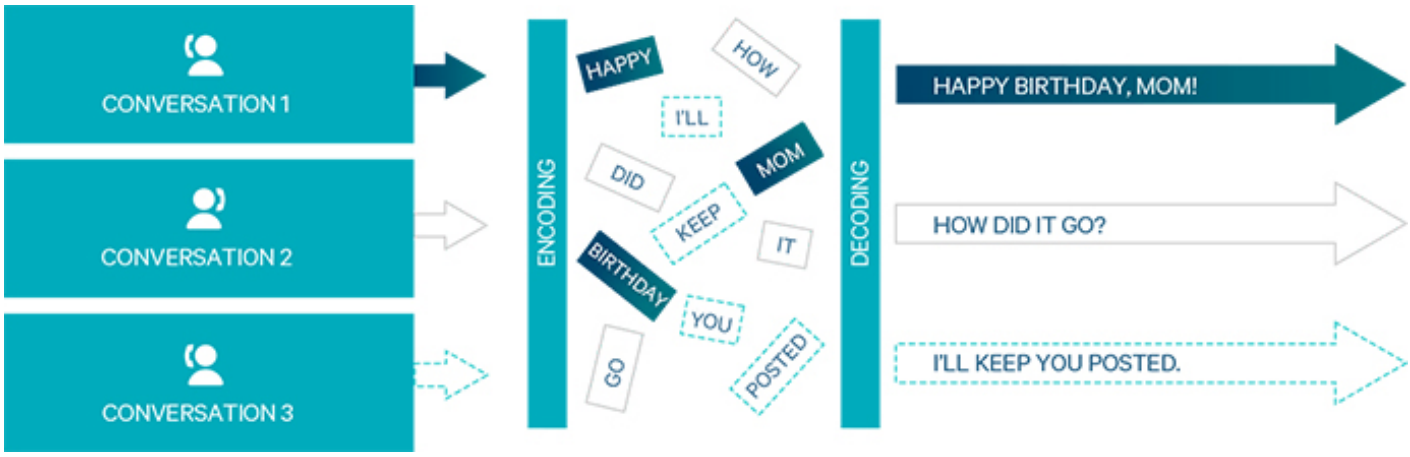
The 1980s was a period in which global telecom standards were transitioning from analog to 2G. In the analog age, each user's cell phone call was assigned a channel in which a single call could be transported. Since spectrum is a finite resource, the number of calls that can be made at the same time is limited. To put it differently, a highway is only so wide, which means only a certain number of cars can simultaneously fit across it before space runs out. Much like physical infrastructure, the constraints imposed by analog infrastructure meant that it could not accommodate huge volumes of calls and data.

The breakthrough in 2G technologies was that it allowed for multiple calls on the same channel, thereby circumventing the constraints of limited spectrum space. At the time, most of the world focused on a radio transmission technology called Time Division Multiple Access (TDMA). This technology improved spectrum usage efficiency by dividing the channel into

multiple time slots and assigning them to different calls on the same radio channel. It was a solution to maximize the usage of channel space that was often wasted or under-utilized during a phone call. It is essentially the equivalent of allowing multiple cars to run on different schedules in the same lane on the highway, except it's for multiple radio transmissions over a digital highway.

Qualcomm, founded in 1985 by UC San Diego professor Irwin Jacobs, pioneered another method that came to be known as Code Division Multiple Access (CDMA). This technology assigned each call with a code, and that call is then disassembled, transmitted, and then reassembled at the receiver's end by using the code. Because of the code identifier, the call is no longer limited to stay in one channel but can hop on other channels as needed. Therefore, multiple users can speak at the same time. To use the highway analogy again, cars no longer need to stay in a single lane and can now also use other lanes whenever there is availability, allowing for more efficient use of all radio frequencies (see Figure 1).

Figure 1. How CDMA Works



Source: Qualcomm.

CDMA might be more technologically fitting for cellular communication than TDMA, but it was too late to the party. By the time Jacobs successfully prototyped the standard in 1989, the telecom industry had already sunk millions of dollars into TDMA infrastructure and incorporated it into a Global System for Mobile Communications (GSM). Moreover, since GSM was developed through a collaborative effort by multiple European telecom companies, there was more buy-in of the standard from industry stakeholders.

According to International Telecommunication Union's report, by the end of 1998, roughly 130 million phones around the world were running on the GSM standard. Almost 90% of mobile phones in Europe, 35% in Asia, and 88% in Africa subscribed to the GSM standard. In contrast, Qualcomm's 2G standard, which came to be called cdmaOne, had only 20 million subscribers with a minimal footprint on every continent in the world except for North America.

Just as the world was becoming more connected through mobile phones, Qualcomm appeared to be on the losing end of the standards competition. This is why seizing the China market was so integral to Qualcomm's strategy. At the time, China was a market that had very low mobile penetration, and if hundreds of millions of Chinese started adopting devices with Qualcomm hardware, that would turn the tables on the 2G standards competition. In the standards race, the name of the game is market share.

Qualcomm CEO Jacobs reached the same conclusion when he visited China for the first time in 1992, the year Deng Xiaoping embarked on his "Southern Tour" to revive flagging economic reforms. Jacobs immediately set about finding ways to enter the China market. But Qualcomm's initial overtures met resistance from a Chinese government that had already set its sights on the GSM standard.

The reason was simple. Beijing preferred GSM because it didn't have to pay hefty royalties. Since GSM was a product of joint development among different countries, it had to be open source to enable collaboration. In contrast, the cdmaOne standard was developed virtually exclusively by Qualcomm, which meant it alone held numerous critical patents. Any manufacturer of CDMA-enabled mobile phones or network equipment had to pay for Qualcomm's IP.

So in 1994, when the Chinese government decided the country needed to advance from analog to 2G standards like the rest of the world, it studied both GSM and cdmaOne as potential candidates and eventually decided to go with GSM for commercial applications.

Even so, that didn't stop Qualcomm from testing the waters. The company recognized early on that it needed to enlist a key domestic constituency in China to support its efforts. It found an unconventional partner: the People's Liberation Army (PLA). Although China had selected the GSM standard for commercial use, the PLA at the time was searching for a radio transmission technology that would be secure for military communication. Because Qualcomm's CDMA technology was based on coded radio transmissions, it seemed like a good fit for what the PLA wanted.

While such a partnership would be unthinkable and forbidden today, the 1990s was a period in which the PLA had more latitude to engage in commercial activities that ranged from automobile manufacturing to pharmaceuticals and hotels.

In the Qualcomm case, the Ministry of Post and Communication (MoPC), the predecessor to the powerful Ministry of Industry and Information Technology (MIIT), ordered its local bureaus to set up a joint venture (JV) with local PLA divisions called "Great Wall." The JV was granted a civil-military dual use license to experiment with cellular networks with the cdmaOne standard on the 800Mhz spectrum in four major cities: Beijing, Shanghai, Xi'an, and Guangzhou. By 1997,

Qualcomm's 2G networks for both commercial and military applications in these cities were up and running, with the potential to expand into other Chinese cities.

But the JV collapsed almost as soon as it was formed. Just a year after the 2G network went live, Chinese President Jiang Zemin issued an order that forbid the PLA from engaging in any commercial activities. The experimental CDMA network remained in place, however, though it never grew to cover more than half a million users, all of whom were later transferred to other networks.

The timing of this episode was peculiar. Some observers even suspected that the MoPC was anticipating this outcome all along and was setting up the JV to fail just to kill the CDMA standard in its infancy. In fact, while granting the CDMA dual-use license to Great Wall, the MoPC simultaneously accelerated the approval of a national GSM license on the 900MHz band to China Telecom, one of the national champions. MoPC's true motivations will never be known, but one thing was clear enough to Qualcomm's Jacobs, who said on the record that because the MoPC fully owned the 900 MHz band used for GSM, the ministry favored GSM.

These early setbacks didn't dissuade Qualcomm from continuing its pursuit of the China market. The environment was different in 1999, when Beijing was wrapped up in intense negotiations to enter the World Trade Organization (WTO), which required winning Washington's acceptance. Qualcomm, like many companies at the time, saw an opportunity to ramp up the pressure on China to open its market. From the Chinese vantage point, liberalizing the telecom sector could go a long way toward mollifying the United States and securing its support for WTO entry. And so, Premier Zhu inked his support for Qualcomm on March 2, 1999 as detailed above and gave his promise to the US delegation, led by Commerce Secretary Bill Daley, that was soon to arrive in Beijing.

Even with Zhu's support for Qualcomm, negotiations were far from over. That's because MoPC's head at the time Wu Jichuan, a major proponent of China's decision to adopt the GSM standard in 1994, was a hard-nosed negotiator. Wu wasn't about to give in until Qualcomm met three demands: develop a new mobile phone model that can run on both CDMA and GSM networks; lower the royalty fees; and share the design of Qualcomm's CDMA chipset. In other words, Wu wanted options to abandon CDMA at will, use the technology cheaply, and own the IP so that China can make its own chipsets.

Wu's demands were of course a non-starter, and he probably knew it but wanted to play hardball anyway. These demands were viewed by the US government as China's disingenuous attitude toward WTO entry. Daley raised the issue multiple times during broader talks with Chinese leaders. While underscoring the White House's determination to push through key trade legislation that will support Beijing's WTO entry, Daley also made it very clear that "among our commercial interests, this [Qualcomm] was a very important one."

With US government support and intensive efforts, Qualcomm finally secured an in-person meeting with Premier Zhu. On October 6, 2000, Jacobs, Wu, Brent Scowcroft (former National Security Advisor who heads the consultancy The Scowcroft Group), and Yang Xiaozu (Chairman of China Unicom) all met in a conference room at Tsinghua University. Zhu was there to broker a compromise among the competing interests. He demanded all the parties present to write down a list of demands and disagreements and to "sort it out."

At this critical juncture, China's quest to enter the WTO was the priority, which meant Wu's ambitious demands had to take a back seat for the time being. The impasse was broken: Qualcomm was allowed in the China market. In March 2002, a decade after Jacob's first trip to China, China Unicom announced its commitment to deliver CDMA services to more than 350 cities.

When Success Comes Back to Bite

In the decade after Zhu lent his support to Qualcomm on the ledge of a memo, the American tech giant's revenue stream from China grew from virtually nil to $2.4 billion, more than twice the amount from its home market and accounting for one-fifth of its global revenue (see Figure 2). Yet that very success led to two unintended consequences that would eventually turn the tables on Qualcomm: 1) It prompted China to nurse a grudge against the company, particularly toward its fee structure; 2) It clarified for Beijing that it needed to raise its game in global standards setting or else accept ponying up licensing fees in perpetuity.

Figure 2. Qualcomm's Revenue Stream from China

Source: Qualcomm's annual reports (1999 – 2008); author calculations.

But Qualcomm wasn't done with profiting from 2G. It was hoping to extend its windfall in China to the 3G era, which officially commenced in January 2009 when Li Yizhong, the head of MIIT, announced China's transition from 2G to 3G. In the months leading up to this announcement, MIIT had already been laying the groundwork by consolidating the state telecom industry from six major carriers to just three: China Mobile, China Unicom, and China Telecom. This "big three" competitive landscape, much like China's state oil industry, is the one that endures today.

One of the rationales behind the industry restructuring was that each state giant would be awarded one of three competing 3G standards: WCDMA that evolved from GSM (license granted to China Unicom); CDMA2000 that evolved from Qualcomm's own cdmaOne (license granted to China Telecom, which took over China Unicom's CDMA network for $16 billion during the restructuring), and China's homegrown Time Division Synchronous Code Division Multiple Access (TD-SCDMA) (license granted to China Mobile).

The "Double Dipping" Fee Structure

However you sliced it, this 3G standards landscape would significantly benefit Qualcomm. That's because with anything that had the acronym "CDMA" in it, chances are Qualcomm owned some of the core IP since it was the original developer of the CDMA technology. According to Qualcomm's 2009 financial report, both the WCDMA and CDMA2000 technologies were derived from CDMA and "are covered by our patents." The company also claimed to hold critical patents for the TD-SCDMA standard.

With a firm grip on the core IP of the 3G era, Qualcomm made money by both licensing the IP and directly selling its own 3G-enabled chips to mobile phone vendors, who would still need to pay a royalty. Selling its own 3G chips may have been a larger contributor to the company's revenue stream, but Qualcomm's real profits were made from its licensing fees. By amassing hundreds of thousands of patents in cellular communication standards, Qualcomm's fingerprints were virtually everywhere in the telecom industry. For years, the telecom industry had a running joke that while death and taxes are two certainties in life, paying royalties to Qualcomm was another certainty in the wireless industry.

Maintaining what amounted to a patent monopoly on 3G standards enabled Qualcomm to leverage a unique and highly lucrative licensing fee structure that is still largely in place today. It basically works like this: mobile phone manufacturers license Qualcomm's technologies and pay the company royalties that are as much as 5% of the final sale price of the phone. This means the royalties increase with the phone price, even if Qualcomm's technology inside the phone remains unchanged. In contrast, other telecom companies, such as Ericsson and Nokia, charge a flat fee for the specific technologies that licensees actually use.

To illustrate, if a basic mobile phone costs $400, then Qualcomm gets 5% of that in royalties, or $20 per phone. If the manufacturer decides to add a high-resolution camera, a bigger screen, or a sleeker case to soup up the phone, the price doubles to $800. Now the manufacturer has to pay $40 in royalties to Qualcomm even though the technologies licensed have not changed.

On top of paying royalties, as a 3G mobile phone manufacturer, you would either need to make your own 3G chips or buy from other chip makers. More likely than not, manufacturers end up buying chips from Qualcomm, so they have to pay the company again. This fee structure came to be known as "double dipping" and, needless to say, has irked many manufacturers.

Figure 3. Evolution of China's Mobile Standards Adoption (1994 – 2014)

Source: Jefferies Equity Research.

This pricing strategy was also applied to the China market. But initially, it was foreign manufacturers that felt the brunt of Qualcomm's fee structure. That's because when the company entered China in 1999, the country was still a technological backwater incapable of producing quality mobile phones. Major carriers like China Unicom had to sign contracts with foreign manufacturers, such as Nokia and Ericsson, to import the phones. (Those manufacturers also relied on Qualcomm chips and IP, so were paying the company.)

It would take about another decade for Chinese manufacturers to acquire the capability to produce low end to "good enough" 3G mobile phones (see Figure 3). And that was when Chinese producers started to directly feel the pinch of Qualcomm's double dipping strategy. Since Qualcomm held patents for all three 3G standards in China, manufacturers had little choice but to pay licensing fees. In addition, few Chinese manufacturers had the ability to make their own 3G chips, so they had to rely on foreign imports, including Qualcomm's.

Figure 4. Mobile Phone's Growth in China, 2000-2014 (in millions)

Source: Statista.

Demand for 3G mobile phones skyrocketed in China after 2009, and has grown 15 times in the 15 years since Qualcomm's official entry into the China market. This led to another windfall for Qualcomm. By 2010, Qualcomm's revenue from China reached $3 billion, surpassing that of South Korea. Just four years later, the company's China market revenuefor the first time exceeded combined revenue from the rest of the world, including the United States.

Yet as Qualcomm's profit margins widened, Chinese mobile phone makers' margins were being squeezed. Domestic original equipment manufacturers (OEMs) already had to keep prices low because of fierce competition that often resulted in price wars. Meanwhile, as Chinese manufacturers started to make more expensive phones with better displays and high-end cameras, they discovered that Qualcomm's licensing fee kept on increasing, even though they were using the same IP.

Unsurprisingly, this did not sit well with Chinese OEMs nor with the Chinese government. What's more, it wasn't exclusive to China. The fee structure irritated many global mobile phone makers, especially as they were under the pressure of product cycles to constantly deliver new and more expensive features such as larger and better displays and fancier cameras. The price of their products went up, and like Chinese manufacturers, they also suddenly found themselves paying Qualcomm double or even triple the royalties for licensing essentially the same technologies.

In subsequent years, Qualcomm's double dipping strategy would become a major source of conflict, not only in the China market but also in the broader telecom industry. Qualcomm has

long argued since the 1990s that no matter what went in the phones, it was their technology that enabled them all. But this argument gained less traction in the 2000s. In the eyes of Qualcomm's customers, the company's technology was contributing less value to mobile phones yet the licensing fees kept on rising.

By 2015, Qualcomm was embroiled in controversies or being fined by regulators in Taiwan ($773 million), South Korea ($1.23 billion), and Europe ($853 million). Even Apple jumped on the bandwagon and fought multi-year legal battles with Qualcomm over this issue, arguing that it was engaged in "illegal practices." By mid-2018, Apple announced that it would manufacture its own chips for the iPhone, completely moving away from Qualcomm's chips.

It didn't help matters that Qualcomm, at times, may have rubbed salt in the wound. In its 2014 annual financial report, the company noted, "particularly in China, certain licensees have disputed or underreported royalties owed to us under their license agreements with us, and certain companies have yet to enter into or delayed entering into license agreements with us for their use of our intellectual property, and such licensees and/or companies may continue to do so in the future." While some Chinese manufacturers certainly found ways to circumvent royalty payments, Qualcomm still had all the chips in its corner.

Complaints in China grew louder and became harder to ignore for Chinese regulators. So they sprang into action. In November 2013, months before the issuance of 4G licenses, the National Development and Reform Commission (NDRC) initiated an investigation into whether Qualcomm's licensing practices violated China's Anti-Monopoly Law, which took effect in 2007.

As the investigation proceeded, Qualcomm was preparing for a fine of 1% to 10% of its previous year's revenue and other remedies.

After the 14-month investigation concluded, Chinese regulators slapped a $975 million fine, equivalent to 3.7% of the company's 2014 revenue, the largest fine ever in China for monopolistic practices. On top of the fine, Qualcomm agreed to lower its royalty rates on 3G devices to 5% and 3.5% for 4G devices, using a royalty base of 65% of the final sale price as opposed to 100%. So the company effectively lowered its royalty rates to 3.3% and 2.3%, respectively, on 3G and 4G devices, lower than in other foreign markets including India. In response, Qualcomm's annual dividend saw a $0.60 cents per share reduction.

China's first failed attempt on standards setting

What resulted was beyond Qualcomm's expectations, but such an outcome should not have been a surprise. The writing was already on the wall four years before the investigation, when MIIT in 2009 unveiled its grand designs on promoting 3G standards.

The lack of domestically developed IP in mobile standards has clearly frustrated Chinese

regulators to no end. They learned first-hand from Qualcomm how having a near monopoly on core technology patents is directly linked to market position and profits. From Beijing's perspective, why should China passively accept standards when it had the market size to come up with competing standards to Qualcomm's?

MIIT's answer to that question was to order the China Academy of Telecommunications Technology (CATT) to collaborate with Germany's Siemens to develop a new 3G standard that would come to be known as TD-SCDMA. In 2001, backed by all three Chinese carriers, TD-SCDMA was approved to join the global 3G standards governing body, the 3G Partnership Project (3GPP). However, it was China Mobile that was granted the TD-SCDMA license. Of the three state carriers, China Mobile was MIIT's favorite and had dominated the 2G market (see Figure 5). But being the favorite also meant that China Mobile had the unenviable task of ensuring that the indigenous but commercially unproven 3G standard becomes a success.

Figure 5. China Mobile Lost Market Share from 2G to 3G

Source: Caixin.

Except the opposite happened. TD-SCDMA turned out to be far less developed than the prevailing 3G standards WCDMA or CDMA2000, both of which had proven to be commercially viable for years. No carrier outside of China ever used TD-SCDMA and even Chinese carriers, including China Mobile itself, sought to disassociate themselves with it.

But MIIT didn't want to give up hope and ordered China Mobile to develop an entire 3G network based on TD-SCDMA. This was ostensibly a last-ditch effort to bolster the homegrown technology, but instead, China Mobile lost 10% of its market share over the four years it was being forced to support the weaker standard. "Other people had a head start and were running ahead of you on the main road. You can't just give up, turn around and dig a separate lane," as a Chinese telecom industry expert commented. "China's TD-SCDMA led Chinese telecom companies to a detour from the mainstream."

China's effort to introduce a domestic 3G standard ended in failure, but its appetite for reducing dependence on foreign core technologies remained as strong as ever. The Chinese government had learned a hard lesson, but did not exactly hide its ambition to have another go at setting standards. It bided its time and largely went with the flow as the world moved to 4G standards.

But even then, two Chinese companies, Huawei and ZTE, had started to make some waves. According to Jefferies Equity Analysis, ZTE held 6% and Huawei 1% of all patents in 4G standards. That would change quickly as Huawei matured and trained its sights directly on Qualcomm. If Huawei's effort to lead in global 5G standards succeeds, it will prove disruptive

for Qualcomm's business in China.

The 5G Race Is On

On July 26, 2018, China's telecom giant Huaweipresented a medal to Dr. Erdal Arikan, a Turkish expert in polar coding theory. The medal was designed and crafted by Monnaie de Paris with a Baccarat crystal. As extravagant as the medal was, its value was negligible compared to the royalties Huawei was about to collect by developing its own IP based on Arikan's theory.

Huawei had been quietly pouring 15% of its annual revenue, ormore than $61 billion, over the past decade to develop technologies that have the potential to become global 5G telecom standards. One such technology is based on Arikan's polar coding theory. To understand why that's important, a brief explanation of 5G standard development is needed.

Just like in the 2G and 3G eras, delegates from the world's major telecom operators, networks, terminals and chipset vendors, and internet companies regularly met at 3GPP, the international governing body of telecom standards, to pitch technical solutions to various 5G challenges. One of the main problems that needed to be solved was reducing data transmission errors as the volume of data grew exponentially. More errors have crept into large volumes of data due to noise, interference, and fading.

A method called channel coding—which is basically repeating a piece of data to reduce errors—was developed to overcome the problem. To oversimplify, channel coding according to MIT basically works like this: if you were trying to transmit a message with only three bits, like 001, you could send it three times "001001001". If an error crept in, and 001011001 was received instead, you could be reasonably sure that the correct string was 001.

Arikan's polar coding theory is one such channel coding method that could be applied to improve data accuracy. So Huawei decided to back polar coding and invested billions into its commercialization. Within the course of eight years, this relatively new theory had become a viable solution in practice, surprising even Arikan.

The direct competitor to polar coding technology is, no surprise, Qualcomm's low-density parity check (LDPC) technology. Compared to polar coding, LDPC has a much longer track record of commercial viability. The theory of LDPC was first introduced in 1963, 45 years earlier than polar coding. In subsequent decades, Qualcomm pioneered LDPC's commercial application and developed critical patents. By the time polar coding was introduced in 2008, applications of LDPC had already been deployed in the real world.

The contest over whether LDPC or polar coding would become the global 5G standard for

channel coding erupted on November 14, 2016 in Nevada, where 3GPP held meetings to vote on accepting a channel coding solution.

Debate was intense at the meeting, with companies picking sides. Western companies, led by Qualcomm, largely fell in line behind LDPC while numerous Asian manufacturers favored Huawei-backed polar coding. In an interview to the *Wall Street Journal*, an expert who was at the meeting recalled, "the Chinese decided this was important. This was one of the biggest political battles we've ever seen."

Eventually, the two sides reached a compromise: both polar coding and LDPC would be adopted as part of the channel coding standard. This was a victory for Huawei as it gained a critical patent in the 5G global standard.

More such battles have been fought, and Chinese telecom companies have made considerable strides in establishing a foothold in 5G standards. According to technology research firm LexInnova, Huawei and ZTE today hold about 10% of critical 5G patents, compared to 15% for Qualcomm (see Figure 6).

Figure 6. Shares of Critical 5G Patents by Company

Source: LexInnova.

To some extent, the global standards race is a zero sum game in that only one technology will be ultimately suited to addressing one critical technical challenge. And the incentives are such that, like Qualcomm, each company is aiming for market dominance, not just market share. Therefore, the very nature of this competition means that Qualcomm increasingly finds its own dominance being chipped away by the emergence of formidable rivals—some of which are Chinese manufacturers who were once Qualcomm's customers but are now using what they learned to compete with it.

As if fending off new competitors isn't tough enough, Qualcomm also had to face pressure from the Chinese government to transfer its knowledge to Chinese companies. Although the government has long dangled the carrot of market access to get foreign companies to share certain technologies, the difference today is that the relative leverage has shifted.

Qualcomm still carries a lot of weight, but it is no longer the only player in town. Beijing has choices now, and if Qualcomm isn't willing to play ball, the market share will go to a competing European firm or better yet, a rising Chinese company. This makes the trade-off challenging for Qualcomm: lose market share to Western tech giants today or lose market share to Chinese upstarts tomorrow.

Competition is also taking place in the area of hardware, namely advanced chipsets that are

capable of supporting 5G data processing speeds. In fact, Moore's Law's famous prediction of computing speed doubling every two years was predicated on fitting ever more microscopic transistors on a chip. That's because computing power is positively correlated with the number of transistors that can be piled onto a chip. The current generation of advanced mobile chipsets use 14-nanometer transistors.

But few Chinese companies have the ability to manufacture such chips. So eight months after NDRC slapped the fine on Qualcomm, the company agreed to form a JV with Huawei and China's Semiconductor Manufacturing International Corp. (SMIC) to develop 14nm chips. This move was widely interpreted as a way to patch up relationships with the Chinese government, with little upside for Qualcomm otherwise.

These chips, however, quickly became obsolete. A true 5G network would enable users to download a full movie in 15 seconds, compared to 6 minutes in 4G. This means that the data processing capacity required for a 5G chip is much higher than that of 4G. The chips need to fit even more transistors, which means their size had to be reduced to at least 10nm.

Even global giants like Intel struggle with developing 10nm chips, let alone Chinese semiconductor fabricators. But Qualcomm in 2017 again decided to help SMIC's subsidiary SJSemi to start the qualification of wafer bumping, a technique in chip manufacturing, to produce 10nm chips. This made SJSemi the first ever chip manufacturer in mainland China to enter the 10nm arena. Qualcomm at the time said that such collaboration "shows our commitment to support the upgrade of China's local IC manufacturing industry and to better serve our Chinese customers."

Currently, Samsung, Huawei, and Qualcomm are leading the pack in developing 5G chipsets. Huawei started its R&D efforts into 7nm processors in 2015 and has invested over $300 million in developing a prototype. On August 15, 2018, Samsung launchedthe first 10nm 5G chipset that's fully compliant with 3GPP standards. Huawei immediately responded by announcing that it would launch its own 7nm 5G chipset Kirin 980 on August 31. Qualcomm, however, quietly launched its own 7nm Snapdragon chip ahead of Huawei on August 22.

New Battles on the Horizon

Qualcomm brought CDMA to China in the early days of Reform and Opening, even as the Chinese government had already decided to go in a different direction. But the American tech giant wasn't defeated, using various leverage points like negotiations over China's WTO entry to get into a market that was crucial to its long-term strategy.

Qualcomm's persistence paid off handsomely: Beyond the billions of profits, without the China market, it would not have been able to dominate two generations of telecom standards. By

having China adopt the 2G and 3G CDMA standards, Qualcomm's market position in global telecom standards was cemented.

The American company's success, however, left lasting impressions on the Chinese government and companies about the importance of leading in global telecom standards primarily through the development of indigenous IP. Qualcomm also didn't help itself by alienating Chinese manufacturers and the telecom industry writ large with its lucrative fee structure that many viewed as unfair. In fact, China's effort to set its own 3G standard with TD-SCDMA, albeit one that ended in failure, was a response to widespread domestic frustration over not having any influence in global standards.

Figure 7. Huawei vs. Qualcomm Patent Wall

Qualcomm image source.

After a stellar run of 15 years in the China market, Qualcomm's rise may be interrupted. As China's telecom firms and mobile phone manufacturers have matured, and having absorbed the previous lessons of failure, they appear ready to challenge the industry leaders. For Chinese companies, Qualcomm's experience taught them that if you win the patents race, you win the standards war. This is reflected in a Chinese company like Huawei, which has taken chapters from the Qualcomm playbook and has been obsessively filing patents (see Figure 7).

Qualcomm's future prospect is arguably more uncertain than it has been in decades. It is stuck in a paradoxical position: the market that today contributes more than 60% of Qualcomm's global revenue also happens to be the market that is most likely to challenge its dominant position. To make matters worse, this is coming at a crucial period of transition to the next-gen 5G standards in which no clear winner has been crowned.

This race is set to intensify, and so will the politics around it because technology is the main source of current US-China tensions. But ultimately, this is a competition between multinational companies—they are both proxies of respective national ambitions and potential collateral in

the escalating conflict between their home countries. For Qualcomm, the battles it has fought and won so far in the China market appear to pale in comparison to the new battles on the horizon.

READ NEXT:

Imagining Xi Jinping's "State of the (Chinese) Union" Address

Damien Ma

China Didn't Invent Asian Connectivity

Evan A. Feigenbaum

English

996.ICU

What is "996"?

A "996" work schedule refers to an unofficial work schedule (9 am - 9 pm, 6 days a week) that has been gaining more popularity. Serving a company that encourages the "996" work schedule usually means work for at least 60 hours a week.

Labor Law of the People's Republic of China

Labor Law of the People's Republic of China Article 41 says:

> The employer can prolong work hours due to needs of production or businesses after consultation with its trade union and laborers. The work hours to be prolonged, in general, shall be no longer than one hour a day, or no more than three hours a day if such prolonging is called for due to special reasons and under the condition that the physical health of laborers is guaranteed. The work time to be prolonged shall not exceed, however, 36 hours a month.

Gaining more popularity and publicity

In early 2019, a Chinese E-commerce company called *Youzan* announced the adoption of the "996" work schedule in the future, at the company's Chinese New Year Party. The CEO of Youzan responded: "This would definitely be a good decision when we look back in a few years time.".

In Mid-March 2019, it was reported that Jingdong started adopting "996" or "995" work schedule in some of the business units. Jingdong PR posted on their maimai(脉脉 , a Chinese real-name business social network platform) account: "(Our culture is) to devote ourselves with all our hearts (to achieve the business objectives)".

Although it is gaining more publicity recently, this work scheudle is a commonly known "secret" practiced in a lot of companies in China.

Compensation and benefits

According to the Labor Law, employees who follow the "996" work schedule deserve to be paid 2.275 times of their base salary. Unfortunately, people who work under "996" rarely get paid that much.

Where does the name of the repo `996.ICU` come from?

If you consistently follow the "996" work schedule, you run the risk of getting yourself into **I**ntensive **C**are **U**nit.

Developers' lives matter.

freeCodeCamp

How to hack a car—a quick crash-course

Kenny
Kuchera

Jun 21, 2017 · 11 min read

Spoofed tachometer, the engine isn't running.

The goal of this article is to get you started hacking cars—fast, cheap, and easy. In order to do this, we'll spoof the RPM gauge as an example.

The following is by no means an exhaustive tutorial. It instead aims to provide just

enough information to get you up and running. If you want to dig deeper you can checkout the must-reads at the end.

If you decide to carry out this tutorial in real life, you'll need a Linux computer (or a virtual Linux machine), and a CAN-to-USB device (which we'll look into later).

A car is a network

A car consists of multiple computers to control the engine, transmission, windows, locks, lights, etc. These computers are called electronic control units (ECU) and communicate with each other over a network.

For example, when you press the button on your steering wheel to increase the volume of the radio, the steering wheel ECU sends a command to increase volume onto the network, the radio ECU then sees this command and acts accordingly.

There are multiple networks in a car, generally at least two:

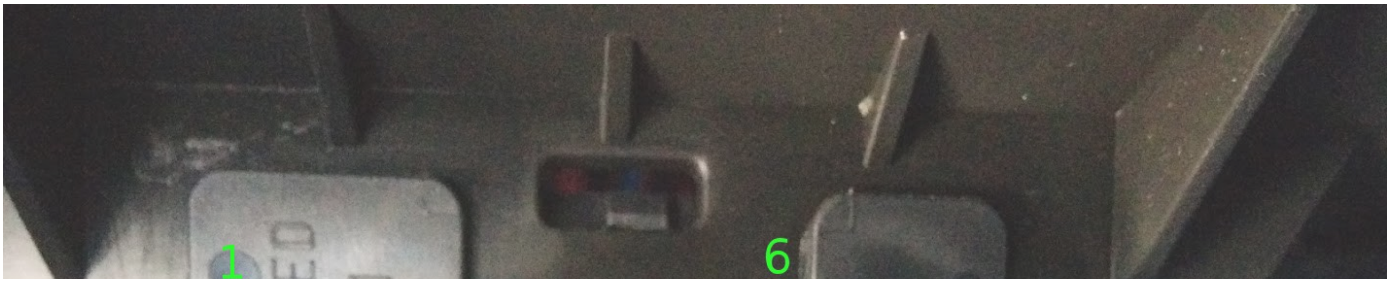
- One for critical data such as engine and powertrain messages
- And one for less critical data such as radio and door locks

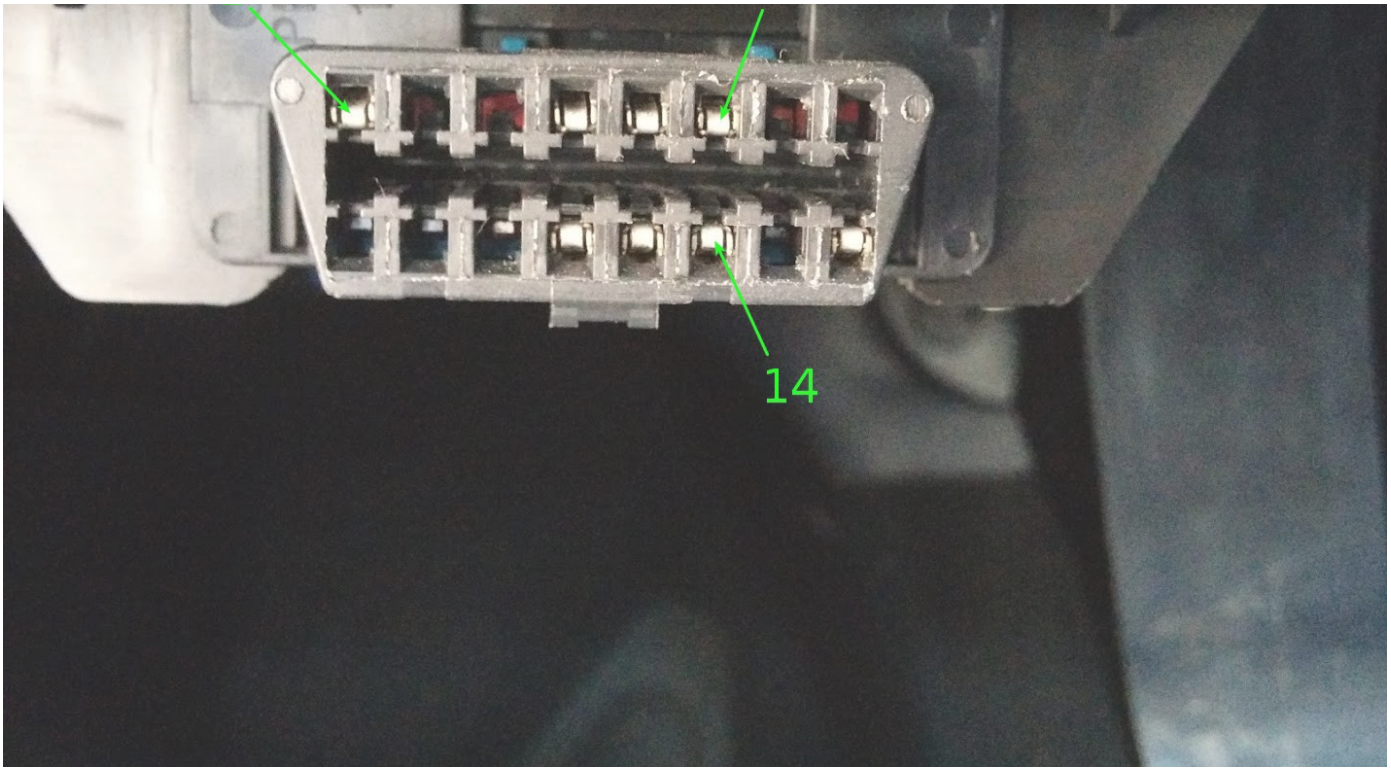
The critical network uses a fast and reliable protocol whereas the non-critical network uses a slower, less reliable but cheaper protocol. The number of networks as well as which ECUs are networked together depends on the car make, model and year. An ECU could also be connected to multiple networks.

Connecting to a network

Some networks can be accessed via the OBD-II port. OBD-II is mandatory on all cars and light trucks built in the US after 1996 and Europe after 2004.

The connector is in arms reach of the driver's seat. You might need to lift off some plastic cover but it is always accessible without tools.





OBD-II connector.

The OBD-II standard allows for five signaling protocols. It's up to the manufacturer to decide which one to use. CAN is the most popular one and is what we will discuss. It is accessible via pins 6 and 14 of the OBD-II connector. If your car has a CAN bus, you will see metal leads on the pins as in the image above.

The CAN bus is a reliable, high speed bus that is used to send critical data. Unfortunately the data packets on the bus are not standardized so you will need to reverse them to know what they mean. The OBD-II standard also leaves room for vendor specific pins that can be used for vendor specific protocols. This makes it easier for the dealer to diagnose problems.

On my car (GM), I have a standard CAN bus on pins 6 and 14, and a vendor specific single wire CAN bus on pin 1. The standard CAN bus is a reliable, high speed (500 kbps) protocol also referred to as high speed CAN (HS-CAN). It is used for critical data. The single wire CAN bus (SW-CAN) or GMLAN is slower (33.3 kbps) and less reliable but cheaper since it only uses one wire. This bus is used for non-critical data.

If you see a vendor specific pin and don't know which protocol is being used, Google "<make> OBD pinout". There is also low speed CAN (LS-CAN) and medium speed CAN (MS-CAN). MS-CAN is generally on pins 3 & 11, running at 125 kbps on Ford and Volvo cars.

Tools

You will need both a device that's capable of interpreting CAN data as well as software to analyze the data

Hardware

In order to receive and transmit CAN packets, you need a device that is capable of this. You will often come across ELM327 based devices. While these have their use, they are terrible for hacking. They are way too slow to monitor the CAN bus.

There are also high-end devices like Kvaser, Peak or EMS Wünsche. These will get the job done but are overkill and pretty expensive.

Some high-end devices also require you to purchase software along with it. The USB2CAN is a native CAN interface for Linux that offers great value for money.

You could also use Cantact or CANUSB. However these aren't native CAN devices in Linux and use an ASCII based protocol. This means that they are slightly more complicated to set up and have lesser performance. On the other hand, they are well supported across multiple operating systems.

I use CANalyze which I've designed for my needs. It is similar to USB2CAN in that it's an affordable native CAN interface but it uses a newer micro controller, is open source and can be built using open source tools. The rest of this tutorial assumes you are using a native CAN interface.

Software

To communicate with the device you need to install the can-utils package on your Linux machine. You can do this via by typing the following into the Linux prompt:

```
sudo apt-get install can-utils
```

Can-utils makes it extremely easy to send, receive and analyze CAN packets. These are the commands that we will use.

- **cansniffer** display only the packets that are changing

- **candump** dump all received packets
- **cansend** send a packet

Linux has CAN support built in to the kernel via SocketCAN. This makes it easy to write your own additional programs. You can interact with the CAN bus in the same way you would interact with any other network i.e. via sockets.

CAN bus

Before you start reversing, you should have some understanding of how the CAN bus works. It consists of 2 wires and uses differential signaling. Since it's a bus, multiple devices can be connected to these two wires. When a CAN frame is sent on the bus, it is received by all ECUs but is only processed if it's useful for the ECU. If multiple CAN frames are sent at the same time, the one with the highest priority wins. A CAN frame has 3 parts that are relevant to us.

- **arbitration identifier** The identifier of a message. An ECU uses it to decide to process or ignore the received frame. It also represents the message's priority. A lower number has a higher priority. So for example, if you'd be an engineer designing the network, you would give the frame for the deployment of airbags a very high priority or a low arbitration ID. On the other hand you'd give a lower priority or higher arbitration ID to data meant for the door locks.
- **data length code (DLC)** Indicates the length of the data field in bytes. A CAN frame can have at most 8 bytes of data.
- **data field** Contains up to 8 bytes of data.

Reversing the CAN bus

The general approach to reversing the CAN bus is to generate the behavior you want to mimic and find the message that causes that behavior. For example, lets say the lane keeping assist system (LKAS) on your car is crap and you've made your own.

In order for it to control the steering, you need to know what messages to send. The way to figure this out is to turn on the original LKAS, monitor the CAN bus and identify the packets responsible for turning the steering wheel. Once you have identified these packets, you can have your own LKAS send these packets onto the

CAN bus to control the steering wheel.

In our case, we want to spoof the tachometer so we need to change the RPM by stepping on the gas with the car on and in neutral and then try to find the packet responsible for changing the RPM.

Setup

Plug the CAN device into the car's OBD-II port and the computer's USB port. Bring up the CAN interface by running the following in your Linux prompt:

```
sudo ip link set can0 up type can bitrate 500000
```

which brings up the `can0` interface (always `can0` if you only have one device connected) at a bit rate of 500 kbps, which is standard.

Identify

When the car is off, the ECUs are usually sleeping so you need to turn on the car or put it in accessory mode. You can look at raw CAN data by running this in your Linux prompt:

```
candump can0
```

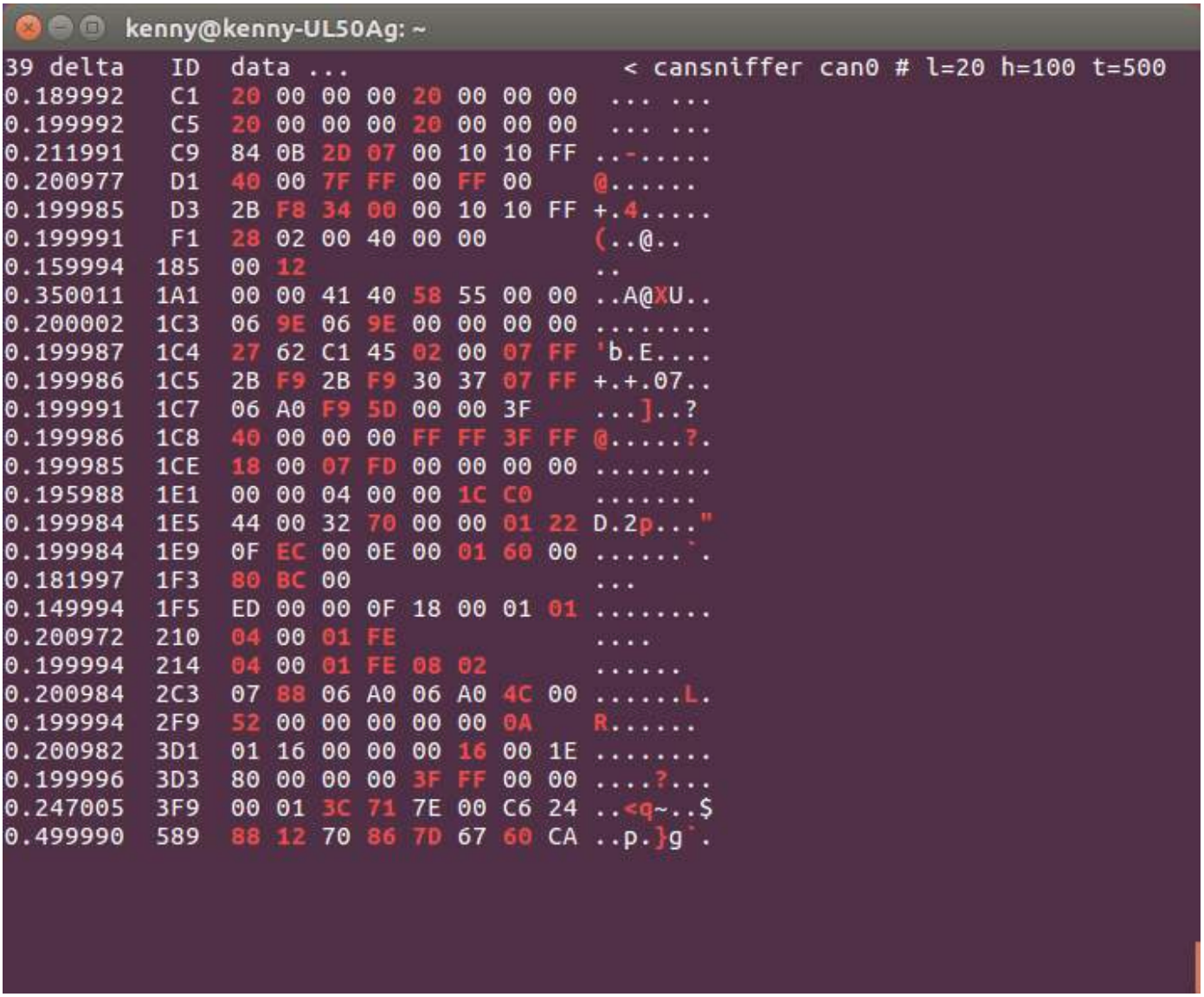
This prints CAN data to the screen as soon as it is received. This however is very unorganized and it is very difficult to see what packets correspond to a certain event. You can press ctrl+c to stop the program. To make the data more readable we use cansniffer which groups the packets by arbitration ID and only shows the packets that are changing. In order to start it run the command in your Linux prompt:

```
cansniffer -c can0
```

where `-c` colorizes the changing bytes and `can0` is the interface to sniff. It takes a few seconds to remove the constant packets.

You should see something similar to the image below, though the numbers will

You should see something similar to the image below, though the numbers will probably be completely different.



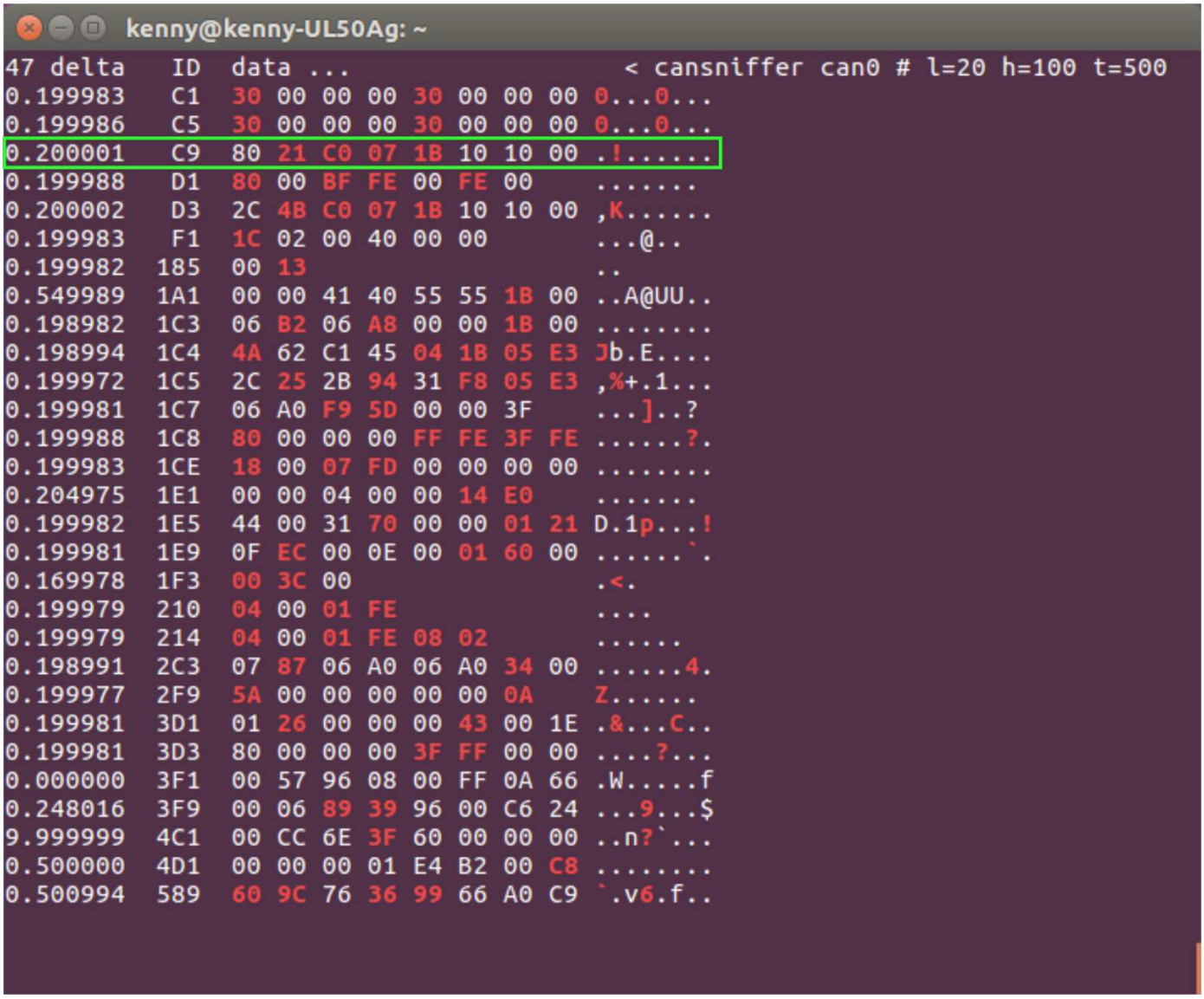
Cansniffer with engine idle.

The first column (delta) shows the rate in seconds at which the packets with that arbitration ID are being received. The second column (ID) contains the arbitration ID. The remaining alphanumeric columns (data …) contain the data bytes. If the data has an ASCII representation, it can be seen to the right, otherwise it's a dot.

When you step on the throttle with the engine running in order to increase RPM, there might be new CAN messages appearing on the screen and/or existing ones changing.

We need to find a CAN message where the changing bytes correlate to the change in RPM. We can probably expect that the value will increase/decrease as the RPM increases/decreases.

The first CAN frame in cansniffer that seems to vary with RPM is the frame with arbitration id `C9`. There are probably multiple potential packets that vary with RPM, this is just the first one.



Detected packet correlating to RPM.

There are 4 bytes that are changing (colored red) in this message but not all of these necessarily indicate the RPM. Variations in the third byte `07` don't seem to correlate to varying RPM. The last byte `1B` does.

However, as soon as we take our foot off the throttle, it goes to `00`. This would indicate that it represents the throttle position and not the RPM.

Finally there are the two bytes `21 C0` that do seem to correspond to a change in RPM. More so, it varies as a 16 byte integer i.e. when the second byte `C0` overflows, the first byte `21` gets increased by one. Also it seems that `21` corresponds to roughly 2000 RPM. This is good to note when you will replay the

message.

Replay

Once you have a candidate, send it onto the CAN bus with the following command in your Linux prompt:

```
cansend can0 0C9#8021C0071B101000
```

where the frame has the format `<arb_id>#{data}` and must be substituted with your own CAN message.

Your car can be running or in accessory mode for this. Be sure to use a packet that you obtained when the engine was non-idle or else you won't see anything change when replaying it while your engine is idle.

If you just send the packet once, you will probably not see anything change on the instrument cluster. This is because the original message is still being sent continuously on the bus at 0.2 second intervals by the ECU so your message will just be ignored.

Recall that the rate is given in the first column of cansniffer. There are two ways to get around this aside from disconnecting the ECU that's generating these messages. One option is to send the packets at a much higher frequency than the ones currently being sent. You can do this by running the following in your Linux prompt:

```
while true; do cansend can0 0C9#8021C0071B101000; sleep 0.002; done
```

and substituting the CAN message with the one you've identified. Press ctrl+c to stop.

Another option is to monitor the bus, and every time you detect the packet that you want to spoof, send your own packet out immediately after. This can be done by running in your Linux prompt:

```
```

```
candump can0 | grep " 0C9 " | while read line; do cansend can0 0C9#8021C0071B101000; done
```

where you need to substitute the CAN message and `0C9` with CAN message you identified and it's arbitration id respectively. You can experiment with both approaches to see which one works better.

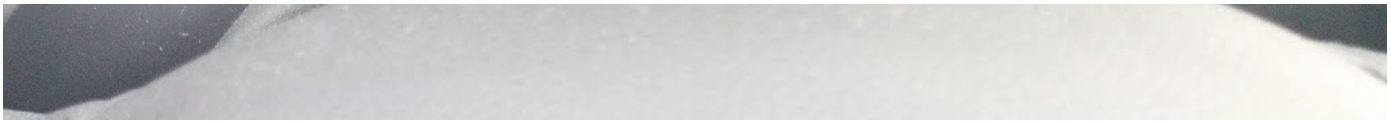
If the tachometer changes, good job, you found it! If not, identify the next message that correlates to RPM and replay it.

Fuzzing

Now that you have the CAN frame that sets the RPM on the instrument cluster, you can play with the data that you send to see what happens. We have noted that the the two bytes that correspond to RPM behave as a 16bit integer so in order to set the tachometer to 8k RPM, we run the following in your Linux prompt:

```
while true; do cansend can0 0C9#00**8000**0000101000; sleep 0.002; done
```

and the result is…



Spoofed RPM with engine turned off.

That's it! You can now try controlling the speedometer, radio, lights, door locks, etc. using the same approach.

Possible issues

- While the CAN bus is the most popular network, it's not the only network. If you can't find the message you are looking for on the CAN bus, try a different network. Especially non-critical messages such as radio, lights and door locks will probably be on a different network.
- As mentioned the exact data transmitted over CAN depends on the car's make, model and year. Some cars use a counter in the CAN message to ensure the same message isn't processed multiple times. This is slightly more difficult but you should be able to do it with the provided information. Some cars also use a checksum to ensure integrity of the data. Calculating this checksum can be difficult. If you have a Toyota, check out Adventures in Automotive Networks and Control Units, p10, Checksum-Toyota. Everyone should really read the whole paper.
- When replaying the identified packet on the bus, your CAN to USB device might go into the "bus off" state. This is part of the CAN standard and happens when the device experienced too many errors. This generally happens when there is a lot of traffic on the bus. In order to get around this you can play with delays and timing, maybe try replaying the message immediately after putting the car in accessory mode, try waiting a bit, try it with the car on, etc. If you've identified what ECU's are connected to the bus, you can also pull their fuse to stop them from sending messages and lower the traffic on the bus.

Must reads

- Car Hacker's Handbook
- Charlie Miller's and Chris Valasek's research, yes all of it
- University of California San Diego's and University of Washington's research.

Be sure to also check out Open Garages and their videos.

Cars | Technology | Tech | Self Improvement | Programming

1.1K claps

15

Kenny Kuchera

Car enthousiast

Follow

freeCodeCamp.org

Stories worth reading about programming and technology from our open source community.

Follow

Never miss a story from **freeCodeCamp.org**

GET UPDATES

A Prehistory of the Ethereum Protocol

Sep 14, 2017

Although the ideas behind the current Ethereum protocol have largely been stable for two years, Ethereum did not emerge all at once, in its current conception and fully formed. Before the blockchain has launched, the protocol went through a number of significant evolutions and design decisions. The purpose of this article will be to go through the various evolutions that the protocol went through from start to launch; the countless work that was done on the implementations of the protocol such as Geth, cppethereum, pyethereum, and EthereumJ, as well as the history of applications and businesses in the Ethereum ecosystem, is deliberately out of scope.

Also out of scope is the history of Casper and sharding research. While we can certainly make more blog posts talking about all of the various ideas Vlad, Gavin, myself and others came up with, and discarded, including “proof of proof of work”, hub-and-spoke chains, “hypercubes”, shadow chains (arguably a precursor to Plasma), chain fibers, and various iterations of Casper, as well as Vlad’s rapidly evolving thoughts on reasoning about incentives of actors in consensus protocols and properties thereof, this would also be far too complex a story to go through in one post, so we will leave it out for now.

Let us first begin with the very earliest version of what would eventually become Ethereum, back when it was not even called Ethereum. When I was visiting Israel in October 2013, I spent quite a bit of time with the Mastercoin team, and even suggested a few features for them. After spending a couple of times thinking about what they were doing, I sent the team a proposal to make their protocol more generalized and support more types of contracts without adding an equally large and complex set of features:

https://web.archive.org/web/20150627031414/http://vbuterin.com/ultimatescripting.html

Ultimate Scripting: A Platform for Generalized Financial Contracts on Mastercoin

0.1. Introduction

Perhaps the key advantage of Mastercoin over the raw Bitcoin protocol is the potential to include much more advanced transaction types, including transactions that specify behavior based on future information well off into the future. For example, Mastercoin joins Ripple in being one of the only two major cryptocurrency networks that include the ability for users to make binding exchange offers as a type of transaction. From there, the Mastercoin Foundation intends to integrate even more complex contracts, including bets, contracts for difference and on-blockchain dice rolls. However, up until this point Mastercoin has been taking a relatively unstructured process in developing these idea, essentially treating each one as a separate "feature" with its own transaction code and rules. This document outlines an alternative way of specifying Mastercoin contracts which follows an open-ended philosophy, specifying only the basic data and arithmetic building blocks and allowing anyone to craft arbitrarily complex Mastercoin contracts to suit their own needs, including needs which we may not even anticipate.

0.2. Specification

The underlying idea behind this specification is to allow anyone to create a contract which pays out according to an arbitrary formula. The formula will be defined in a Bitcoin-like stack-based scripting language, consisting of numbers and opcodes.

The evaluation algorithm is as follows:

```
dataStack = []
opStack = script
while len(opStack) > 0:
    var op = opStack.pop()
    if typeof(op) == 'opcode': eval(dataStack,op)
    else: dataStack.push(op)
return dataStack.pop()
```

Where `eval` is defined for each opcode below. Any error (eg. division by zero) will make the script return `FAIL`, and result in the entire transaction being treated as invalid by the Mastercoin network. All variables will be signed 64-bit integers, and all arithmetic operations wrap around (that is, if the underlying arithmetic operation returns `R`, the value pushed is `((R + 2^63) % 2^64) - 2^63`.

Notice that this is very far from the later and more expansive vision of Ethereum: it specialized purely in what Mastercoin was trying to specialize in already, namely two-party contracts where parties A and B would both put in money, and then they would later get money out according to some formula specified in the contract (eg. a bet would say “if X happens then give all the money to A, otherwise give all the money to B”). The scripting language was not Turing-complete.

The Mastercoin team was impressed, but they were not interested in dropping everything they were doing to go in this direction, which I was increasingly convinced is the correct choice. So here comes version 2, circa December:

https://web.archive.org/web/20131219030753/http://vitalik.ca/ethereum.html

Ethereum: The Ultimate Smart Contract and Autonomous Corporation Platform on the Blockchain

In the last few months, there has been a great amount of interest into the area of using the Bitcoin blockchain, the mechanism that allows for the entire world to agree on the state of a public ownership database, for more than just money. Perhaps the first, and oldest, such alternative application is colored coins, which is a protocol that allows users to label specific bitcoins and treat them as assets representing some real world value - whether company shares, collectibles or even existing currencies like gold and USD. A more independent alternative, Ripple, also includes the ability to create custom currencies and assets, but adds a decentralized exchange. More recently, Mastercoin has started to go even further, allowing more complex financial contracts such as hedging, trust-free dice rolls, binary options and self-stabilizing currencies - essentially, almost any common financial instrument imaginable. Taken together, all of these projects can be thought of as initial efforts toward a sort of "cryptocurrency 2.0" - they are to Bitcoin what Web 2.0 was to the World Wide Web circa 1995.

At the same time, there has been significant interest in "decentralized autonomous corporations" - autonomous entities that operate on the blockchain in a completely transparent and publicly managed way without any central control whatsoever. Rather than the relationships of the investors, owners and employees of the corporation being mediated by a legal contract or a set of organizational bylaws, the funds and corporate resources are managed directly on the blockchain. However, decentralized autonomous corporations are difficult to implement today, simply because the scripting systems of Bitcoin, and even proto-cryptocurrency 2.0 alternatives like Ripple and Mastercoin, are far too limited to allow the kind of arbitrarily complex computation that DACs require. Although these platforms have begun to offer increasingly complex contracts such as financial derivatives, order matching and trust-free bets, the way that the protocols are set up is inherently limited and closed-ended: each of these use cases is treated as a specific transaction type, not allowing any way for users to build contracts that the developers have not specifically chosen to include.

What this project intends to do is take cryptocurrency 2.0, and generalize it - create a fully-fledged, Turing-complete (but heavily fee-regulated) cryptographic ledger that allows participants to encode arbitrarily complex contracts, autonomous agents and relationships that will be mediated entirely by the blockchain. On-chain currencies, futures contracts, prediction markets, Namecoin-style domain name systems and even provably fair gambling sites will become trivial to implement, existing as simple, hundred-line-of-code contracts on the chain.

Basic Building Blocks

Here you can see the results of a substantial rearchitecting, largely a result of a long walk through San Francisco I took in November once I realized that smart contracts could potentially be fully generalized. Instead of the scripting language being simply a way of describing the terms of relations between two parties, contracts were themselves fully-fledged accounts, and

had the ability to hold, send and receive assets, and even maintain a permanent storage (back then, the permanent storage was called “memory”, and the only temporary “memory” was the 256 registers). The language switched from being a stack-based machine to being a register-based one on my own volition; I had little argument for this other than that it seemed more sophisticated.

Additionally, notice that there is now a built-in fee mechanism:

Whenever ether is sent to a script, the following happens:

1. The ether's endowment increases by the amount sent
2. All registers are reset to zero.
3. The sender is placed into R0.
4. The value sent is placed into R1.
5. The fee is placed into R2.
6. The index pointer is set to zero, and `STEPCOUNT = 0`
7. Repeat forever:
    - set `TOTALFEE = 0`
    - set `STEPCOUNT <- STEPCOUNT + 1`
    - if `STEPCOUNT > 16`, set `TOTALFEE <- TOTALFEE + STEPFEE`
    - see if the command at the index pointer is a valid command and not `STOP`. If it is invalid or `STOP`, HALT and break out of the loop
    - see if the command will do any modifications to the contract. If so, set `TOTALFEE <- TOTALFEE + DATAFEE`
    - see if the command will fill up a previously zero memory field. If so, set `TOTALFEE <- TOTALFEE + MEMORYFEE`
    - see if the commend will zero a previously used memory field. If so, set `TOTALFEE <- TOTALFEE - MEMORYFEE`
    - see if the command is `EXTRO` or `BALANCE`. If so, set `TOTALFEE <- TOTALFEE + EXTROFEE`
    - see if the command is `MKTX` or `RAWTX`. If so, set `TOTALFEE <- TOTALFEE + (transaction's value plus transaction's fee)`
    - if `TOTALFEE > contract's endowment`, HALT and break out of the loop
    - else, subtract `TOTALFEE` from contract's endowment. Note that `TOTALFEE` may be negative in some cases, in which case the endowment would actually increase
    - run the command

At this point, ether literally was gas; after every single computational step, the balance of the contract that a transaction was calling would drop a little bit, and if the contract ran out of money execution would halt. Note that this “receiver pays” mechanism meant that the contract itself had to require the sender to pay the contract a fee, and immediately exit if this fee is not present; the protocol allocated an allowance of 16 free execution steps to allow contracts to reject non-fee-paying transactions.

This was the time when the Ethereum protocol was entirely my own creation. From here on, however, new participants started to join the fold. By far the most prominent was Gavin Wood, who reached out to me in an about.me message in December 2013:

Gav Wood sent you a message on about.me
1 message

i@gavwood.com <i@gavwood.com> Thu, Dec 19, 2013 at 11:53 AM
Reply-To: i@gavwood.com
To: vbuterin@gmail.com

me

Hi **Vitalik**!
View Dashboard

Gav Wood sent you a message

"Johnny gave me the heads up - I can do C++ (e.g. github/gavofyork). How far are you with ethereum?

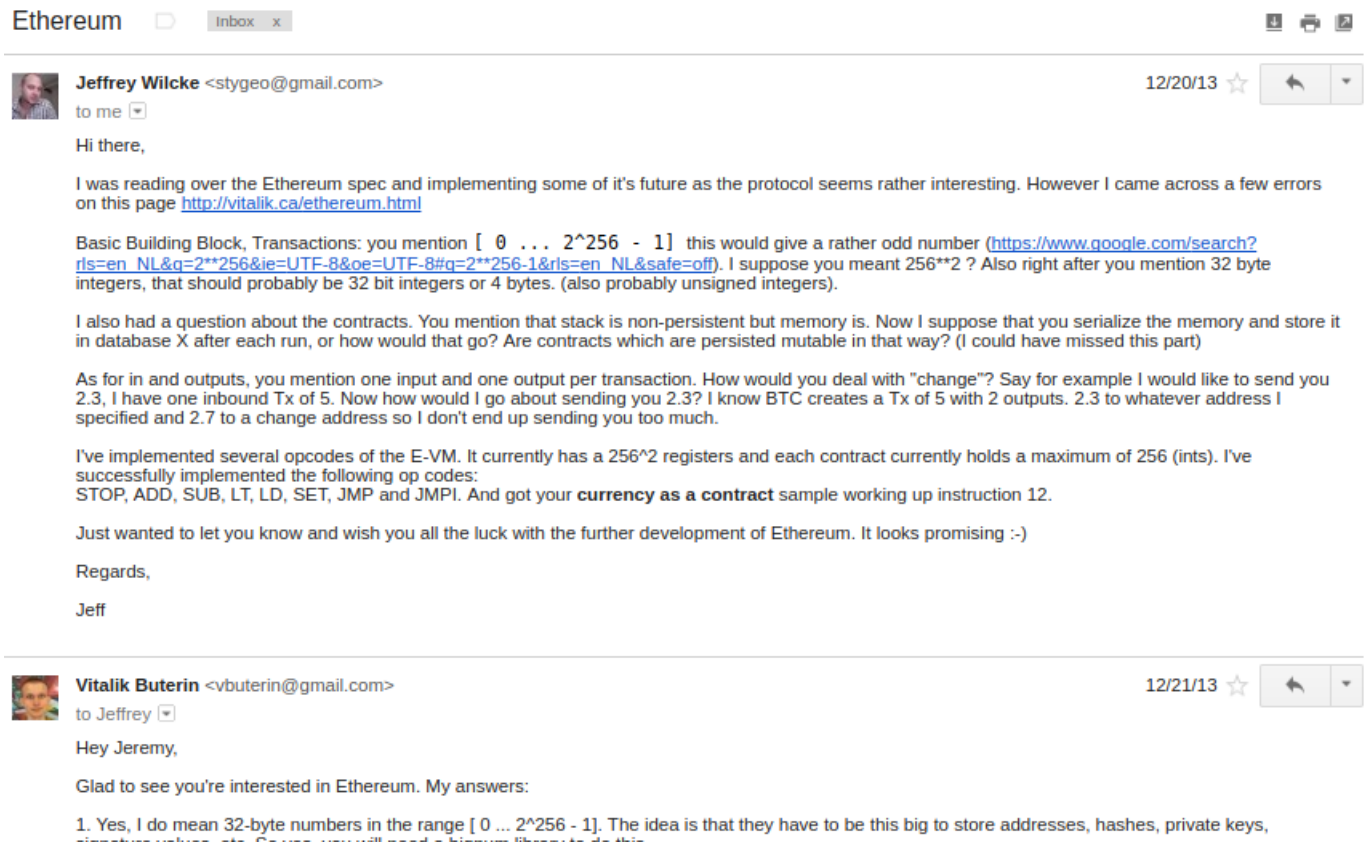
REPLY TO GAV

This email was sent to you by about.me/gavwood, and is not an official communication from about.me.

Cheers,
The about.me team

Don't want these emails? One Click Unsubscribe
Terms of Service | Privacy Policy
about.me 2601 Mission St San Francisco, CA 94110

Jeffrey Wilcke, lead developer of the Go client (back then called “ethereal”) also reached out and started coding around the same time, though his contributions were much more on the side of client development rather than protocol research.



Ethereum Inbox x

Jeffrey Wilcke <stygeo@gmail.com> 12/20/13

to me

Hi there,

I was reading over the Ethereum spec and implementing some of it's future as the protocol seems rather interesting. However I came across a few errors on this page http://vitalik.ca/ethereum.html

Basic Building Block, Transactions: you mention `[ 0 ... 2^256 - 1]` this would give a rather odd number (https://www.google.com/search?rls=en_NL&q=2**256&ie=UTF-8&oe=UTF-8#q=2**256-1&rls=en_NL&safe=off). I suppose you meant 256**2 ? Also right after you mention 32 byte integers, that should probably be 32 bit integers or 4 bytes. (also probably unsigned integers).

I also had a question about the contracts. You mention that stack is non-persistent but memory is. Now I suppose that you serialize the memory and store it in database X after each run, or how would that go? Are contracts which are persisted mutable in that way? (I could have missed this part)

As for in and outputs, you mention one input and one output per transaction. How would you deal with "change"? Say for example I would like to send you 2.3, I have one inbound Tx of 5. Now how would I go about sending you 2.3? I know BTC creates a Tx of 5 with 2 outputs. 2.3 to whatever address I specified and 2.7 to a change address so I don't end up sending you too much.

I've implemented several opcodes of the E-VM. It currently has a 256^2 registers and each contract currently holds a maximum of 256 (ints). I've successfully implemented the following op codes:
STOP, ADD, SUB, LT, LD, SET, JMP and JMPI. And got your **currency as a contract** sample working up instruction 12.

Just wanted to let you know and wish you all the luck with the further development of Ethereum. It looks promising :-)

Regards,

Jeff

Vitalik Buterin <vbuterin@gmail.com> 12/21/13

to Jeffrey

Hey Jeremy,

Glad to see you're interested in Ethereum. My answers:

1. Yes, I do mean 32-byte numbers in the range [0 ... 2^256 - 1]. The idea is that they have to be this big to store addresses, hashes, private keys,

"Hey Jeremy, glad to see you're interested in Ethereum…"

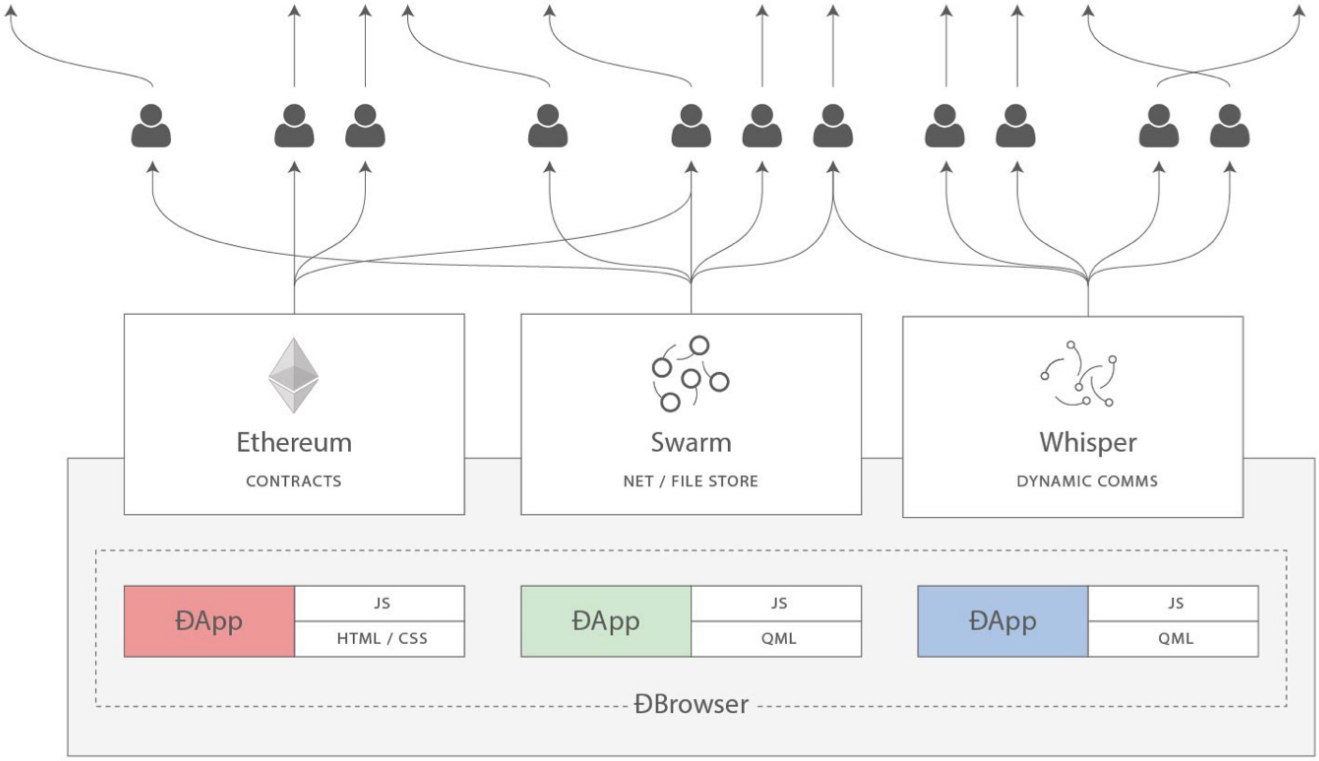
Gavin's initial contributions were two-fold. First, you might notice that the contract calling model in the initial design was an asynchronous one: although contract A could create an "internal transaction" to contract B ("internal transaction" is Etherscan's lingo; initially they were just called "transactions" and then later "message calls" or "calls"), the internal transaction's execution would not start until the execution of the first transaction completely finished. This meant that transactions could not use internal transactions as a way of getting information from other contracts; the only way to do that was the EXTRO opcode (kind of like an SLOAD that you could use to read other contracts' storage), and this too was later removed with the support of Gavin and others.

When implementing my initial spec, Gavin naturally implemented internal transactions synchronously without even realizing that the intent was different - that is to say, in Gavin's implementation, when a contract calls another contract, the internal transaction gets executed immediately, and once that execution finishes, the VM returns back to the contract that created the internal transaction and proceeds to the next opcode. This approach seemed to both of us to be superior, so we decided to make it part of the spec.

Second, a discussion between him and myself (during a walk in San Francisco, so the exact details will be forever lost to the winds of history and possibly a copy or two in the deep archives of the NSA) led to a re-factoring of the transaction fee model, moving away from the "contract pays" approach to a "sender pays" approach, and also switching to the "gas"

architecture. Instead of each individual transaction step immediately taking away a bit of ether, the transaction sender pays for and is allocated some “gas” (roughly, a counter of computational steps), and computational steps drew from this allowance of gas. If a transaction runs out of gas, the gas would still be forfeit, but the entire execution would be reverted; this seemed like the safest thing to do, as it removed an entire class of “partial execution” attacks that contracts previously had to worry about. When a transaction execution finishes, the fee for any unused gas is refunded.

Gavin can also be largely credited for the subtle change in vision from viewing Ethereum as a platform for building programmable money, with blockchain-based contracts that can hold digital assets and transfer them according to pre-set rules, to a general-purpose computing platform. This started with subtle changes in emphasis and terminology, and later this influence became stronger with the increasing emphasis on the “Web 3” ensemble, which saw Ethereum as being one piece of a suite of decentralized technologies, the other two being Whisper and Swarm.



There were also changes made around the start of 2014 that were suggested by others. We ended up moving back to a stack-based architecture after the idea was suggested by Andrew Miller and others.

On 12/19/2013 03:40 PM, Andrew Miller wrote:
> Hi Vitalik,
> I'd really like to talk with you more about this. I'm really
> interested in extending the functionality of Bitcoin beyond trivial
> money-shoving financial transactions and up to user-customizable
> contracts, and I'm pretty stoked that you're taking a shot at this as
> a serious project.
>
> Here are some specific concerns/questions about your transaction language:
>
> 1. (Note: this is my most superficial criticism) Why did you design
> your own register language? What's wrong with a stack based language
> similar to Bitcoin? You can have a turing-complete and higher order
> stack language (look at Joy, Factor or Forth). If anything I'd
> recommend a lambda-calculus based language. From Stack-based to
> Register-based is such a superficial change and there's absolutely no
> motivation for it, yet most of your document is about minutiae related
> to this. When you present your contract examples, you're writing in
> pseudocode that isn't really any closer to ASM than to stack-based or
> functional anyway. You might also look at E, a language basd on
> javascript that was explicitly designed for the purpose of writing
> smart contracts. http://www.erights.org/elang/

Charles Hoskinson suggested the switch from Bitcoin's SHA256 to the newer SHA3 (or, more accurately, keccak256). Although there was some controversy for a while, discussions with Gavin, Andrew and others led to establishing that the size of values on the stack should be limited to 32 bytes; the other alternative being considered, unlimited-size integers, had the problem that it was too difficult to figure out how much gas to charge for additions, multiplications and other operations.

The initial mining algorithm that we had in mind, back in January 2014, was a contraption called Dagger:

https://github.com/ethereum/wiki/blob/master/Dagger.md

Algorithm specification:

Essentially, the Dagger algorithm works by creating a directed acyclic graph (the technical term for a tree where each node is allowed to have multiple parents) with a total of 2^{23} - 1 nodes in sequence. Each node depends on 3-15 randomly selected nodes before it. If the miner finds a node between index 2^{22} and 2^{23} such that this resulting hash is below 2^{256} divided by the difficulty parameter, the result is a valid proof of work.

Let `D` be the underlying data (eg. in Bitcoin's case the block header), `N` be the nonce and `||` be the string concatenation operator (ie. `'foo' || 'bar' == 'foobar'`) . The entire code for the algorithm is as follows:

```
D(data,xn,0) = sha3(data)
D(data,xn,n) =
    with v = sha3(data + xn + n)
         L = 2 if n < 2^21 else 11 if n < 2^22 else 3
         a[k] = floor(v/n^k) mod n for 0 <= k < 2
         a[k] = floor(v/n^k) mod 2^22 for 2 <= k < L
    sha3(v ++ D(data,xn,a[0]) ++ D(data,xn,a[1]) ++ ... ++ D(data,xn,a[L-1]))
```

Properties:

Objective: find `xn`, `n` such that `n > 2^22` and `D(data,xn,n) < 2^256 / diff`

Dagger was named after the “directed acyclic graph” (DAG), the mathematical structure that is used in the algorithm. The idea is that every N blocks, a new DAG would be pseudorandomly generated from a seed, and the bottom layer of the DAG would be a collection of nodes that takes several gigabytes to store. However, generating any individual value in the DAG would require calculating only a few thousand entries. A “Dagger computation” involved getting some number of values in random positions in this bottom-level dataset and hashing them together. This meant that there was a fast way to make a Dagger calculation - already having the data in memory, and a slow, but not memory intensive way - regenerating each value from the DAG that you need to get from scratch.

The intention of this algorithm was to have the same “memory-hardness” properties as algorithms that were popular at the time, like Scrypt, but still be light-client friendly. Miners would use the fast way, and so their mining would be constrained by memory bandwidth (the theory is that consumer-grade RAM is already very heavily optimized, and so it would be hard to further optimize it with ASICs), but light clients could use the memory-free but slower version for verification. The fast way might take a few microseconds and the slow but memory-free way a few milliseconds, so it would still be very viable for light clients.

From here, the algorithm would change several times over the course of Ethereum development. The next idea that we went through is “adaptive proof of work”; here, the proof of work would involve executing randomly selected Ethereum contracts, and there is a clever reason why this is expected to be ASIC-resistant: if an ASIC was developed, competing miners would have the incentive to create and publish many contracts that that ASIC was not good at executing. There is no such thing as an ASIC for general computation, the story goes, as that is just a CPU, so we could instead use this kind of adversarial incentive mechanism to make a proof of work that essentially was executing general computation.

This fell apart for one simple reason: long-range attacks. An attacker could start a chain from block 1, fill it up with only simple contracts that they can create specialized hardware for, and rapidly overtake the main chain. So… back to the drawing board.

The next algorithm was something called Random Circuit, described in this google doc here, proposed by myself and Vlad Zamfir, and analyzed by Matthew Wampler-Doty and others. The idea here was also to simulate general-purpose computation inside a mining algorithm, this time by executing randomly generated circuits. There’s no hard proof that something based on these principles could not work, but the computer hardware experts that we reached out to in 2014 tended to be fairly pessimistic on it. Matthew Wampler-Doty himself suggested a proof of work based on SAT solving, but this too was ultimately rejected.

Finally, we came full circle with an algorithm called “Dagger Hashimoto”. “Dashimoto”, as it was sometimes called in short, borrowed many ideas from Hashimoto, a proof of work algorithm by Thaddeus Dryja that pioneered the notion of “I/O bound proof of work”, where the

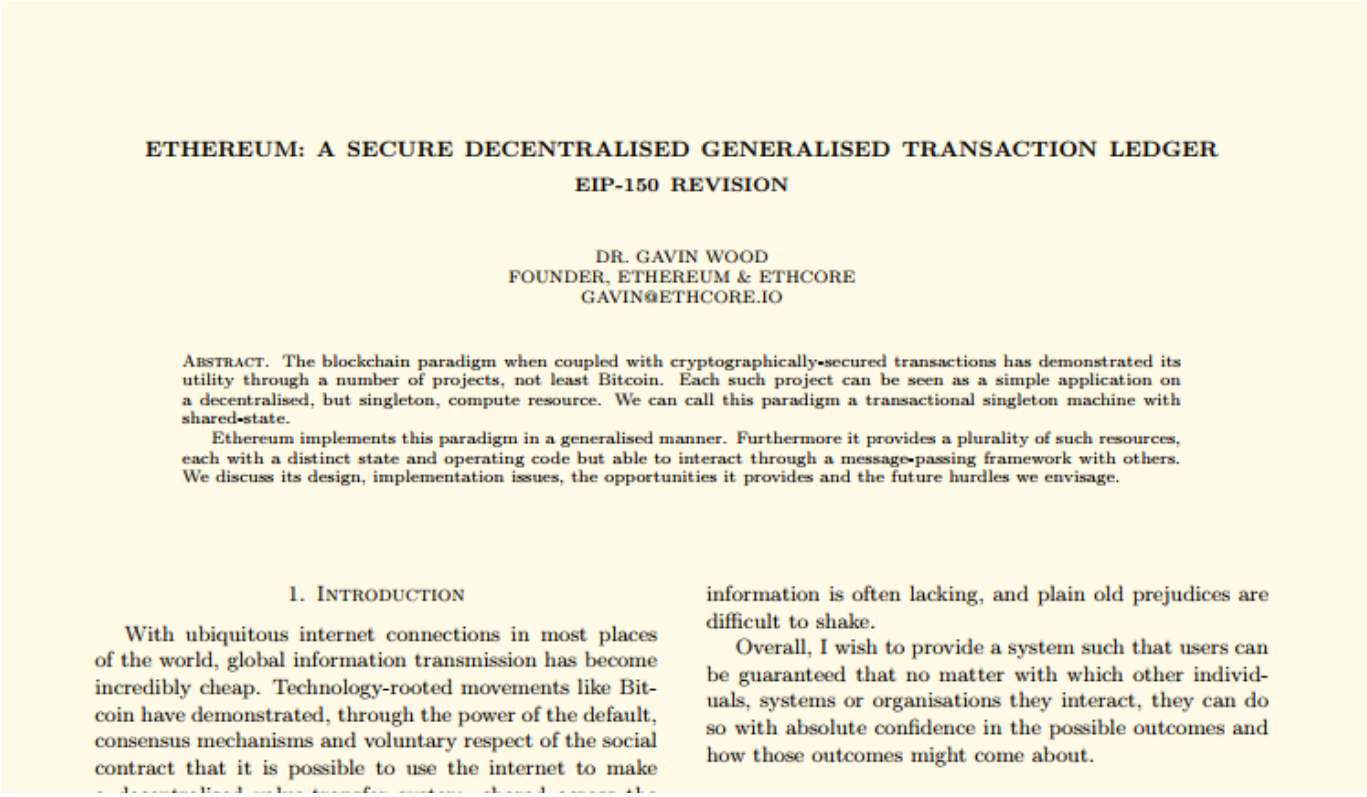
dominant limiting factor in mining speed was not hashes per second, but rather megabytes per second of RAM access. However, it combined this with Dagger's notion of light-client-friendly DAG-generated datasets. After many rounds of tweaking by myself, Matthew, Tim and others, the ideas finally converged into the algorithm we now call Ethash.

```
def hashimoto(header, nonce, full_size, dataset_lookup):
    n = full_size / HASH_BYTES
    w = MIX_BYTES // WORD_BYTES
    mixhashes = MIX_BYTES / HASH_BYTES
    # combine header+nonce into a 64 byte seed
    s = sha3_512(header + nonce[::-1])
    # start the mix with replicated s
    mix = []
    for _ in range(MIX_BYTES / HASH_BYTES):
        mix.extend(s)
    # mix in random dataset nodes
    for i in range(ACCESSES):
        p = fnv(i ^ s[0], mix[i % w]) % (n // mixhashes) * mixhashes
        newdata = []
        for j in range(MIX_BYTES / HASH_BYTES):
            newdata.extend(dataset_lookup(p + j))
        mix = map(fnv, mix, newdata)
    # compress mix
    cmix = []
    for i in range(0, len(mix), 4):
        cmix.append(fnv(fnv(fnv(mix[i], mix[i+1]), mix[i+2]), mix[i+3]))
    return {
        "mix digest": serialize_hash(cmix),
        "result": serialize_hash(sha3_256(s+cmix))
    }

def hashimoto_light(full_size, cache, header, nonce):
    return hashimoto(header, nonce, full_size, lambda x: calc_dataset_item(cache, x)

def hashimoto_full(full_size, dataset, header, nonce):
    return hashimoto(header, nonce, full_size, lambda x: dataset[x])
```

By the summer of 2014, the protocol had considerably stabilized, with the major exception of the proof of work algorithm which would not reach the Ethash phase until around the beginning of 2015, and a semi-formal specification existed in the form of Gavin's yellow paper.



ETHEREUM: A SECURE DECENTRALISED GENERALISED TRANSACTION LEDGER

EIP-150 REVISION

DR. GAVIN WOOD
FOUNDER, ETHEREUM & ETHCORE
GAVIN@ETHCORE.IO

Abstract. The blockchain paradigm when coupled with cryptographically-secured transactions has demonstrated its utility through a number of projects, not least Bitcoin. Each such project can be seen as a simple application on a decentralised, but singleton, compute resource. We can call this paradigm a transactional singleton machine with shared-state.

Ethereum implements this paradigm in a generalised manner. Furthermore it provides a plurality of such resources, each with a distinct state and operating code but able to interact through a message-passing framework with others. We discuss its design, implementation issues, the opportunities it provides and the future hurdles we envisage.

1. Introduction

With ubiquitous internet connections in most places of the world, global information transmission has become incredibly cheap. Technology-rooted movements like Bitcoin have demonstrated, through the power of the default, consensus mechanisms and voluntary respect of the social contract that it is possible to use the internet to make

information is often lacking, and plain old prejudices are difficult to shake.

Overall, I wish to provide a system such that users can be guaranteed that no matter with which other individuals, systems or organisations they interact, they can do so with absolute confidence in the possible outcomes and how those outcomes might come about.

In August 2014, I developed and introduced the uncle mechanism, which allows Ethereum's blockchain to have a shorter block time and higher capacity while mitigating centralization risks. This was introduced as part of PoC6.

Discussions with the Bitshares team led us to consider adding heaps as a first-class data structure, though we ended up not doing this due to lack of time, and later security audits and DoS attacks will show that it is actually much harder than we had thought at the time to do this safely.

In September, Gavin and I planned out the next two major changes to the protocol design. First, alongside the state tree and transaction tree, every block would also contain a "receipt tree". The receipt tree would include hashes of the logs created by a transaction, along with intermediate state roots. Logs would allow transactions to create "outputs" that are saved in the blockchain, and are accessible to light clients, but that are not accessible to future state calculations. This could be used to allow decentralized applications to easily query for events, such as token transfers, purchases, exchange orders being created and filled, auctions being started, and so forth.

There were other ideas that were considered, like making a Merkle tree out of the entire execution trace of a transaction to allow anything to be proven; logs were chosen because they were a compromise between simplicity and completeness.

The second was the idea of "precompiles", solving the problem of allowing complex cryptographic computations to be usable in the EVM without having to deal with EVM overhead. We had also gone through many more ambitious ideas about "native contracts",

where if miners have an optimized implementation of some contracts they could “vote” the gasprice of those contracts down, so contracts that most miners could execute much more quickly would naturally have a lower gas price; however, all of these ideas were rejected because we could not come up with a cryptoeconomically safe way to implement such a thing. An attacker could always create a contract which executes some trapdoored cryptographic operation, distribute the trapdoor to themselves and their friends to allow them to execute this contract much faster, then vote the gasprice down and use this to DoS the network. Instead we opted for the much less ambitious approach of having a smaller number of precompiles that are simply specified in the protocol, for common operations such as hashes and signature schemes.

Gavin was also a key initial voice in developing the idea of “protocol abstraction” - moving as many parts of the protocol such as ether balances, transaction signing algorithms, nonces, etc into the protocol itself as contracts, with a theoretical final goal of reaching a situation where the entire ethereum protocol could be described as making a function call into a virtual machine that has some pre-initialized state. There was not enough time for these ideas to get into the initial Frontier release, but the principles are expected to start slowly getting integrated through some of the Constantinople changes, the Casper contract and the sharding specification.

This was all implemented in PoC7; after PoC7, the protocol did not really change much, with the exception of minor, though in some cases important, details that would come out through security audits…

In early 2015, came the pre-launch security audits organized by Jutta Steiner and others, which included both software code audits and academic audits. The software audits were primarily on the C++ and Go implementations, which were led by Gavin Wood and Jeffrey Wilcke, respectively, though there was also a smaller audit on my pyethereum implementation. Of the two academic audits, one was performed by Ittay Eyal (of “selfish mining” fame), and the other by Andrew Miller and others from Least Authority. The Eyal audit led to a minor protocol change: the total difficulty of a chain would not include uncles. The Least Authority audit was more focused on smart contract and gas economics, as well as the Patricia tree. This audit led to several protocol changes. One small one is the use of sha3(addr) and sha3(key) as trie keys instead of the address and key directly; this would make it harder to perform a worst-case attack on the trie.

There are useful parallels between this refund loop and the publish-subscribe function illustrated in Miller's thesis. He demonstrates several hazards that are present when the `publish` callbacks are run synchronously:

- exceptions raised during the callback would prevent execution of later callbacks
- reentrancy hazards if the callback itself executes publish(), subscribe(), or unsubscribe(): repeated actions, missing actions, and inconsistent delivery of messages

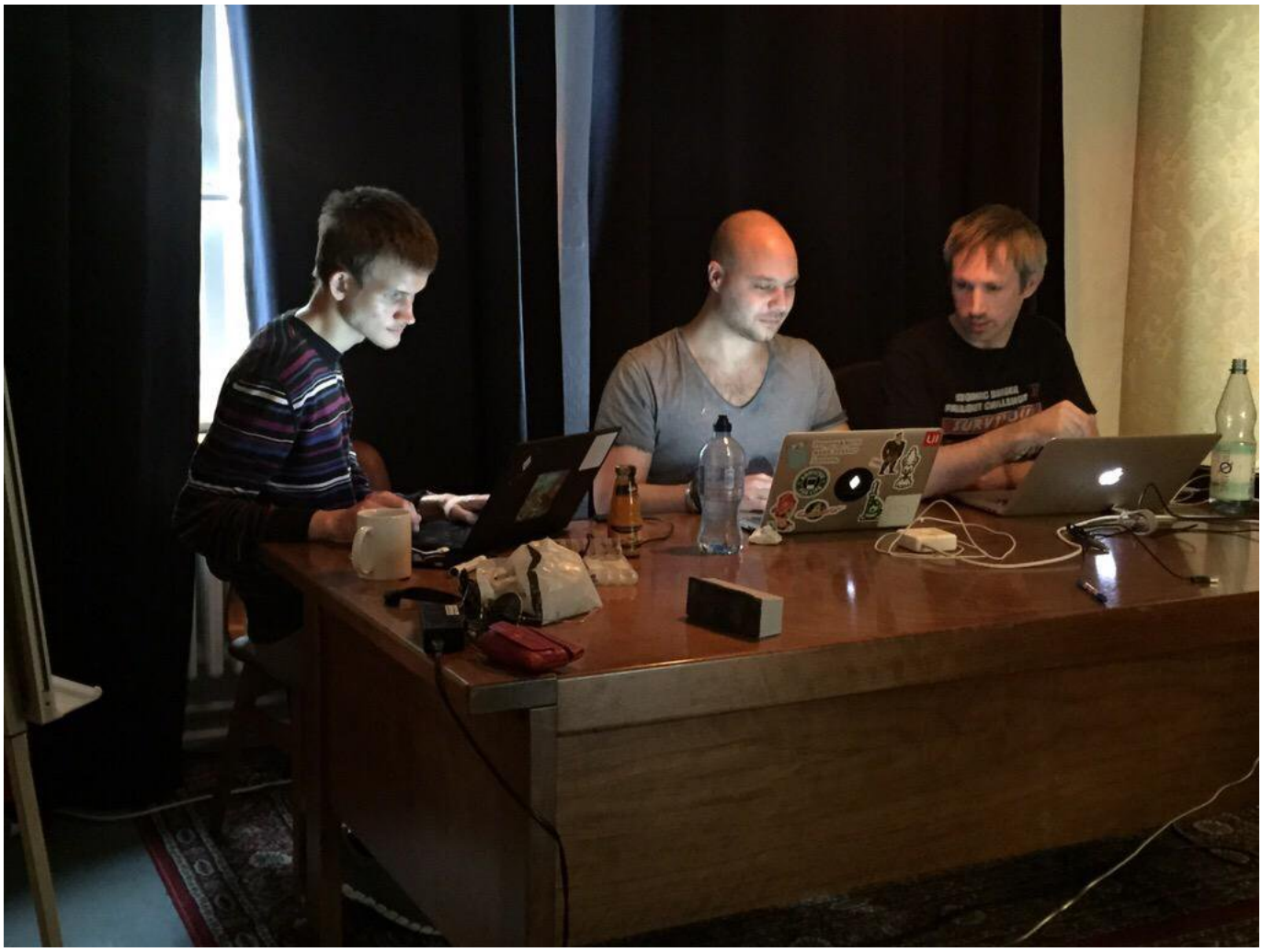
Some analogous issues in the crowdfund example are:

- delivering a contribution, after the funding deadline, with just enough gas to allow some refunds to go through, but not all: the contract could be left in a state where it was unable to refund the remaining contributions
- if the refund was triggered by a contract at the end of a long call stack, the `send` instructions will fail. However the example appears to ignore the return value of the `send`, so execution will continue. All records will be cleared, and the funds can never be recovered.
- the refund callback could make a new donation, triggering another refund cycle, potentially double-refunding the earlier contributions, or failing to refund later ones

And a warning that was perhaps a bit too far ahead of its time…

Another significant thing that we discussed was the gas limit voting mechanism. At the time, we were already concerned by perceived lack of progress in the bitcoin block size debate, and wanted to have a more flexible design in Ethereum that could adjust over time as needed. But the challenge is: what is the optimal limit? My initial thought had been to make a dynamic limit, targeting 1.5 times the long-term exponential moving average of the actual gas usage, so that in the long run on average blocks would be 2/3 full. However, Andrew showed that this was exploitable in some ways - specifically, miners who wanted to raise the limit would simply include transactions in their own blocks that consume a very large amount of gas, but take very little time to process, and thereby always create full blocks at no cost to themselves. The security model was thus, at least in the upward direction, equivalent to simply having miners vote on the gas limit.

We did not manage to come up with a gas limit strategy that was less likely to break, and so Andrew's recommended solution was to simply have miners vote on the gas limit explicitly, and have the default strategy for voting be the 1.5x EMA rule. The reasoning was that we were still very far from knowing the right approach for setting maximum gas limits, and the risk of any specific approach failing seemed greater than the risk of miners abusing their voting power. Hence, we might as well simply let miners vote on the gas limit, and accept the risk that the limit will go too high or too low, in exchange for the benefit of flexibility, and the ability for miners to work together to very quickly adjust the limit upwards or downwards as needed.



After a mini-hackathon between Gavin, Jeff and myself, PoC9 was launched in March, and was intended to be the final proof of concept release. A testnet, Olympic, ran for four months, using the protocol that was intended to be used in the livenet, and Ethereum's long-term plan was established. Vinay Gupta wrote a blog post, "The Ethereum Launch Process", that described the four expected stages of Ethereum livenet development, and gave them their current names: Frontier, Homestead, Metropolis and Serenity.

Olympic ran for four months. In the first two months, many bugs were found in the various implementations, consensus failures happened, among other issues, but around June the network noticeably stabilized. In July a decision was made to make a code-freeze, followed by a release, and on July 30 the release took place.

Vitalik Buterin's website

Vitalik Buterin's website

vbuterin
VitalikButerin

Lil Wayne-Young Moula Baby-2008

Language English

Identifier Lil_Wayne-Young_Moula_Baby-2008
Scanner Internet Archive Python library 1.0.10

Reviews

Add Review

There are no reviews yet. Be the first one to write a review.

8,568 Views

1 Favorite

DOWNLOAD OPTIONS

INFORMATION 1 file

ITEM TILE	1 file
JPEG	1 file
OGG VORBIS	20 files
SIMPLE FILE VERIFICATION	1 file
TORRENT	1 file
VBR M3U	1 file
VBR MP3	20 files

SHOW ALL

90 Files

29 Original

IN COLLECTIONS

Hip Hop Mixtapes

Folksoundomy: A Library of Sound

Uploaded by

Sketch the Cow

on November 5, 2016

ISSN 1551-3483

https://scale.qihardware.org