Amsterdam University of Applied Sciences

FACULTY OF DIGITAL MEDIA AND CREATIVE INDUSTRIES
HBO – Information and Communication Technologies

The theoretical design of a QPU

Quantum Stack

Lilith Bertens

June 13, 2024

1 Introduction

As quantum computers are known today, they are big bulky machines which can take up entire rooms. Well, the devices themselves commonly are not this big, but all the equipment which accompanies really adds up. So a similar process might be expected for the quantum computer, it will shrink in size until it can fit in your pocket.

This paper aims to explore the possibilities of shrinking current technology in size to let it fit within a desktop computer. Where this small quantum computing device will be referenced to as a QPU, Quantum Processing Unit. Something akin to a GPU. Where the aim is to design one with as few assumptions and leaps in technology as possible.

The second chapter will talk about what current technology is out there and determine which would be the most fitting for a QPU. Chapter 3 will delve into what current technology is used to let a motherboard communicate with a GPU and how it could be replicated. Chapter 4 will take a look at what modules would be necessary for the QPU to work and combine all of these into a full design and chapter 5 will bode the conclusion to the previous chapters and results.

2 What would fit?

There are plenty of types of quantum computers out there. All with their upsides and downsides. Lets first take a look at what technologies are out there:

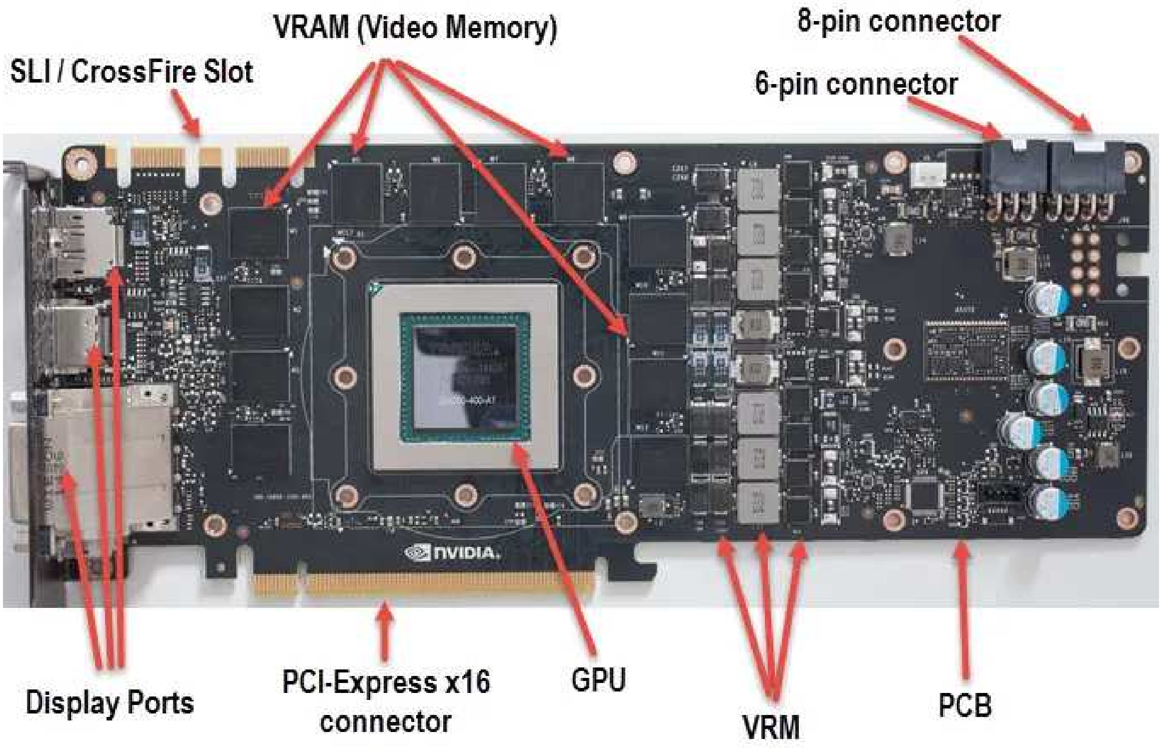
- Superconducting
- Trapped ion
- Silicon dot
- Photonic
- Neutral atoms
- NV diamond

Where, to be able to fit one of these technologies into a QPU without too many assumptions, it would need to be able to function at room temperature and pressure. The choice already gets limited a lot here, as only NV diamonds can function at room temperature and pressure. This gives us the exact technology to be used for a QPU.

How does it work? The nv diamond quantum computing method is named after the NV diamond faults it uses. An NV fault is a type of fault which can both be naturally present in diamonds or be introduced. Where one carbon is replaced with a nitrogen and another connected carbon is completely removed, creating a vacancy. Which is where it gets its name from, Nitrogen Vacancy diamond faults [1].

These faults are then used together with a magnetic field or other method of delivering energy to steer and adjust the faults. So they can be operated on and measured. That being said, there is very little info on how this is done exactly out there, a lot of it is kept behind bars at the moment. So assumptions will have to be made regarding how this is to be controlled and how much voltage a Q-bit will need [2].

Figure 1: The layout of a GPU



3 What do GPU's do?

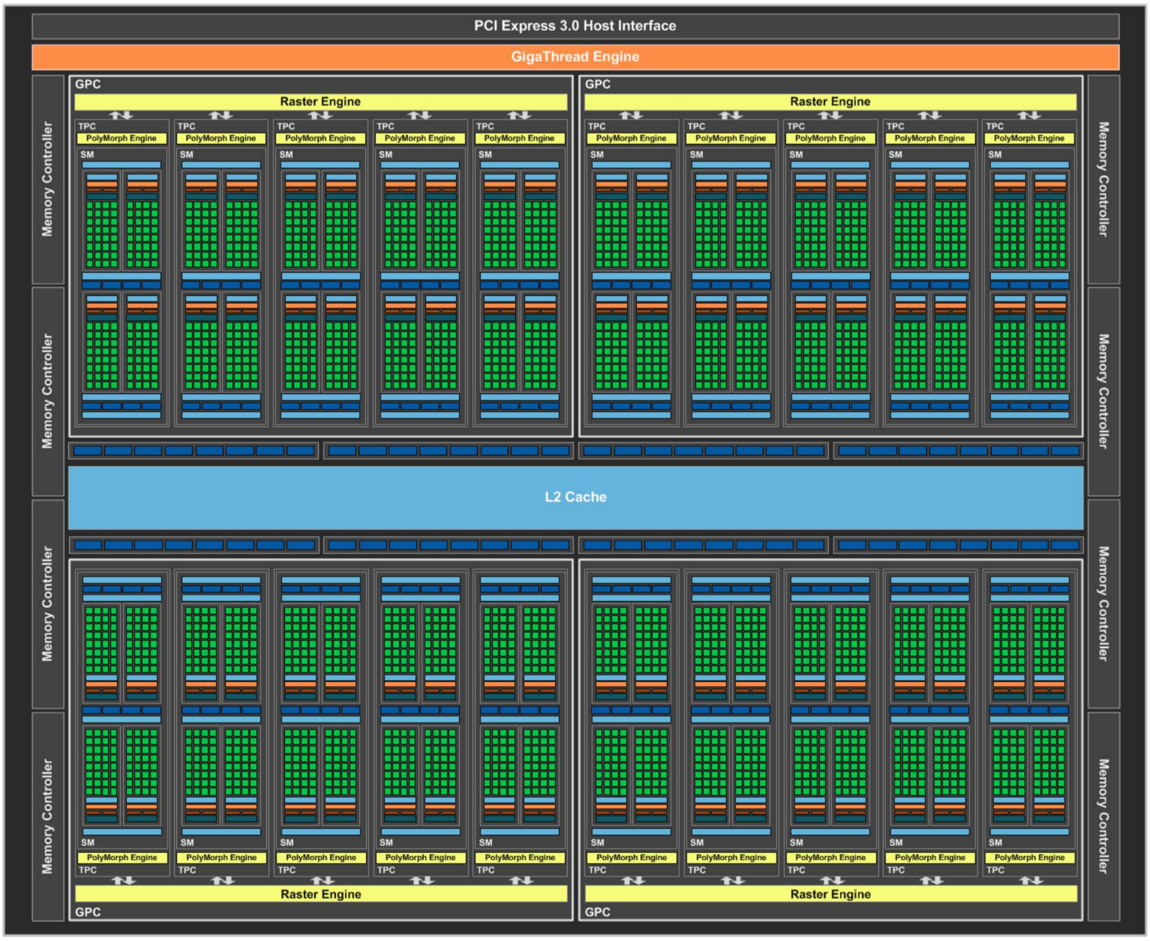
A GPU is used to process a 3D scene into an image which can be displayed to the screen. It gets the vertexes, materials, shaders and all other things which influence the look of a scene and turn it into an image, a frame. It does this 60 times per second most commonly, sometimes even faster if the screen is made for it and the GPU can handle that workload [3].

The GPU consists of a couple things, see figure 1 for reference:

- a PCIe slot
- GPU chip
- Display ports
- VRAM
- a PCB
- VRM
- x-Pin connectors

This list is not exhaustive. but will do for the purposes of explaining how a GPU works. Indeed, the GPU is both the name of the device and the chip on the device. This is due to the trivial name of the device being linked to the most important component. The GPU is a chip which

Figure 2: Layout of a GPU chip



has many smaller sub area's for parallel calculation as can be seen in figure 2, as a scene can contain a lot of objects, pixels and effects. Which can be somewhat calculated in parallel, keep in mind that on stage follows after another. Shaders are done after meshes for example, but multiple meshes can be calculated at the same time [3].

The second most important component is VRAM. This is in essence just RAM but is located on the graphics card, which allows it to access this RAM quicker. This ram is used to store the input and output of the GPU. An example of what it would store is: information on the 3D scene and the partially completed image it is rendering. Where the amount of VRAM the card has will affect the complexity the scene may have before the card starts having issues with a bottleneck.

Just to not go on too long, The PCB is the plate all of the components are soldered on. The VRM regulates the voltage going to components, the display port outputs the frames the GPU calculates to a screen. The PCIe slot is where the GPU slots into the motherboard and where it communicates with the CPU. The x-pin connectors describes the amount of pins it needs, the more power a GPU needs the more pins it will require. These pins then connect back to the power supply in a desktop PC.

4 Modules and total design

Combining a GPU with the currently available quantum computing technique should give a vague description as to what a full QPU should look like. Keep in mind, that due to the fact that no data-sheets or exact hardware details on what these quantum chips would require and how they would communicate is available. Meaning that the design can only be up to a description of what components would be necessary and how it could work.

First off the quantum chip which would be used. For the sake of making an assumption, it will be using 5 volts for most components. Together with a transformer to power the part which needs to generate a magnetic field at 24 volts. This can be entirely off and is just an estimate. To keep the complexity relatively low, 5 of these smaller diamond chips will be used. The components to power this chip would be relatively common and not a huge hassle or cost. Besides of-course the diamond chip itself, which has to be produced in a lab under high pressure and heat.

There are several other components present on a GPU which will also be needed. These components are: VRM, PCIe, RAM, a PCB and some pin connectors. The VRM, PCB both somewhat speak for themselves, it needs to be able to distribute voltage around and the components should be soldered to something. The RAM is required as it needs to be able to both store results from running quantum circuits and to be able to store circuits it still needs to execute. To be able to take in the information of the circuits it needs to execute and to send back the results it will need a PCIe slot. This will also let the QPU fit into an everyday desktop setup. At last a set of pin connectors is needed, these will deliver the power to the QPU, it can be assumed that at max an 8-pin connector will be needed, although this might not be needed.

It should be kept in mind that this would only give you a QPU. A device capable of speeding up a select set of classical problems via quantum methods. It would not be a replacement to a GPU and neither would it be a replacement to a CPU. Both of those would still be necessary, although they might not need to be as powerful when some tasks can be handed of to the QPU. This would not matter in the case of a desktop PC but could be vital in regards to a virtual reality headset or even a laptop, where weight and room come in short supply. Where a simple quantum chip would cost less space then the cooling system required for more powerful GPU and CPU's.

The full design would look akin to a regular GPU like in figure 1 as a lot of the component it needs are also needed for a QPU. Which would allow for a relatively setup for a new production line in companies like NVIDIA (or the production companies which manufacture for them). Seeing as the assembly lines to produce these GPU's are already present and it would only require some smaller modifications.

5 Conclusions

In conclusion, a room temperature QPU which can fit in a desktop should be possible by current technologies. In fact, there is a company out there which is currently working on this. They are called Quantum brilliance, an Australian-German company. They currently have a server rack sized quantum computing unit out which has 5 Q-bits and are working on a GPU sized version with around 50 Q-bits [2]. Meaning that this concept is not outside of the realm of the possible. Sadly a full, detailed design will not be possible as no data is available on how these chips would work, communicate and how much power they would draw.

References

[1] Multiple, “Nitrogen vacancy center.”

[2] L. Blain, “Quantum computing at room temperature.”

[3] Nvidia, “How do gpu’s work.”