Hello everyone, my name is Alex Kor, and I want to tell you about my work on DNA sequencer for decoding DNA. The market price of such a device is about $ 150,000.

A brief digression to genetics. If suddenly you remember, in 2003 a sensational statement was made: scientists finally deciphered the human genome. The genome is built from DNA, and DNA is the source code of the organism. DNA is a double strand consisting of 4 types of nucleotides that are repeated about 3 billion times in the human genome. Just as in bits all the information on your computer is encrypted, instructions about the assembly of all proteins of the human body are encoded in the nucleotides. That is, knowing in what sequence nucleotides are located in DNA, we can theoretically gather all the necessary proteins and get a human model. So, in the standard sense, scientists did not decipher the DNA, but simply transferred the chemical sequence to a set of zeros and ones on the computer. What to do next is a separate conversation. For example, at the moment, only 5% of the entire genome array is clear for us (this is protein coding). What do the remaining 95%, we can only guess.

In 2003, the cost of sequencing human DNA was about $ 100 million. Over time, this figure has decreased and now it is approaching a thousand dollars. You pay, your DNA is sequenced, and they give you a hard disk with 3 GB of information - your genome in digital form.

Today there are three main sequencers on the market. The most productive, Hiseq, and its receiver NovaSeq, provides the cheapest (fluorescent) sequencing. One of its launch lasts several days, and during this time the genomes of several people are processed at once. However, the launch itself costs about ten thousand dollars. By the way, the device itself costs about $ 1 million, and since it becomes obsolete in about 3 years, in order for it to pay off, it must bring you $ 1,000 a day.

The second device appeared on the market just a couple of years ago. It is called Nanopore and is based on a very interesting technology, when DNA is sequenced by passing through a nanopore. The cheapest version of Nanopore is positioned as a disposable home sequencer and costs $ 1000.

The third device is PGM, a semiconductor sequencer that costs $ 50,000. The sequencing process on it takes about a few hours.

https://www.thermofisher.com/order/catalog/product/4462921

My instrument is based on a similar principle, but sequencing occurs in real time. First, briefly on how semiconductor sequencing occurs. The entire DNA strand is divided into fragments of 300-400 nucleotides in length, called reads. Then the reads attach to small spheres and are copied many times - as a result, a whole bunch of identical DNA fragments hangs on each sphere. Copying is necessary to enhance the signal from each particular read. A set of different spheres is called a DNA library.

The heart of the PGM is a disposable chip - a matrix similar to the matrix in the camera, only instead of pixels that react to light, here the pH transistors react to changes in the acid-base balance. The resulting library of DNA is loaded onto a chip containing 10 million wells, at the bottom of each of them there is a pH transistor. Only one sphere can fit in a well and, therefore, reads of only one type (with one specific nucleotide sequence). Next, reagents are fed to the chip in such a way that the DNA starts copying itself. And it is copied linearly, that is, the nucleotides are attached to the newly created chain in the order in which they stand in the parent chain. Therefore, one type of nucleotide is fed to the chip - and a change in pH is immediately recorded in some wells (this means that the nucleotide was attached to them). Then another type of nucleotide is supplied and the change in pH in the wells is recorded, etc. Thus, by submitting all 4 types of nucleotides to the chip many times, we can get information about the sequence of nucleotides in each read. Then, in mathematical ways, read short segments are collected on a computer in a single chain. To collect it more or less confidently, each read must be read about 100 times.

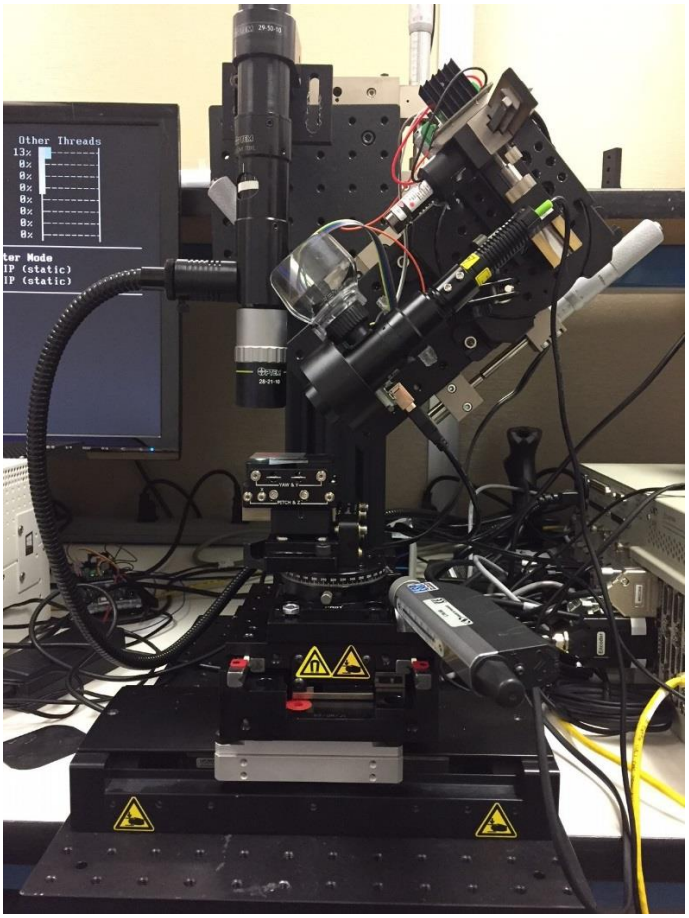
Device. There is, as we already know, a chip, as well as a reagent supply system and a motherboard. All sequencing is done on the chip — the rest of the device only transmits certain signals to it, delivers reagents, reads analog signals from it, digitizes them, and drives the resulting information stream to a computer, where data is accumulated and processed.

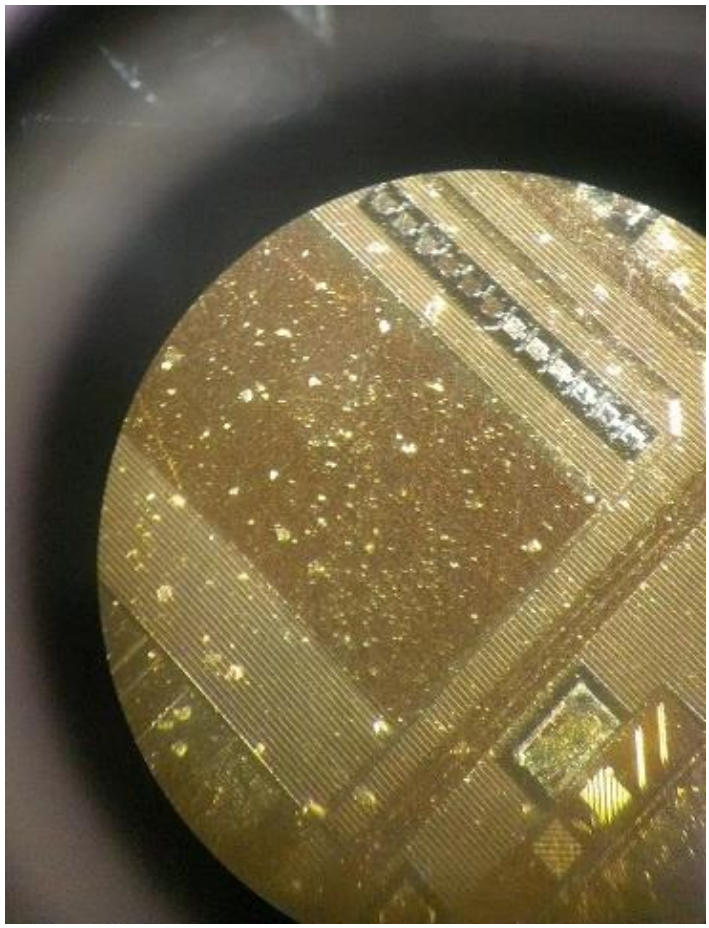
The chip is positioned as a disposable and is discarded after use. Accordingly, where PGM works, such chips can be obtained free of charge in any quantity. Why get them, you ask? The fact is that I have already managed to use the chip many times. In fact, it is eternal: to rinse it well enough - and you can apply it again and again. In terms of accuracy, it will not differ from the new. My very idea was to make a device for this conditionally free chip.

So, I faced the challenge of reverse engineering of the chip. Of course, it was impossible to find any documentation for the cherished microcircuit - the manufacturer was not going to share production secrets but wanted to quietly sell their devices for $ 50,000. To begin with, I did the most obvious and simple thing: I rang the contacts with a tester. It became clear where the digital and analog inputs-outputs, power, and so on. Some information was obtained from the patents on the chip. But all this, of course, was not enough to create a complete product. I still fiddled with the chip, checked my various guesses, experimented with signals, but I didn’t get anywhere in principle. I had to pause the project.

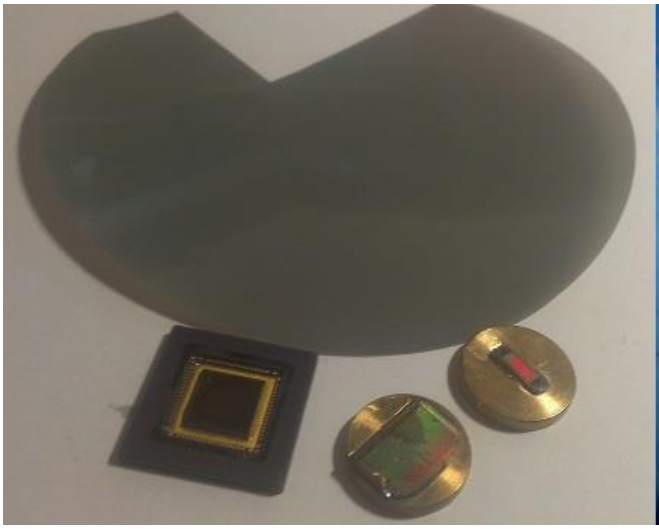
And then suddenly on Habrahabr I came across an article by the famous blogger BarsMonster about how he does reverse-engineering of chips! I was inspired, wrote to him, wrote to other enthusiasts, contacted with Monocristal with a hope to use their technology, they said that they could not polish in layers, they could only capture the top layer, and since my chip is multi-layered, it will not be clear where the tracks from the contacts go. Then I met one guy who also deals with reverse engineering of chips, sent him this chip, but even then, the matter didn't go further than photographing the upper layer. Then I came across an article on the Internet about those who were able to reverse the Sony PlayStation chip, etc. I decided to write them with questions, found their nicknames - and immediately realized that one of them was familiar to me. Recently, a friend brought me to a friend who "is also engaged in genetics at the amateur level," we talked to this friend on Skype and ended the dialogue. And now I understand that my new friend is a mega-cool master of reverse-engineering chips.

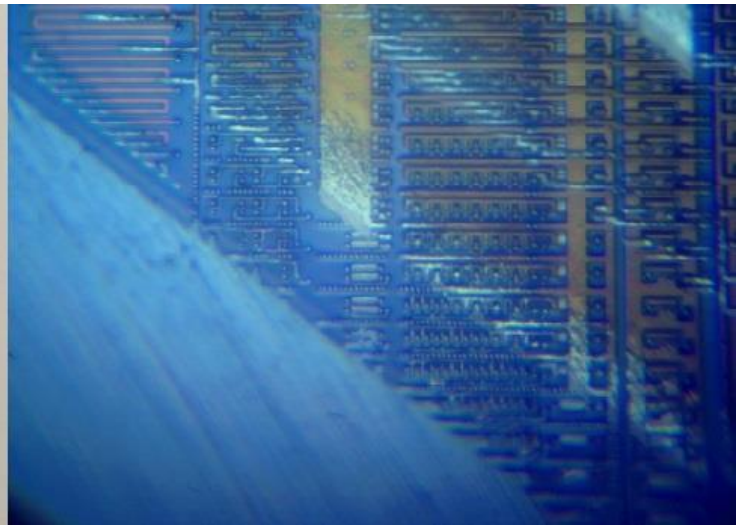
Then I still had to master the technique of polishing myself. The difficulty of polishing is to remove the metal layers with a thickness of about 1 micron - and the chip width is 1 centimeter. For comparison, I will say that this is about the same as allowing a 1 km error of no more than 10 cm. I tried very hard. For this I made the setup for flatness control.





The results of my work are presented in the following photo:

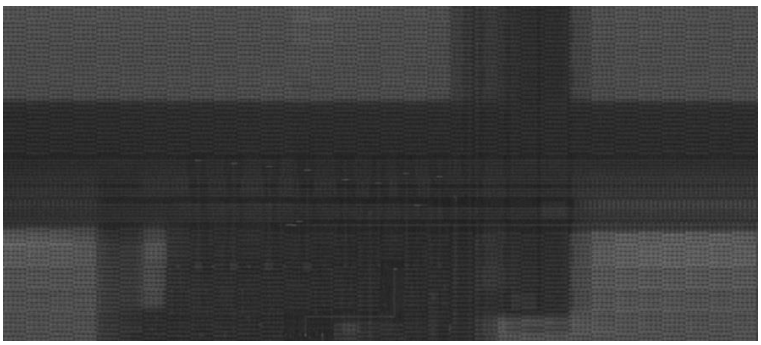


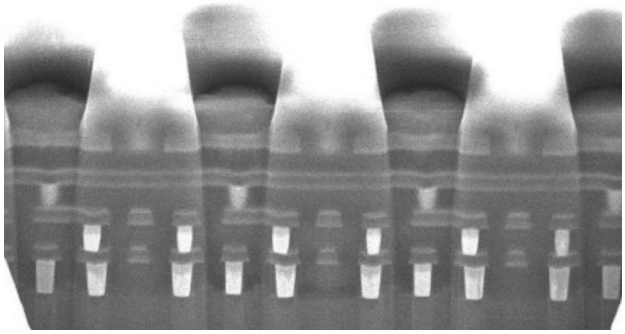


The lower silicon layer, the top layer with transistors, the first, second, third, and fourth layers of metal are quite clearly visible.

The chip consists of repeating zones (such as shift registers), and from such pictures it was very convenient to analyze it: it immediately became clear what was happening on different layers. I "reversed" the most "stuffed" areas with an abundance of logic that were repeated many times. But it turned out to be the most difficult to track the tracks running across the chip, to understand which external contact belongs to what. "Photosession" crumbs in 1 square. cm was a 50 GB of black and white photographs.

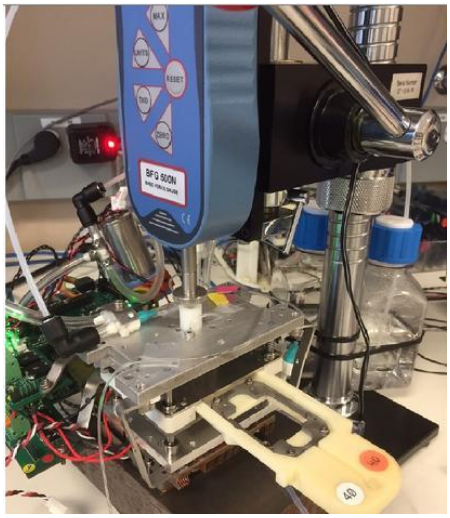
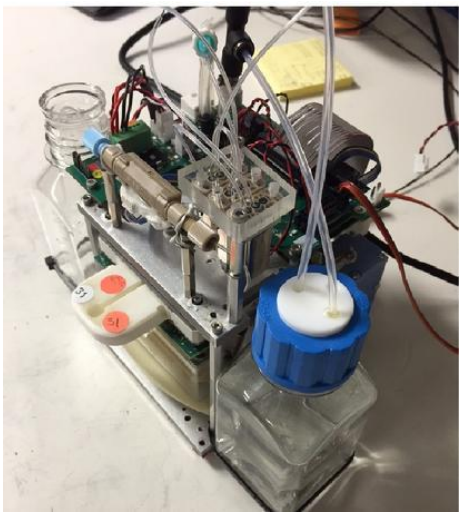
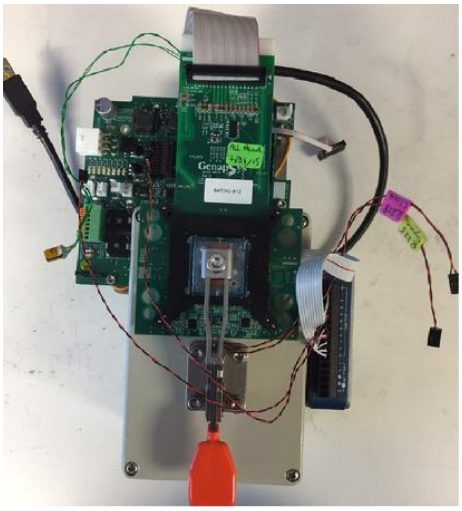
Now all these separate photos needed to be somehow united into one whole picture. Almost the same day I wrote a program on the Python that generated the HTML file — when I opened it in the browser, I got what I needed. Then I wrote another program on javascript that allows me to compare layers, smoothly move between them, align them, adjust the scale, etc. Finally, I had all the tools for solving the main tasks. I traced the chip penetrating tracks and restored its entire structure to the last transistor.



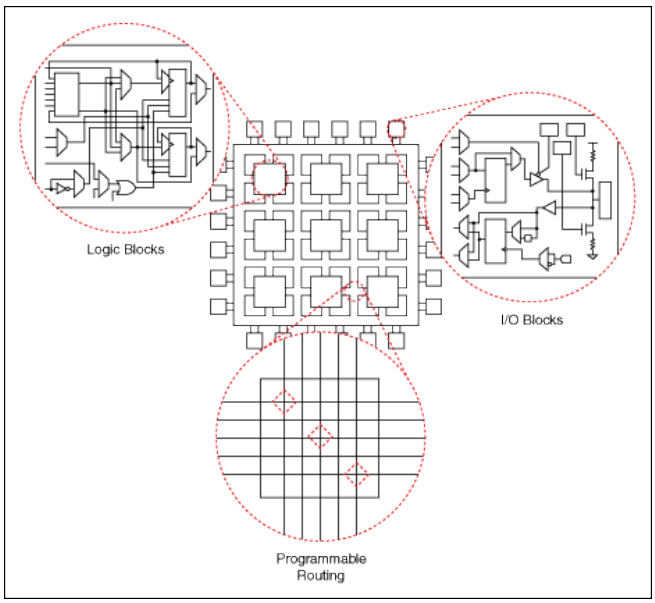


The holes are clearly visible where the spheres with the reads fall. Below are three layers of metal, and even lower - a layer with transistors.

The next was the creation of a motherboard chip. I used GenapSys FPGA board. (FPGA is, roughly speaking, an array of 10 000 universal logic elements; by programming FPGA, we can get any logic that easily processes gigabit information flows.) I wrote the firmware for FPGA myself, and besides, for dynamic system management I wrote software, which sets the entire configuration for FPGA.

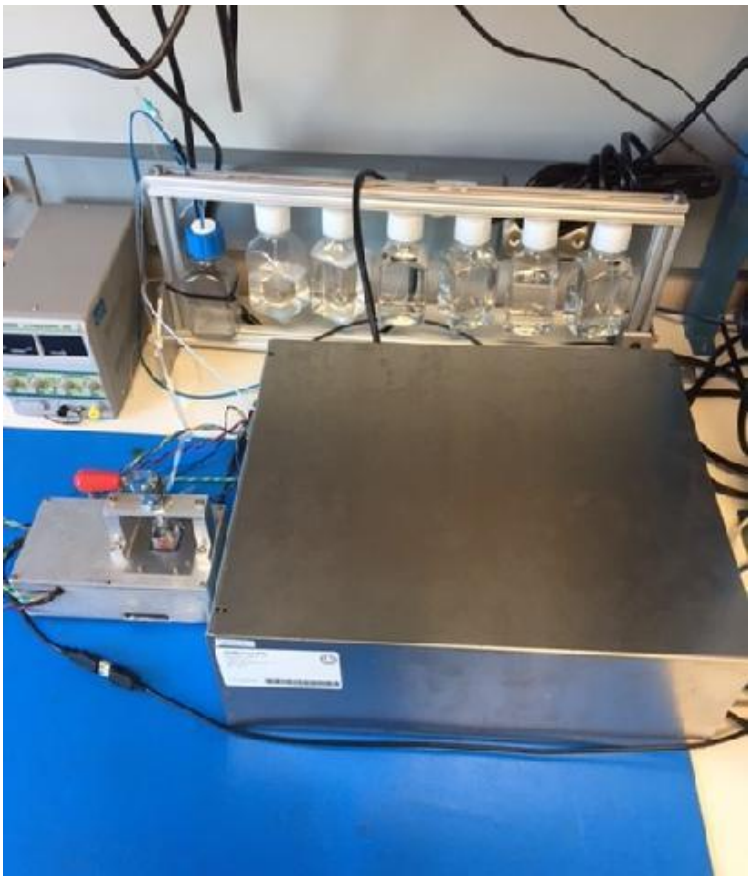
FPGA resources are grouped in slices to create configurable logic blocks. A slice contains a set number of LUTs, flip-flops and multiplexers. A LUT is a collection of logic gates hard-wired on the FPGA

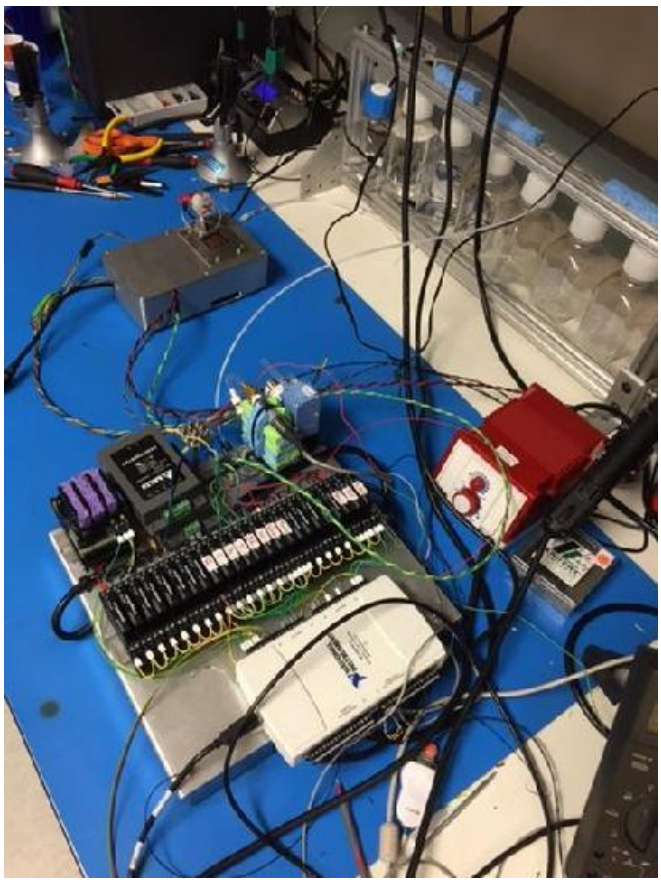


.

http://www.ni.com/pdf/manuals/374546a.pdf

Now I needed to digitize this picture and transfer it to a computer. I put together such an installation:



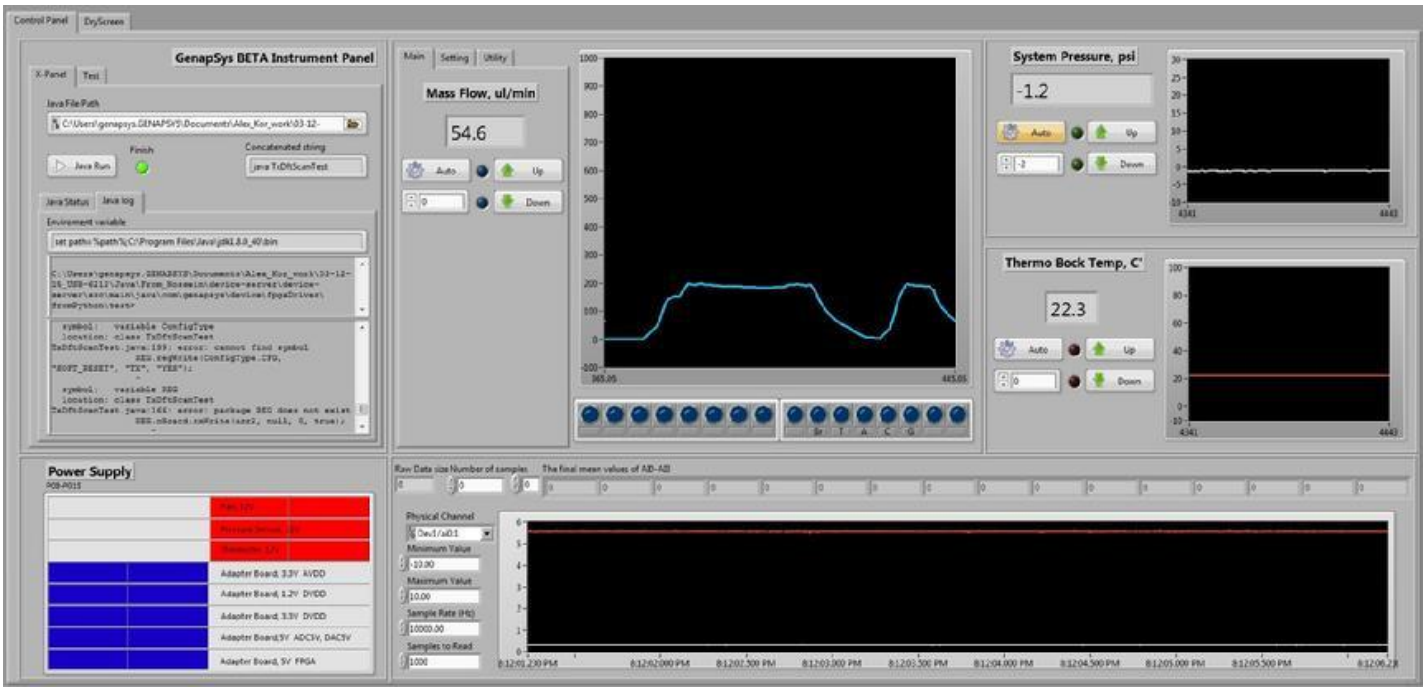


The excitation signal generated by DAQ module U6212 with sample rate of 400 kS/s and timing resolution 50 ns. Signal with timing accuracy of 50 ppm of sample rate applied to the object through Aout directly to Rx signal points of the chip. The programmed resistor Rl is used at the output of the amplifier A1 in order to limit the maximum value of current through the object. Response signal Ux are converted to signals U1(t) and U2(t) and applied to ADCs in the DAQ module through Ain1 and Ain2. Current Ix(t) is converted to voltage U1(t) by currentto-voltage converter (CVC) realized with amplifier A3. Range of the CVC can be changed with resistor Rr in order to fit the measured signal U1(f) to the range of the ADC.

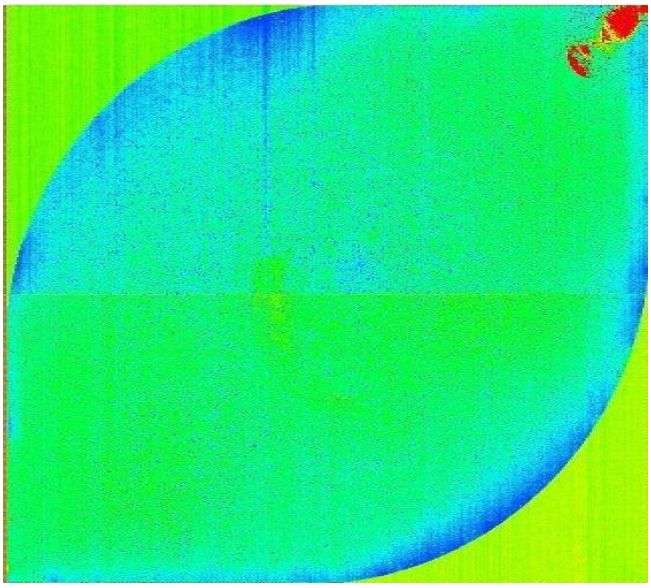
Test Bench device developed as a sequencing instrument that collected data with implemented method of dielectric spectroscopy also known as impedance spectroscopy, and also known as electrochemical impedance spectroscopy, measures the dielectric properties of a medium as a function of frequency. It is based on the interaction of an external field with the electric dipole moment of the sample, often expressed by permittivity. In BETA case it is an experimental method of characterizing electrochemical systems. This technique measures the impedance of a system over a range of frequencies, and therefore the frequency response of the system, including the energy storage and dissipation properties, is revealed.

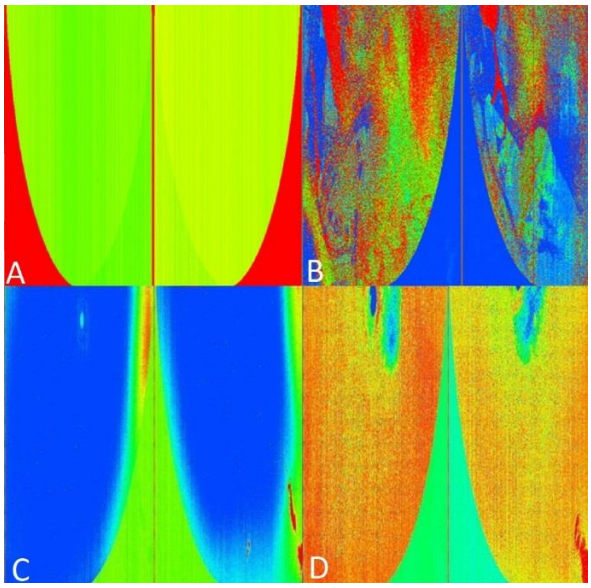
There is a computer that provides control data to the board with FPGA. The board generates digital signals and sends them to the chip. The signal from the chip goes to the amplifier, then - to the ADC on the board, digitized and transmitted via the COM port to the computer. In general, the capacity of the COM port is small: 15 kilobits per second (because there is from 1 million to 10 million "pixels" in one chip, and the maximum transmission rate is 115200 baud). Nevertheless, the picture on the computer eventually falls.

I made a combination of the software for control and acquisition in one GUI:



When the DNA library is fed to the used b / w chip, the chip is filled unevenly: at the edges - to a lesser extent. Different colors are due to different voltages on the pH transistors. That is, we can clearly distinguish those wells into which the spheres with reads are located — subsequently this will help us to control the flushing of the chip.





Accordingly, the next task was flushing the chip. It was necessary to ensure that he became like new. Fortunately, I had a completely new chip as a reference sample. In the picture A it can be seen that in the active area such a chip is almost the same color (vertical repeating stripes are just noises, pickups).

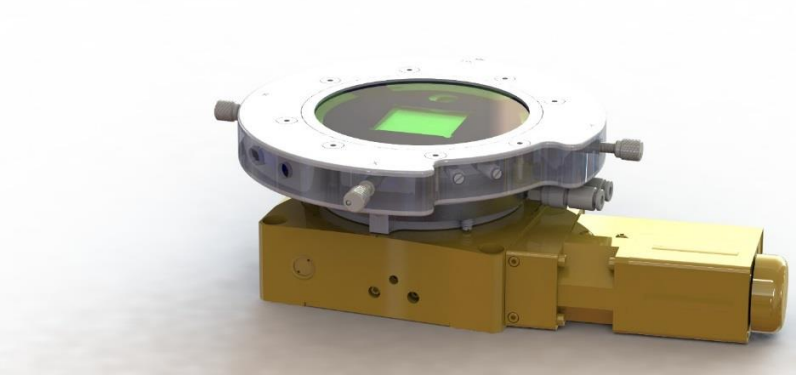
In a picture B, an unsuccessfully washed chip - it is multi-colored. In the picture D - used, but well washed chip. It is seen that the gradient at the edges has disappeared. Nevertheless, it would be worthwhile to prove that it is really clean and can be reused.

I think I need to tell the Lawyers that may be searching for Proprietary Law violation: No physical resources, land, non-human creatures or intellectual property were used for my research. GenapSys was concentrated on the method of Impedance Spectroscopy, so it was not interested in the development of my method, all my work was independent with not using of the proprietary information of GenapSys and all conducted during my personal time.

Since the DNA libraries are attached to the tantalum chip coating in an acidic medium and are detached in an alkaline (i.e., at high pH), the chip is washed with special semi-automatic pipettes with solutions with different pH. Today I managed to achieve almost complete chip cleaning.

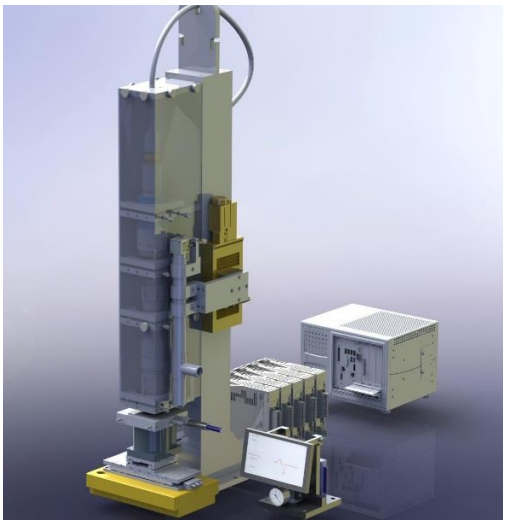
I was asked why, when I fully understood the structure of the chip, I did not order to make it, but preferred to continue to search and retrieve the used ones, to mess around with their flushing, and so on. Yes, because it costs a lot of money, millions of dollars, and a substantial part of this amount goes to physically debugging the resulting product: fitting, tuning all the parameters of transistors, etc. That is, just copying the logic circuit is not enough. Therefore, I take a conditionally free, ready-made - designed, manufactured, debugged - microcircuit and thus save significant funds, seriously reduce the cost of the project.

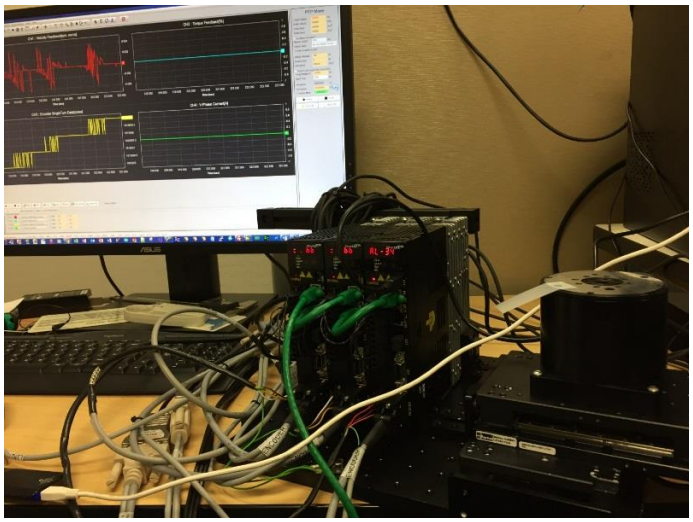
The next task was to create the whole structure with the nest and the heating-cooling system. My experience of several years in GenapSys as a Test and Fabrication Engineer gave me some clues how to make the efficient holder for the chip with all necessary temperature, pressure and adjustment parameters:



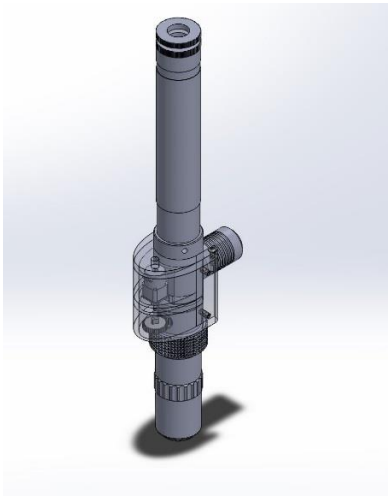
For instant heating and cooling I am using the IR and UV Laser with combination of peltier. The pressure that is necessary to hold the ions consistent is about 80psi, so the nest structure and the optic flat must be efficiently stable. The scanning is synchronized with physical chip moving. The XY stages realized on Newport stages with adjusted Nano resolution.

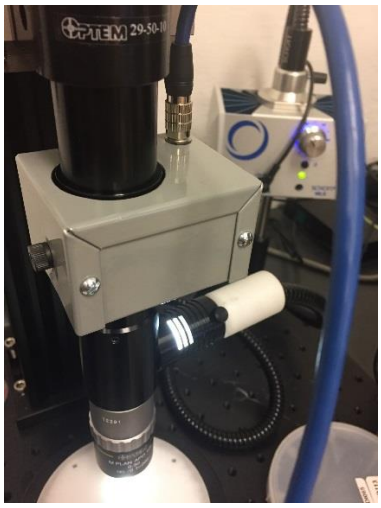
I was truing the Parker Hannagan stages and created the project for it; however, the speed of the encoder transferring is limited by Scan Engine function that is 500MHz and it turned out to be slow for my task.





For faster acquisition for control I made the Auto Zoom microscope.



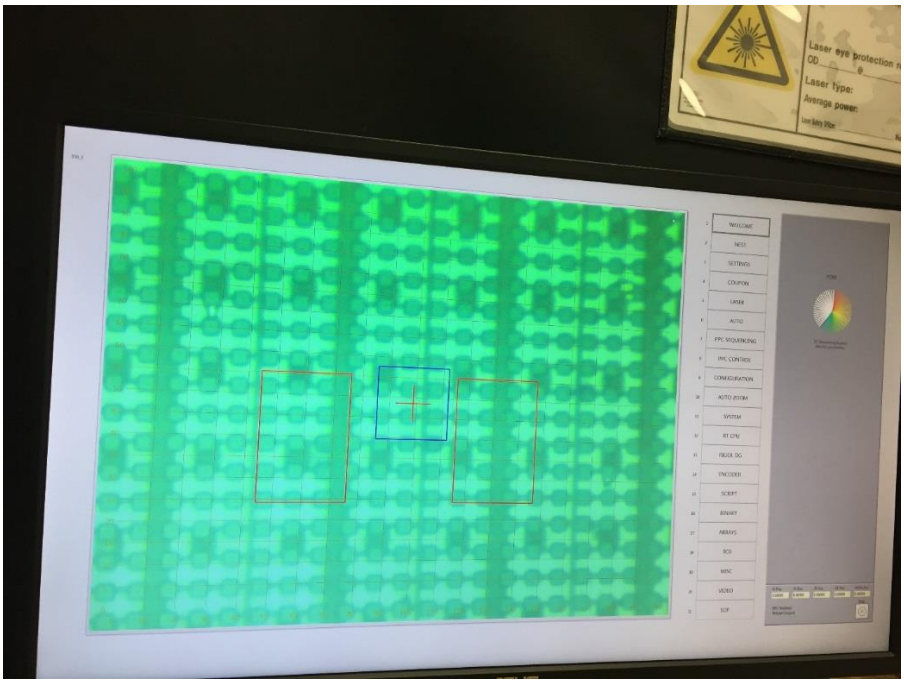


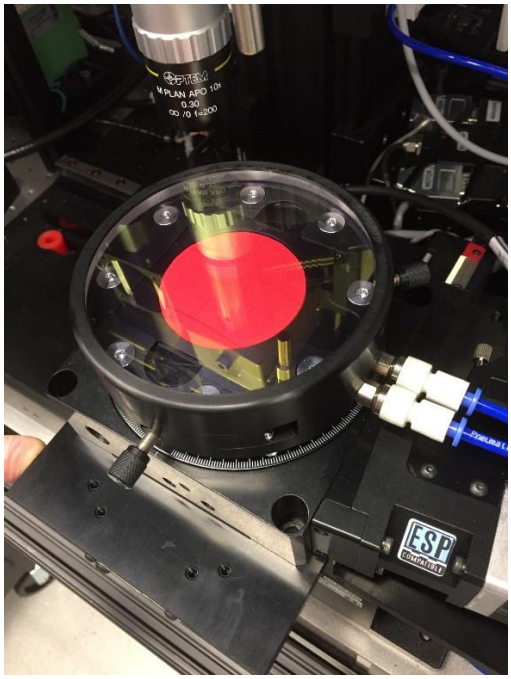
Of course, that the final design is always different and usually much worse than the original idea 😊 but the functionally, the Auto Adjustment Focus is working perfect and saving a lot of time for chip positioning.

The next task was to increase the speed of the sequencing. The regular PC with any operating system was not compatible with this task. So it was necessary to transfer the software to the independent board with own firmware. For this purpose, I user the NI PXIe-1072 chassis, with

the latest controller PXIe-8880. This allowed me to use the digitizer PXIe-6545 with up to 250MHz speed of triggering and PXIe-7976 on Kintex-7 410T with 136, configurable as 136 single-ended, and 68 Differential general-purpose channels for working with the chip. (FPGA is, roughly speaking, an array of 63550 universal logic elements; by programming FPGA, we can get any logic that easily processes gigabit information flows.) I wrote the firmware for FPGA myself, and besides, for dynamic system management I wrote software, which sets the entire configuration for FPGA. For peltier, lasers, Newport stages control, pressure Festo controller gage, and other infrastructure I used PXIe-6361 controller.

The GUI of my DNA decoder and setup:





The newest device that is going to the market is S5 sequencer.

According to manufacturer:

> Speed
>
> Some research decisions can’t wait. A few hours could make all the difference in your quest for the right answer. With other light-based sequencers you could wait almost 60 hours for results. With the Ion S5 System you get results in as little as 24 hours (with a 2.5-4 hour sequencer run time and overnight Ion Chef System run).

https://www.thermofisher.com/order/catalog/product/A27212

Very good result, but still is not enough.

This device is still far away from the RT processing due configuration of the architecture and cannot be used for the creation of the parallel synthesis of the transformation of the genome, as my DNA decoder does.

Thanks for your attention!