EmbeNewsMaker 27 July 2021 at 13:47

The way we made an external PCIe RAM disk based on the DDR memory

High performance *, Cloud computing *, Data storage *, Computer hardware , Data storaging

RAM disk, this is a disk based on RAM memory chips. This kind of disk is not able to retain data after the power is turned off (unless a supporting battery is used), but has an exceptionally high read/write speed (especially for random access) and an unlimited lifespan. It is important in tasks that need a lot of cycles to write over information, even professional SSD drives don't live long. To the operating system the RAM disk is indistinguishable from an SSD or HDD disk and no special drivers or setup is required. Unlike a disk that is virtually located in the computer's RAM memory, where the maximum memory capacity is limited to 128-256 GB in the best consumer motherboards, a RAM disk for a PCIe slot, in general, has no volume limits and can work in any MB with a PCIe slot.

This project has started several year ago as part of a DC program (*distributed computing). With time, the internet speed has grown and the major problem of DC speed has been the low speed of HDDs. SSDs partially addressed this issue, but the higher wear of SSDs due to continuous job overwrites drove away new DC network users. Nobody likes to ruin their expensive SSD for a noetic goal. However, resource and fast write/read speeds with consistent lag are very critical for DC computing, where delay or failure in each segment slows down the whole compute network.

At that moment, our team generated an imaging of the future RAM disk using refan DDR memory. The thing is that such memory is cheap enough – about 40 cents per gigabyte for DDR3 – and readily accessible in big quantities now, during the silicon crisis. Such memory manufacturers as Kllisre, ZIFEI, TANBASSH, Rasalas etc. use the refan chips. These are also mounted in most budget Chinese laptops and tablets. The refan chips can bring the cost of the end device to a reasonable level without any loss of features or reliability.

The second evident condition is that the future drive works in the same way as HDD or SSD drive and is completely compatible with the OS and software products. That is why the Silicon Motion SM2262EN chip was used. The chip supports PCIe x4, NVME 1.3, native DDR memory for clutter buffering and 8 channels of NAND/TLC memory. This allows it to use all available PCIe x4 speeds with no loss of speed.

Using memory slots would require more board area. The number of passive components required for impedance matching would increase. And of course the cost would have increased many times over.

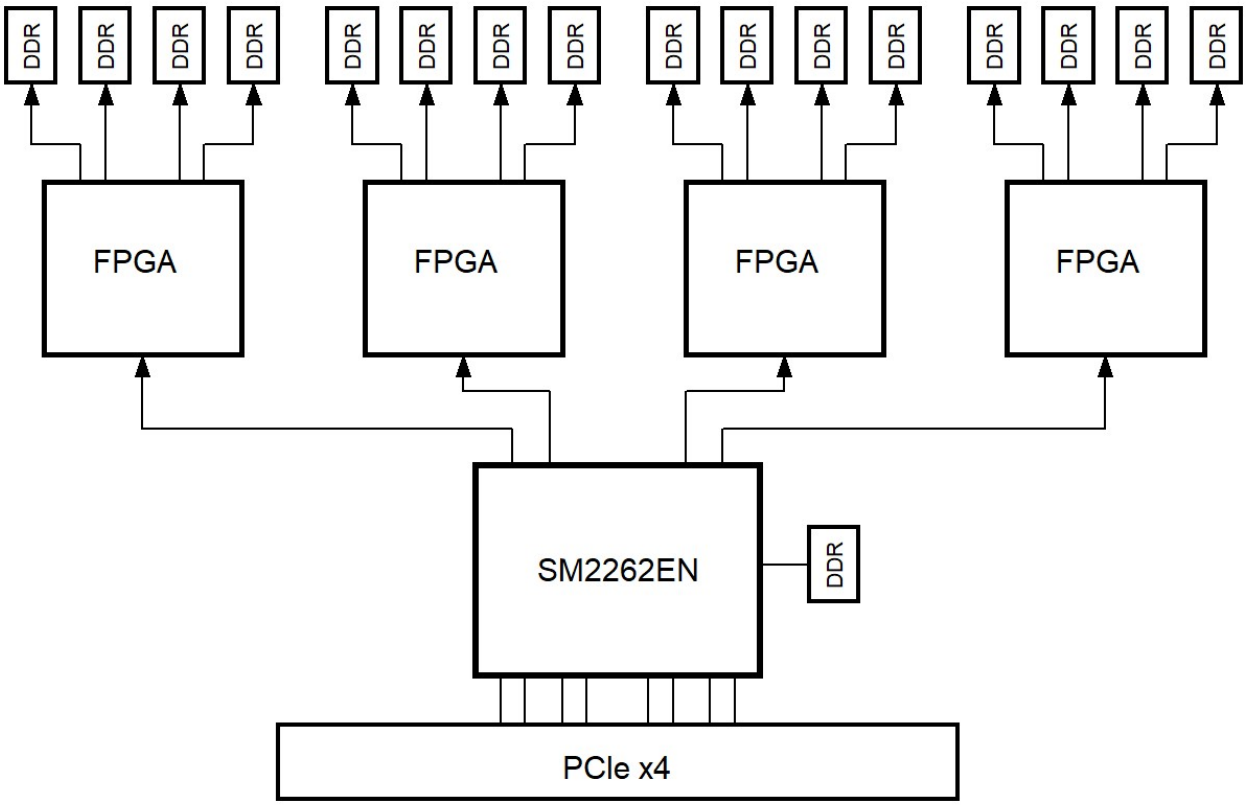
The first problem lurked in the fact that the interfaces of NAND and DDR memory are very

much different and just plugging in DDR instead of NAND would not work.

The second problem was the length of the wires between the DDR chip and the SM2262EN cannot be infinite. And to build up to 1 TB we would need 256 DDR memory chips of 4GB each.

An FPGA as an intermediate between SM2262EN and DDR was called for, but **the third problem** was that a hard memory controller of the FPGA cannot address that many DDR memory chips.

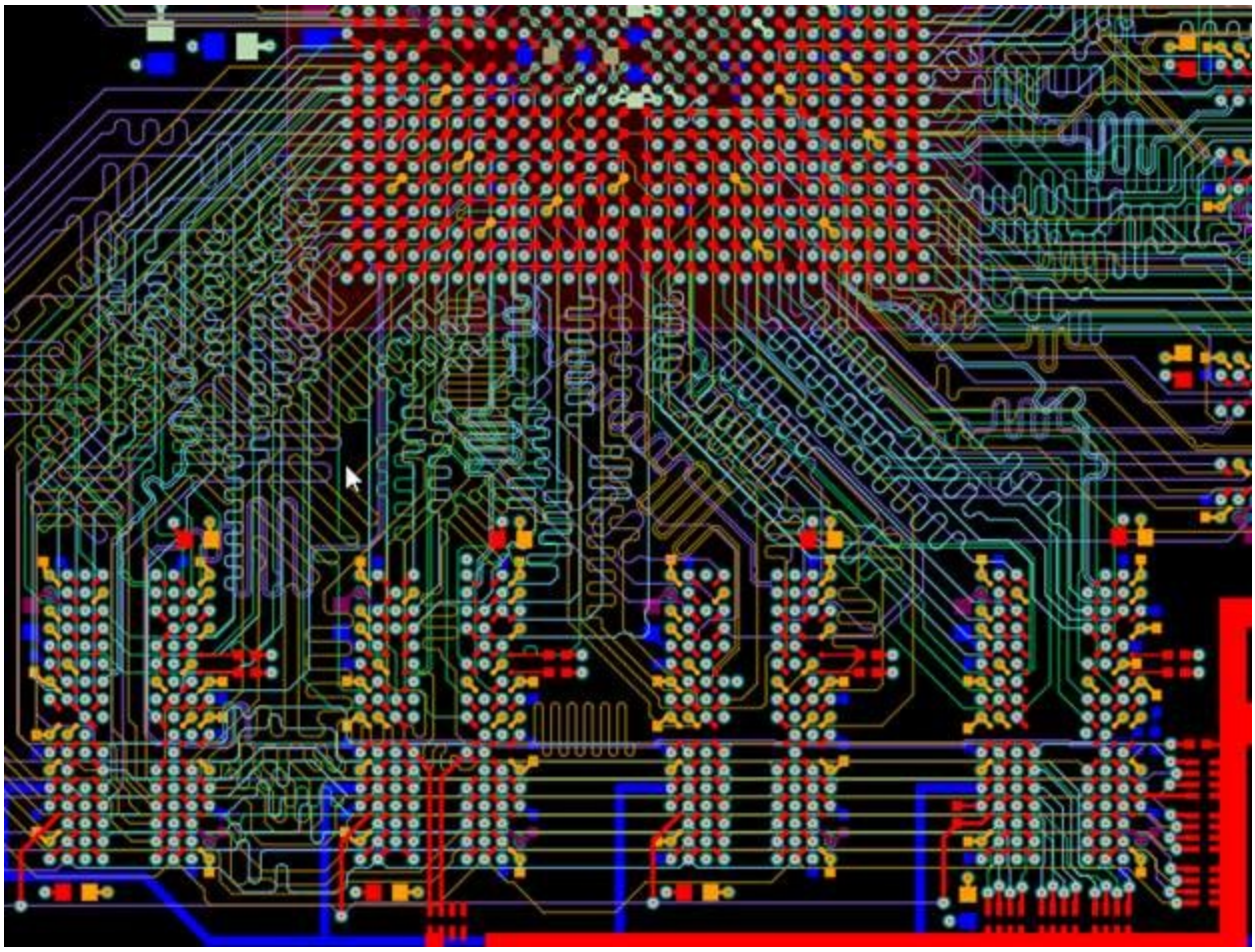
I would not waste your time on the background of our quest, but in result we came to an architectural solution with four FPGAs evenly spread over the card's PCB, each of them connected to two banks of 32 DDR chips and programmatically emulated the work of two NAND chips each.



Disk architecture

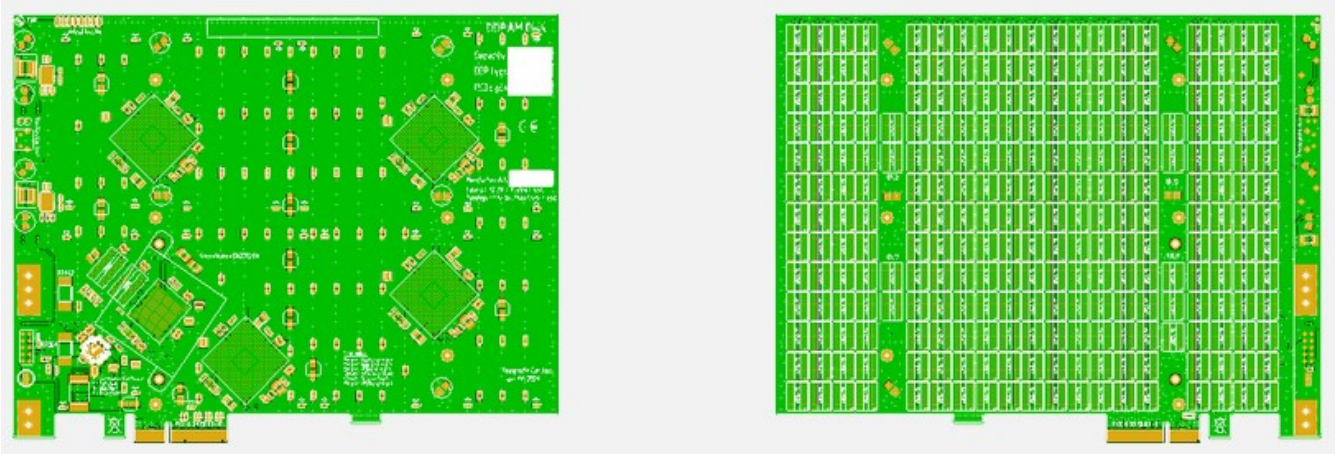
This solution ensures that the SM2262EN does not detect a trick and works with the memory the same way it does with NAND chips. The lengths of the wires separating the FPGA and the furthest DDR chip in the bank are within tolerance and their impedances can be matched using resistors and DCI (*Digitally Controlled Impedance) technology. The larger 12-layer board area and abundant free space has allowed us to reduce a lot of passive components using interlayer capacitance and the own wires resistance. This solution was carefully tested

in HyperLynx and then confirmed to work in the prototype. The example of this tracing is shown in Fig below.

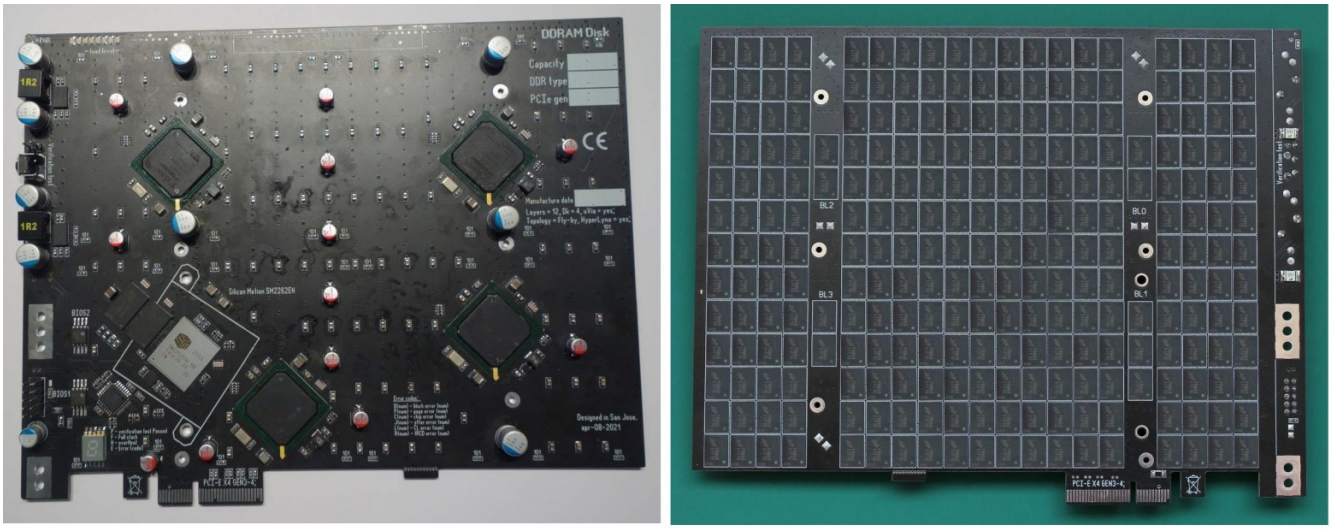


An example of DDR tracing

Additionally, the card is fitted with a Power Management Module (PIMC) and a boot system with two BIOS memory chips which contain bitstreams and configuration information. Two of these chips were installed for reliability reasons. The card also includes a load light column and a screen that displays the status of the drive’s operation, self-test codes, and errors. This small helper lets you quickly troubleshoot and find out problems. The below pictures show the PCB project and manufactured PCBs.

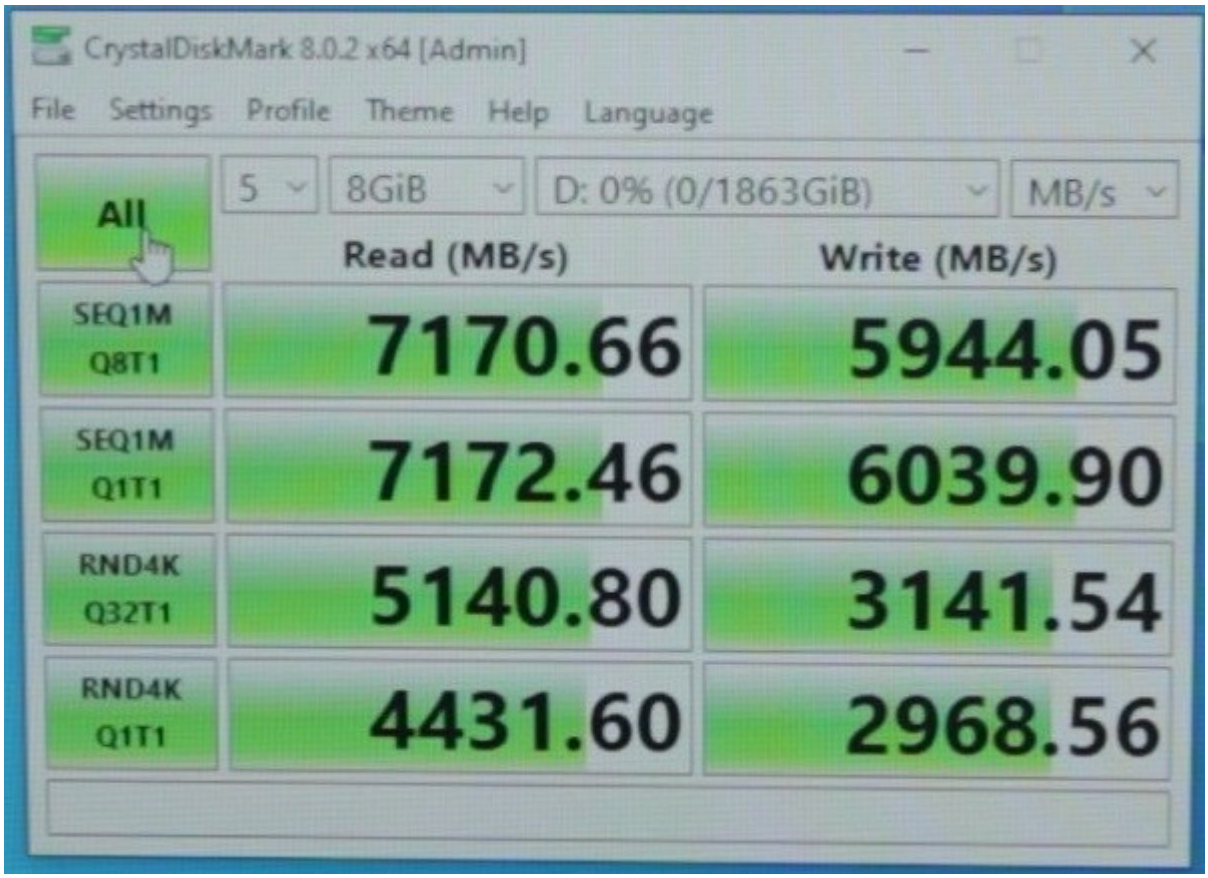


PCB project



Manufactured boards

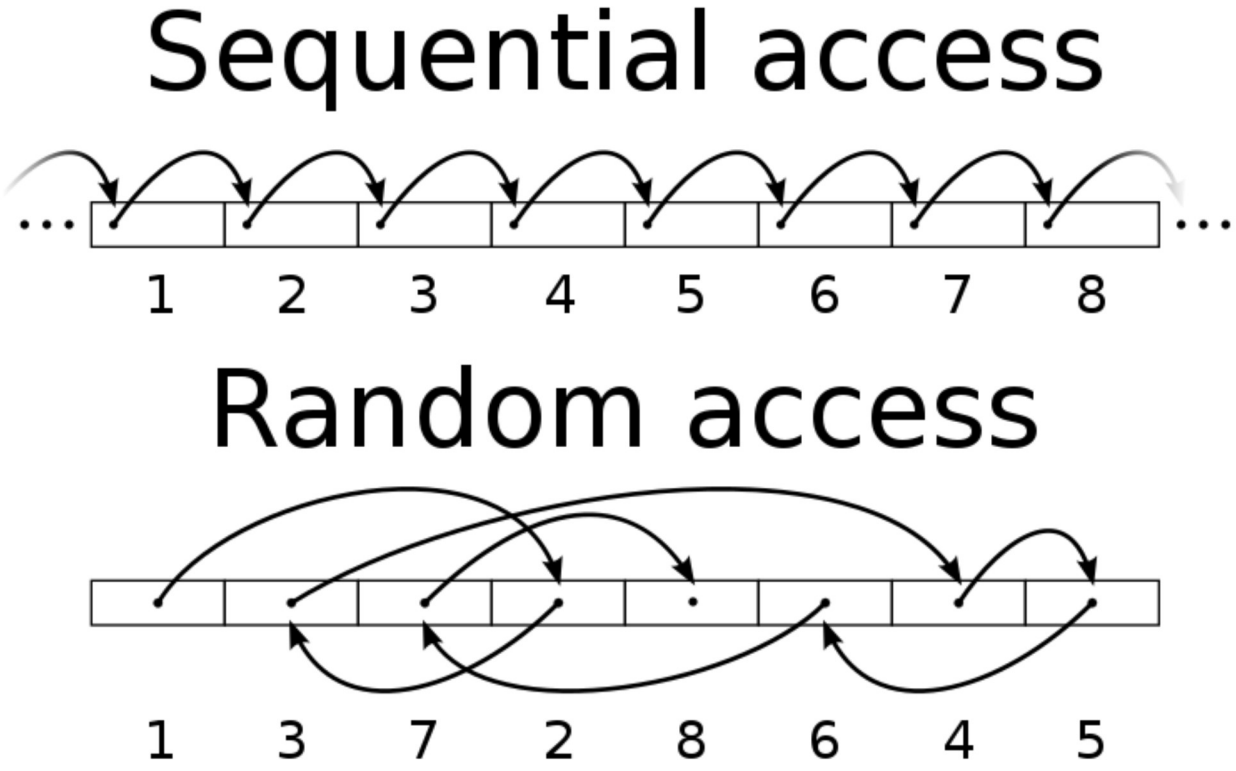
The disk test showed read/write speeds of 7000/6000 MB/s respectively, which is the limit for PCIeX4 gen 4 and approaching the speed limit of DDR3 memory (see the screenshot from this video).



Speed tests results

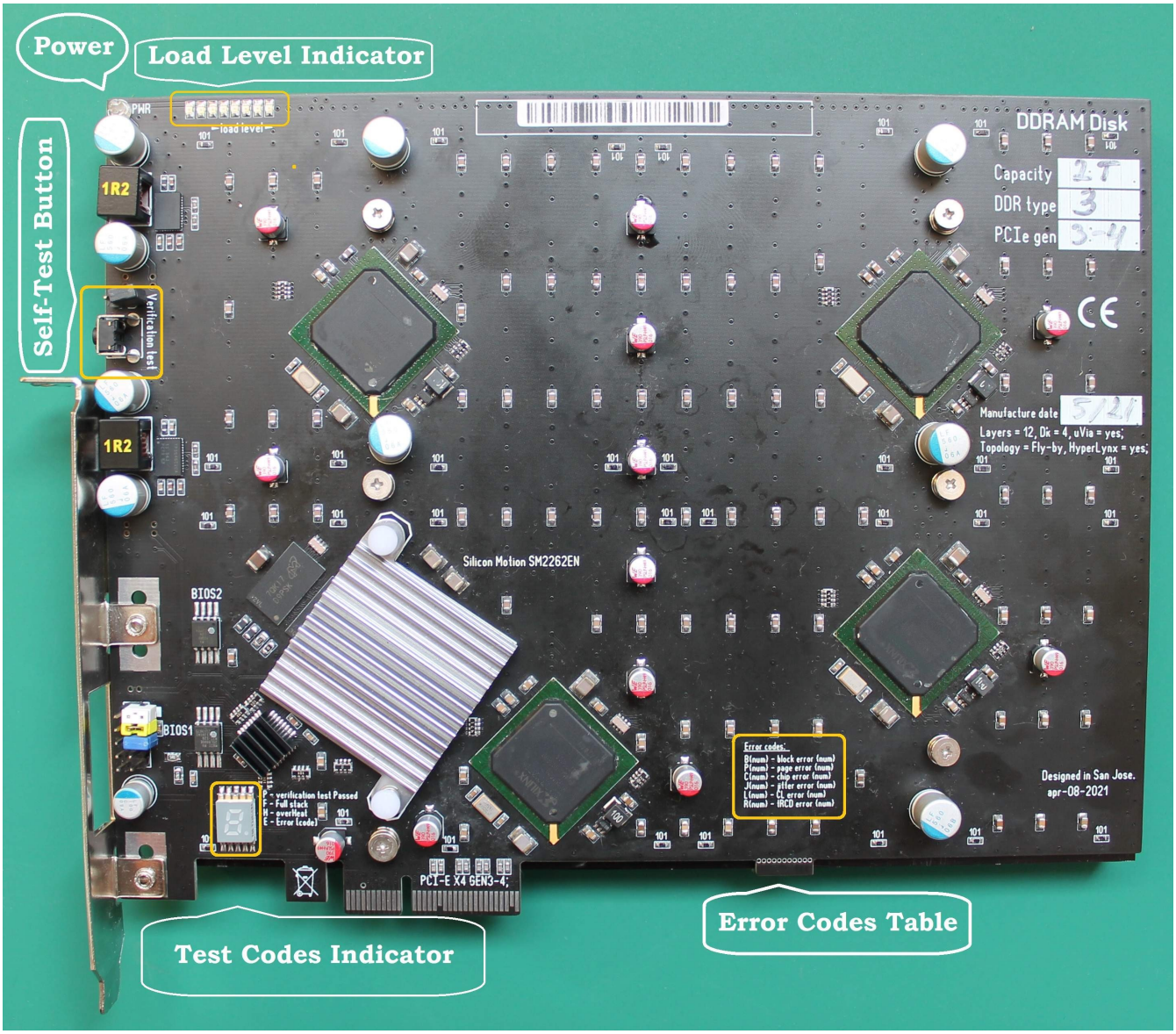
Especially remarkable are the results of the RND4K tests (Q1T1 may be inaccurate due to the applied transport layer) – such a speed is reached due to specifics of the DDR chip bank,

representing a grape cluster of small-volume media, which allows working with them in parallel and simultaneously, as opposed to one single NAND chip. We know of no SSD drives that could show a relative speed in such tests.



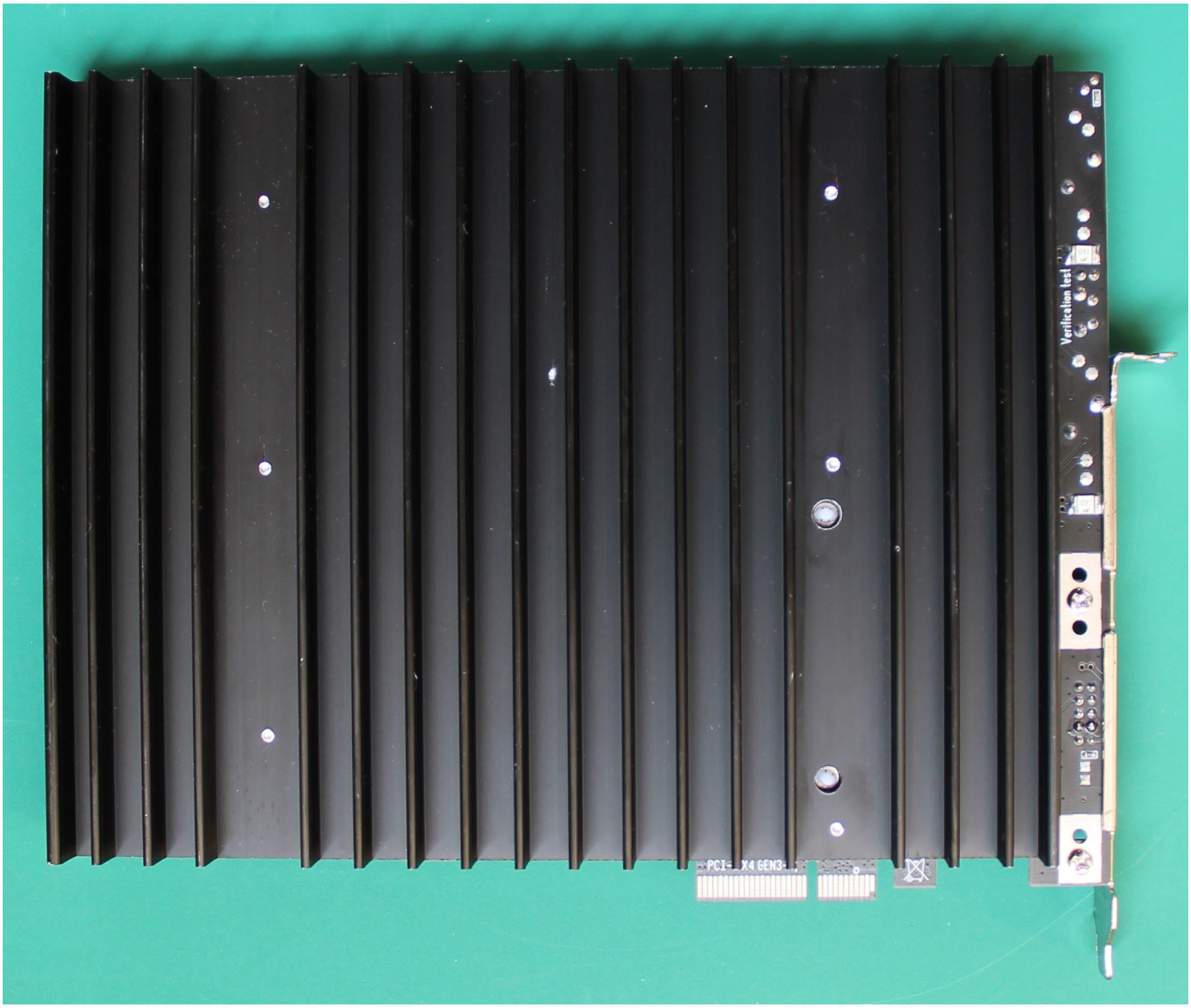
Types of tests

The below picture shows the basic controls of the DDRAM disk. These are the load light column, back-up battery connector, the display and the self-test button. After pressing this key, the drive shuts off all PCIe exchange and starts the test of all of its components, including memory chips. The test ***will destroy all the data on the drive*** regardless of back-up battery availability. The indicator shows the memory bank number of the memory being measured during the test. When the test passes, the light will show P = passed. If not, a fault code will be shown. For the user's comfort, an error code interpretation table is painted on the drive's PCB at the right side of the indicator. The indicator will flash the decimal point at a rate of 2 Hz during operation. This is a sign of normal operation of all drive systems.



Top side and main controls

On the back side of the card is a DDR memory chips plantation and a heatsink. The heatsink is needed, because the chips are installed very tight and from below are heated additionally by FPGA chips and other device elements. Not only does the radiator removes excess heat, but also equalizes the temperature of the different areas of the disk card, that eliminates the skewed characteristics of DDR chips because of uneven heating. A dry thermal interface based on 3M thermal pad is used.



Reverse side with heatsink

The disk has a ledge on the board, leaning against the connector edge, to ease the mechanical stress on PCIe slot and to simplify mounting the disk to board. In case of need there is a possibility to break this ledge, for this a line of holes is provided which contains the breaking line. But we have not yet met a motherboard where it gets in the way.

During the resource tests the disk ran for 200 hours continuously under maximum load and alternating read/write operations. No changes in drive characteristics were noted.

WBR, DDRamDisk team.

Tags: RAM, SSD, HDD, ram disk, hight speed, storage, storage systems, storage spaces, mining, mining rig

Hubs: High performance, Cloud computing, Data storage, Computer hardware, Data storaging