ARTICAL

WORKING OF SSD FLASH MEMORY

KARTIK KALAL

@kartik-kalal

MOTIVATED BY : BRANCH EDUCATION

INTRODUCTION

We do wonder how our smart devices can store countless images, audio, or films. Or, we do wonder when we download a song to our laptop, where does it actually get stored?

In this article, I will open up a smartphone and describe the microscopic view of memory storage microchips. This microchip is also used to store data in solid-state drives, or SSDs in your computer, in tablets, and inside flash drives. The technology is called V-NAND which is incredible how engineers were able to fit such an insane amount of memory storage capacity in such a small space.

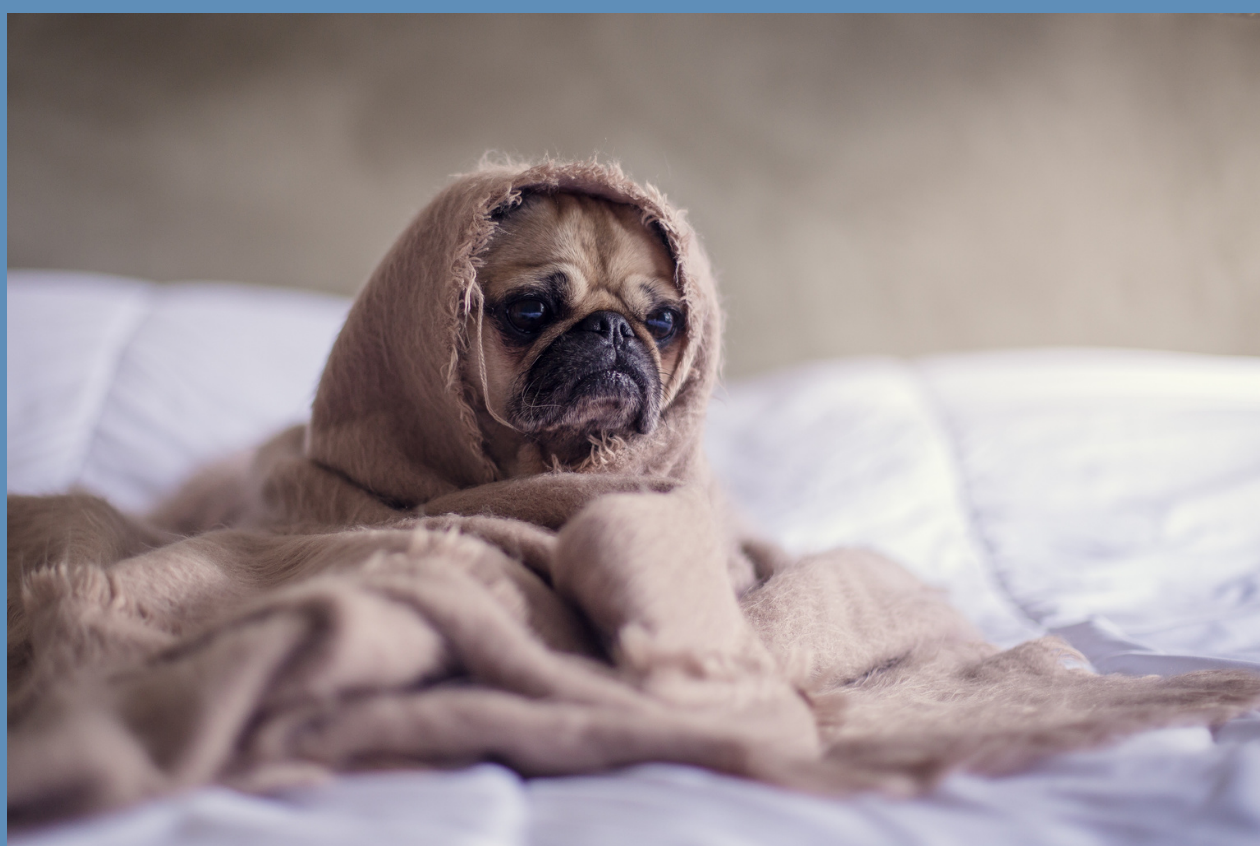
But it might not seem so surprising when you see
the complexity inside your smartphone,
or the inside of this one terabyte solid state drive
commonly found in laptops or computers.

First, this picture is made up of pixels and each pixel has a color so let's zoom in so that we can see the individual pixels.

The color of every pixel is defined by a combination of 3 numbers, ranging from 0 to 255, each representing red, green, or blue.

Each of these 3 numbers from 0 to 255 is represented by 8 bits in binary, or eight ones and zeros ya know, because computers work in binary.

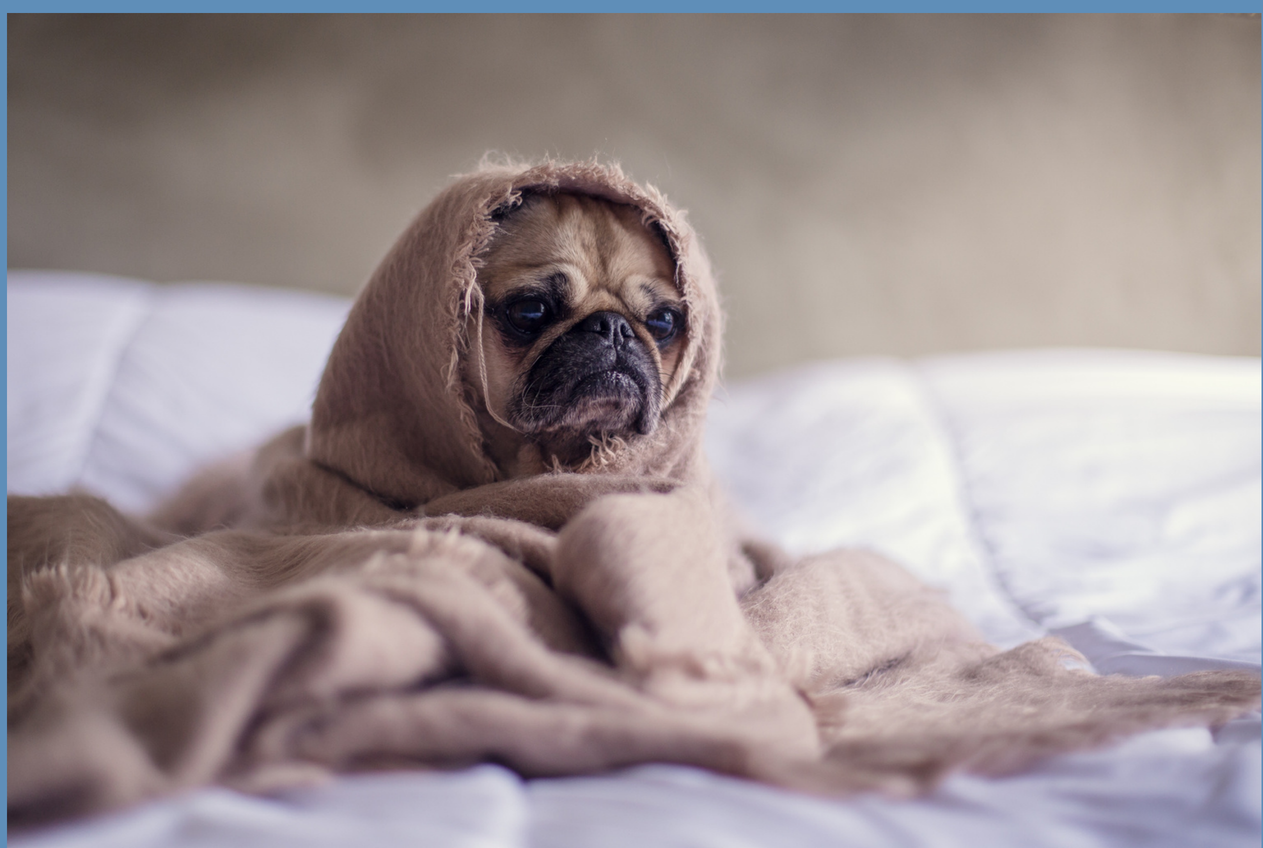
So, 3 colors, red, green and blue, and 8 bits each, means each pixel takes 24 bits to define its color.



First, we see the picture, which is 3024 pixels wide and 4032 pixels tall, which is a total of around 12 million pixels, or, 12 megapixels. Which relates to the resolution of the 12 megapixel camera on a smartphone.

Next, by doing some multiplication
we calculate that an array of this size,
where each pixel is defined by 24 bits,
or 24 0s or 1s only requires 293 million bits
or a unique set of 293 million 0s or 1s.

That's a ton of bits, so let's figure out how your
smartphone or this solid-state drive seamlessly
stores every single one of them.

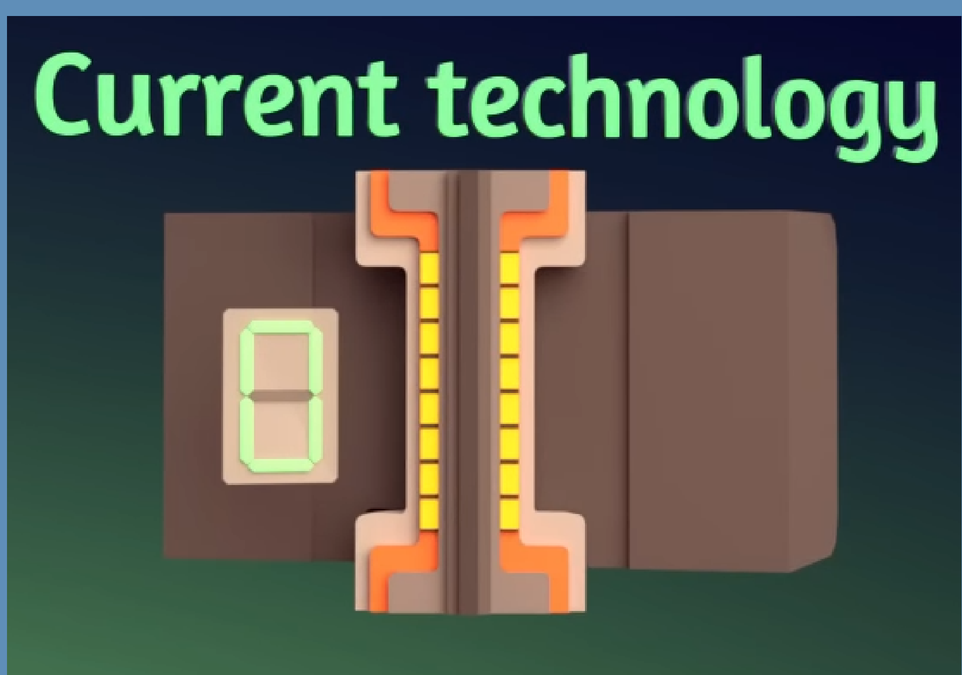


Charge Trap Flash Memory

This is the basic unit of a computer's long term memory storage and it's called Charge Trap Flash Memory-
So how does it work?

Well, in each cell we can store information by placing different levels of electrons onto a charge trap, which is the key component inside the memory cell. Older technology could only store two different levels of electrons, a lot of electrons or very few electrons, which were used to store a single bit as a 1 or a 0.

Most memory cells in 2020 can hold 8 different levels, but newer technology can have 16 different levels of electrons. This means that a single cell, instead of holding only one bit as a lot of electrons or no electrons, can now hold 3 or more bits but, for this example, let's stick with 3 bits.



There are 8 different levels for all the various amounts of electron charges that our charge trap can be set or written to.The key to the charge trap is that it is specially designed so that after it gets charged with electrons, it can hold onto those electrons for decades, which is how information is saved or written to the solid-state drive.

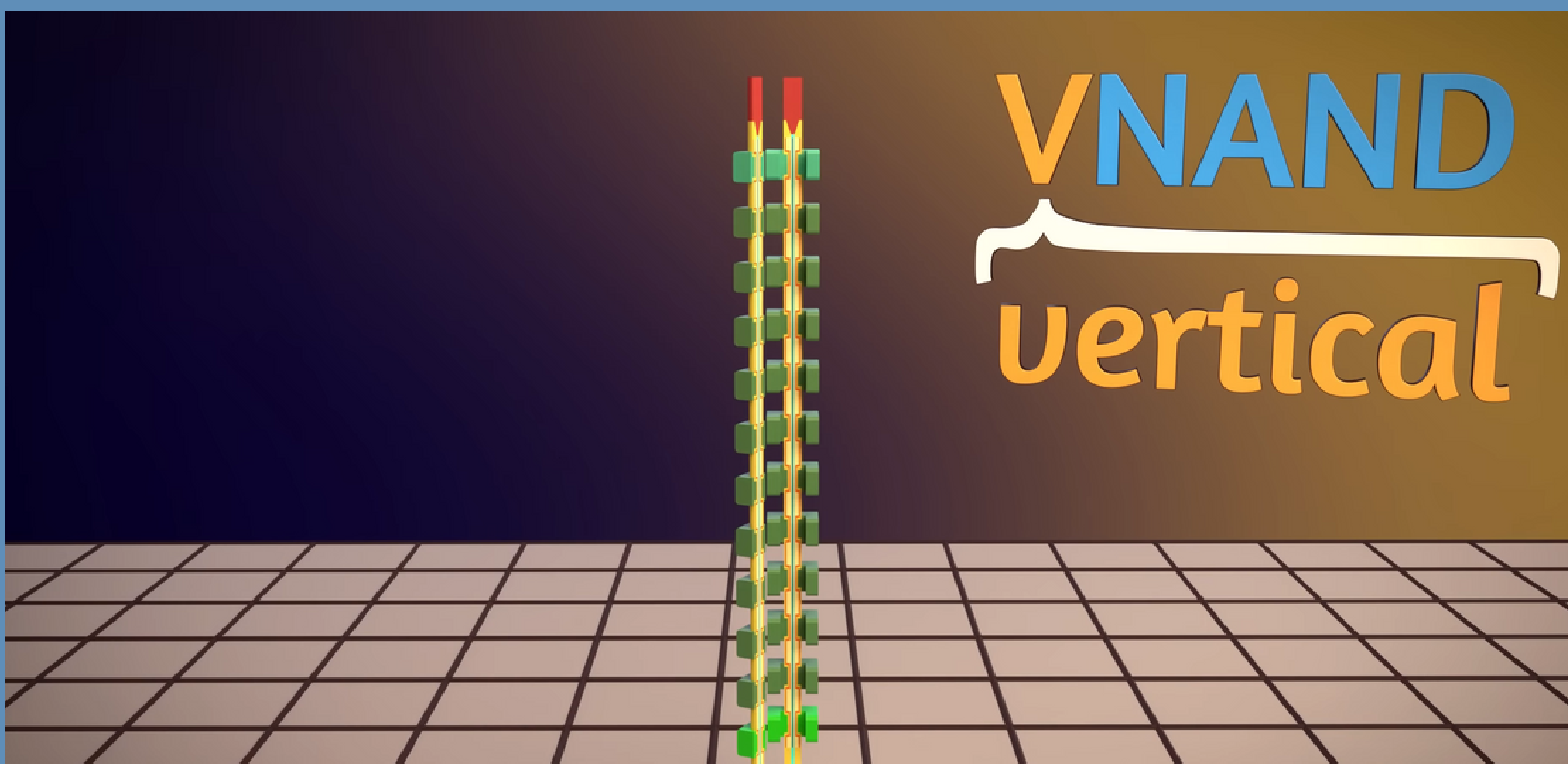
I mean- it’s called a charge trap for a reason.
It traps electrons, or charges for years on end,
and in order to read the information,
the electron charge level is measured,
and the amount of charge on the charge trap is unchanged.

String

After we zoom out a little,
you can see that the memory cells are stacked vertically.
This is where the vertical part in Vertical NAND or VNAND comes from.

This stack of memory cells, what is technically called a string is composed of 10 charge trap flash cells layered one top of another.

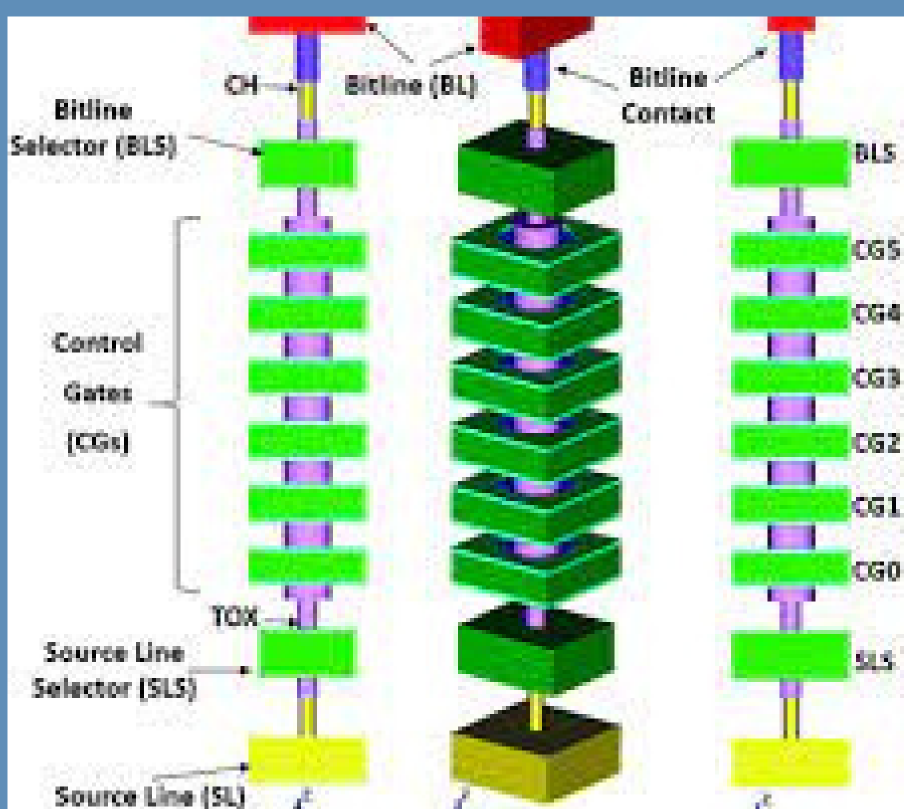
When information is written to or read from a string,
only one cell can be activated at any given time,
and to do that we use separate control gates attached to every layer in the string. It asks the bottom control gate first charge trap 1 the electron charge level.



Then the bottom cell sends that information through the center of the string up to the information highway at the top, which is technically called a bitline.

Then the next control gate for the 2nd layer asks for the charge level in the 2nd cell, and so on, up the string, each cell sending their information up to the highway or bitline.

The main thing is that only one layer in the string is either written to or read from at any given time.
Let’s move on in complexity, next we duplicate this string 32 times, and this gets us a page of strings.

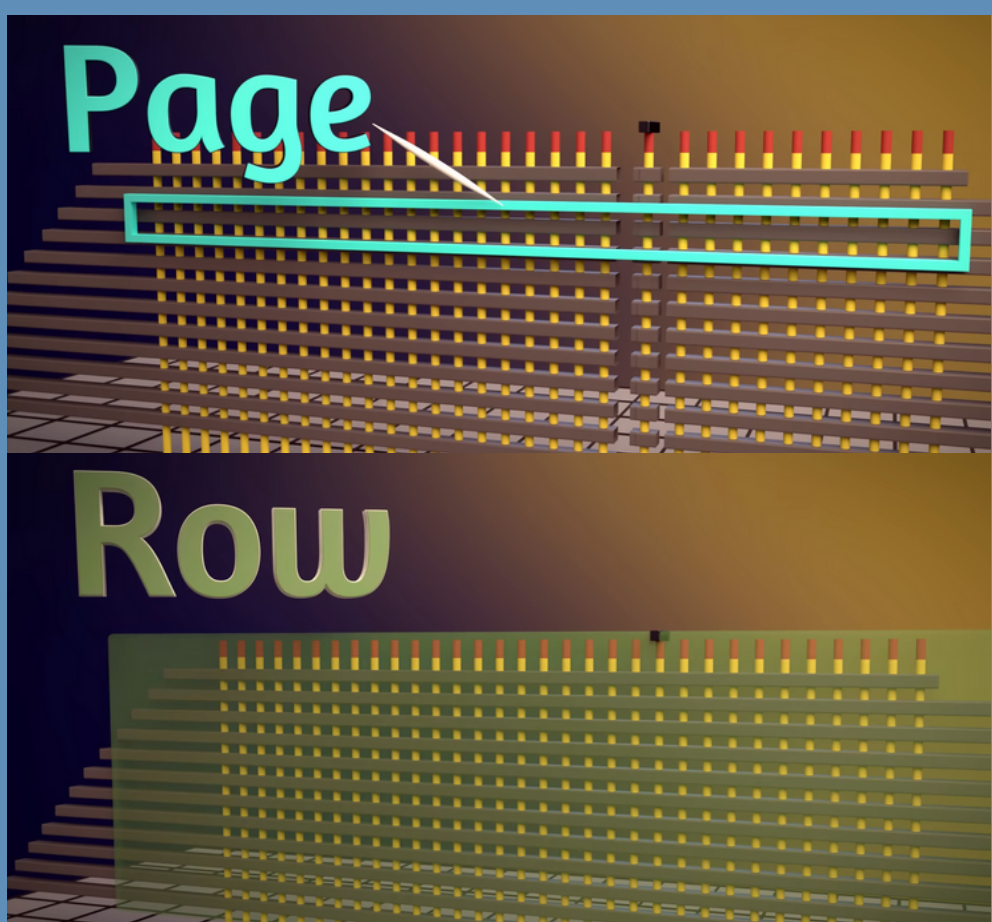


Page and Row

Let's move on in complexity, next we duplicate this string 32 times, and this gets us a page of strings.

And now here we have a page, and we are going to call this entire page of strings a row. When we duplicate the string, we also duplicate the bitline 32 times, however rather than duplicate the control gates, we are going to have every cell in the same page share a common control gate.

This makes it such that when information is written to or read from a row, an entire page composed of 32 adjacent cells, all in the same layer, are activated at the same time.

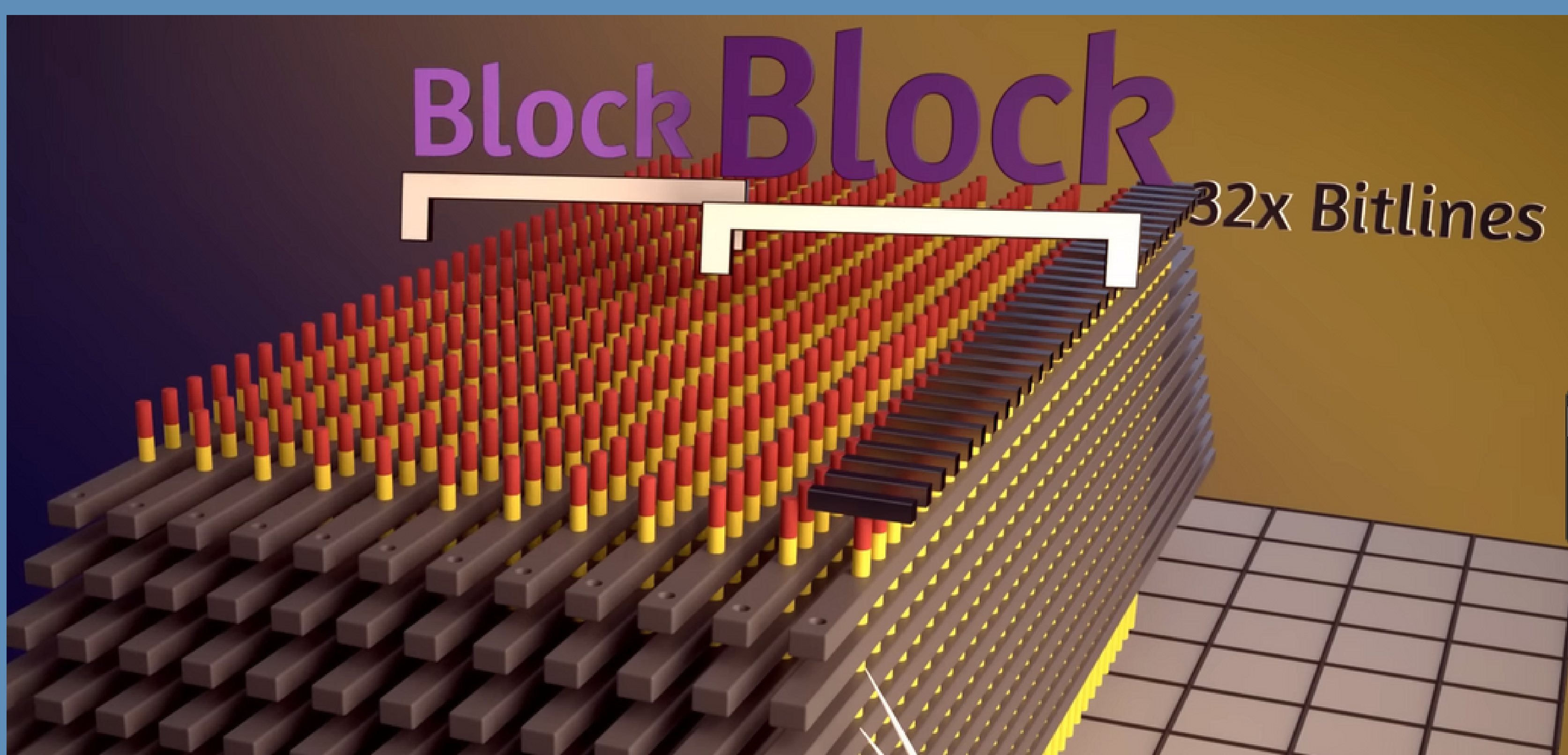


Block

Let’s step up in complexity again:
We duplicate the rows 6 times until we get a block, but we are going to do it 12 times so we can see 2 blocks.

We are going to connect the tops of each string in a column together, so they all share the same bitline and our bitline is looking more like a highway now.

In addition, we have to add a control gate that selects between rows, so that only one row is using the bitline at a time. These are called bitline selectors.

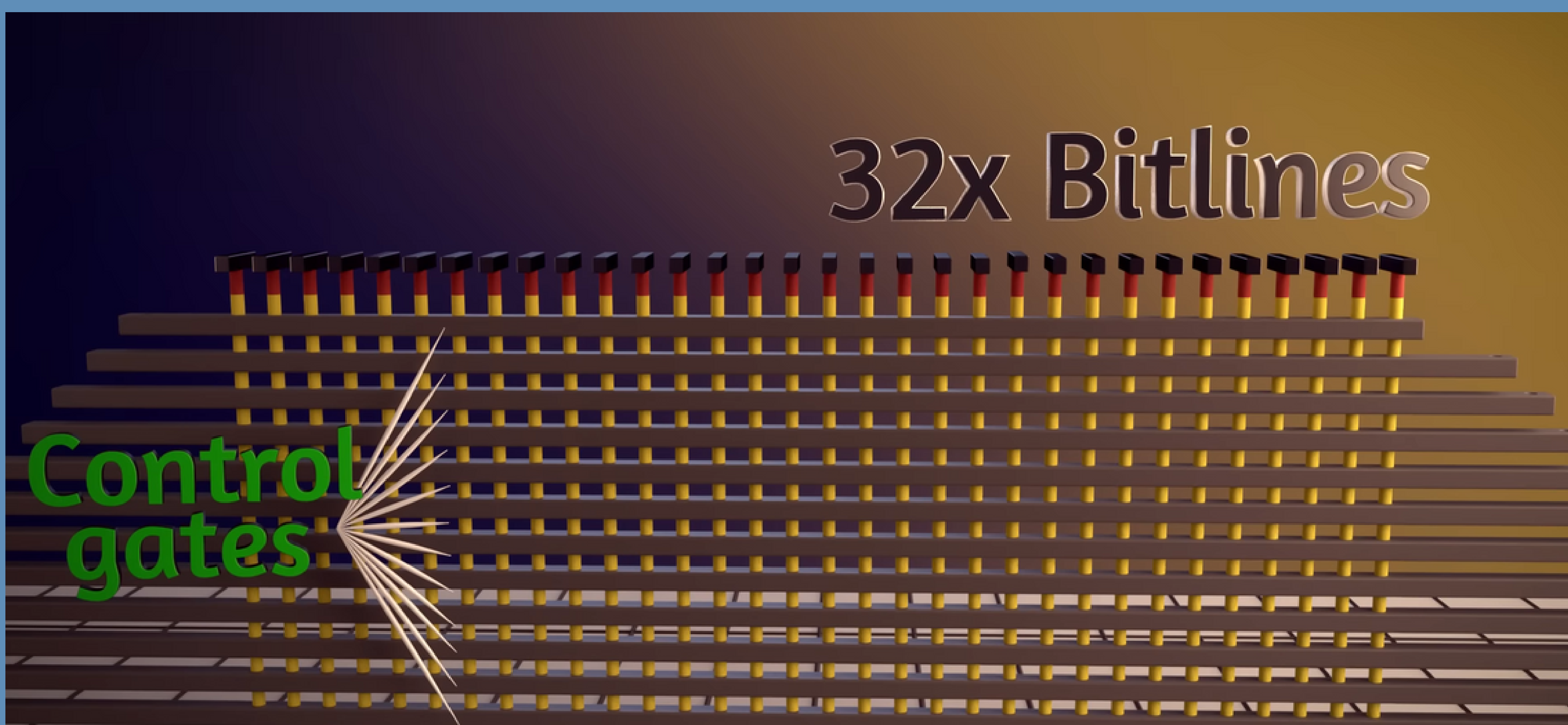


Bitline and Control gate selector

As discussed these bitlines are like highways, and the selectors at the top act as traffic lights that mediate the flow of information so that only a single row can use the highway, or is active at a time.

Similarly, the control gates attached to each layer act as traffic lights for the layers. With bitline selectors along the tops of each row, and control gate selectors along each layer, the solid state drive can read from or write to a single page at any given time.

Additionally, in order to connect to the bitline selectors and control gate selectors there are wires that drop down from above and run perpendicular to the bitlines.



Calculations

8 different levels of electrons are placed on charge traps in order to store 3 bits of information.
These charge trap flash memory cells are stacked into strings 10 cells tall, which are duplicated into pages of 32 strings in a row.

Next, those pages of strings are duplicated until we have a block 6 rows deep, and here we are showing 2 blocks. Doing some quick multiplication we find that there are 3,840 memory cells here capable of storing a total of 11,520 bits.

With each pixel in our picture requiring 24 bits,
that means that we can store 480 pixels,
That means you need about 25 thousand times the size of this layout to store the contents of this single picture.

$$\frac{10\text{ cells}}{1\text{string}} \times \frac{32\text{strings}}{1\text{ row}}$$

$$\times \frac{6\text{ rows}}{1\text{block}} \times 2\text{blocks}$$

$$= 3{,}840 \times \frac{3\text{ bits}}{1\text{cell}}$$

$$= 11{,}520\text{ bits}$$

THANK YOU