Storing and Preserving Digital Images

Brian Reid
13 August 2011

I gave this talk at the International Center of Photography to an audience of photographers, most of whom had film cameras with them. I added these notes the next day so that people who didn't hear the talk might better understand what I'm trying to say here.

A story that will never be told

"I went to my grandparents' farmhouse in Pennsylvania, which had been rented out for decades after they died.

I found in the barn a box containing a hard drive. There were thousands of perfectly-preserved JPEGs on it, all family treasures."

OK, how about this version?

“... I found their old digital camera in the barn and and the film card was undamaged and was in a format that my computer could read ...”

Responsibility has shifted

- When cameras produced permanent media,
- and prints were made on permanent media,
  - preservation just happened.
- Now you have to take explicit action to preserve your pictures.
  - Preservation doesn't "just happen".

Oh, what ever can I do?

- I’m willing to believe that I have to take specific action or my photographs will be lost.
- What should I do?
- Start by understanding the obstacles. Know your enemy.

Let's travel in time

- It's August 2111
- NiCad Palin is President
- The US is receiving foreign aid and food shipments from China
- The gigapixel Colt Leica M16A2 was just released (cost is only ¥ 25 million)

The numbering sequence of Leica M camera and Colt military rifles have overlap. But surely the M16 rifle will have become a studio prop by the time Leica issues an M16 camera.

You found a box of old pictures from 2013

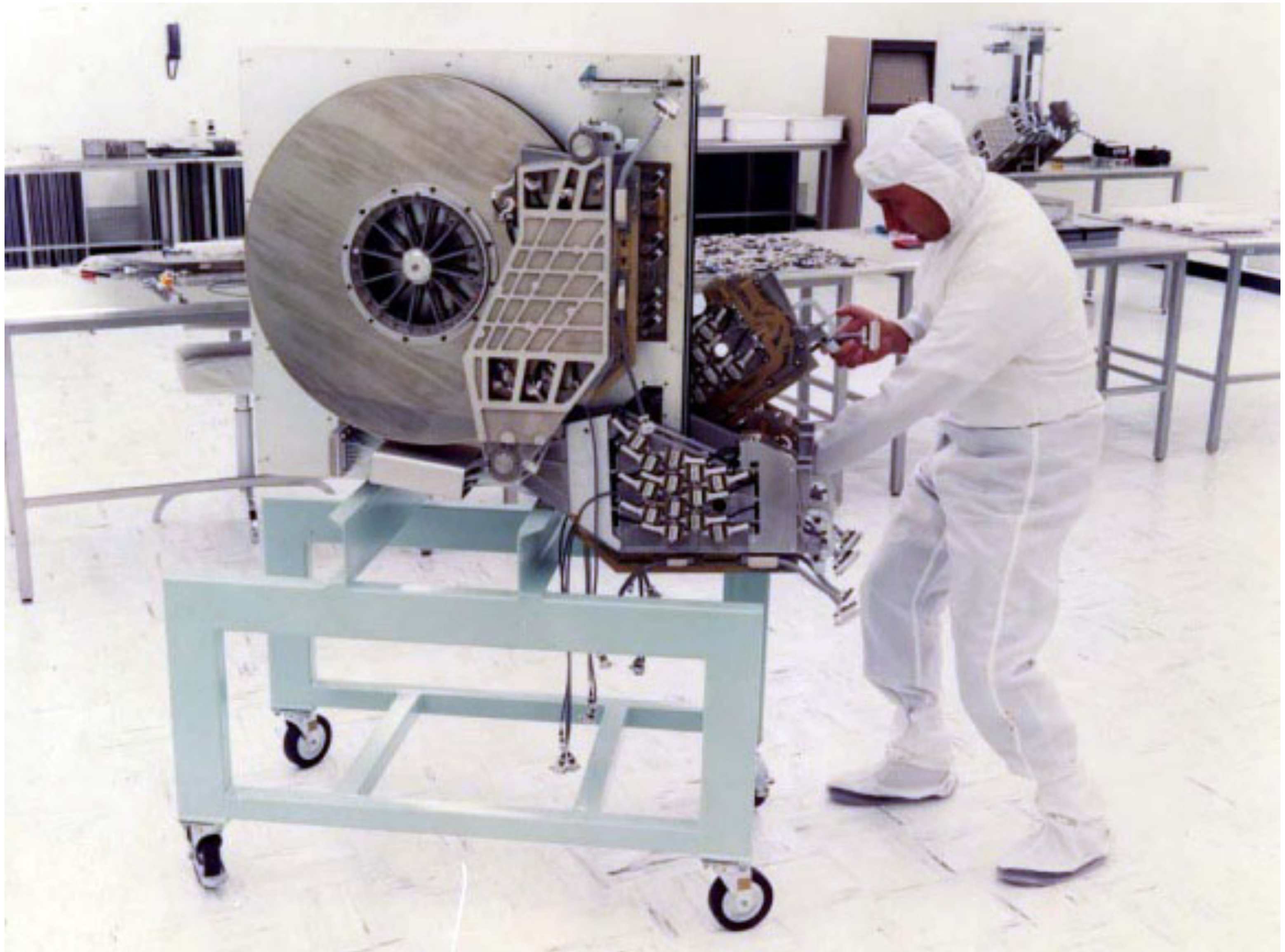
- It contains a stack of DVDs,
- a couple of hard drives,
- a jump drive
- and some inkjet prints in a Ziploc bag

Hard drive evolution

From right to left, front to back: an iPod drive, a laptop drive, a desktop computer drive, a first-generation home computer hard drive, and two mainframe hard drives from about 1980 and 1975.

Hard drive prehistory

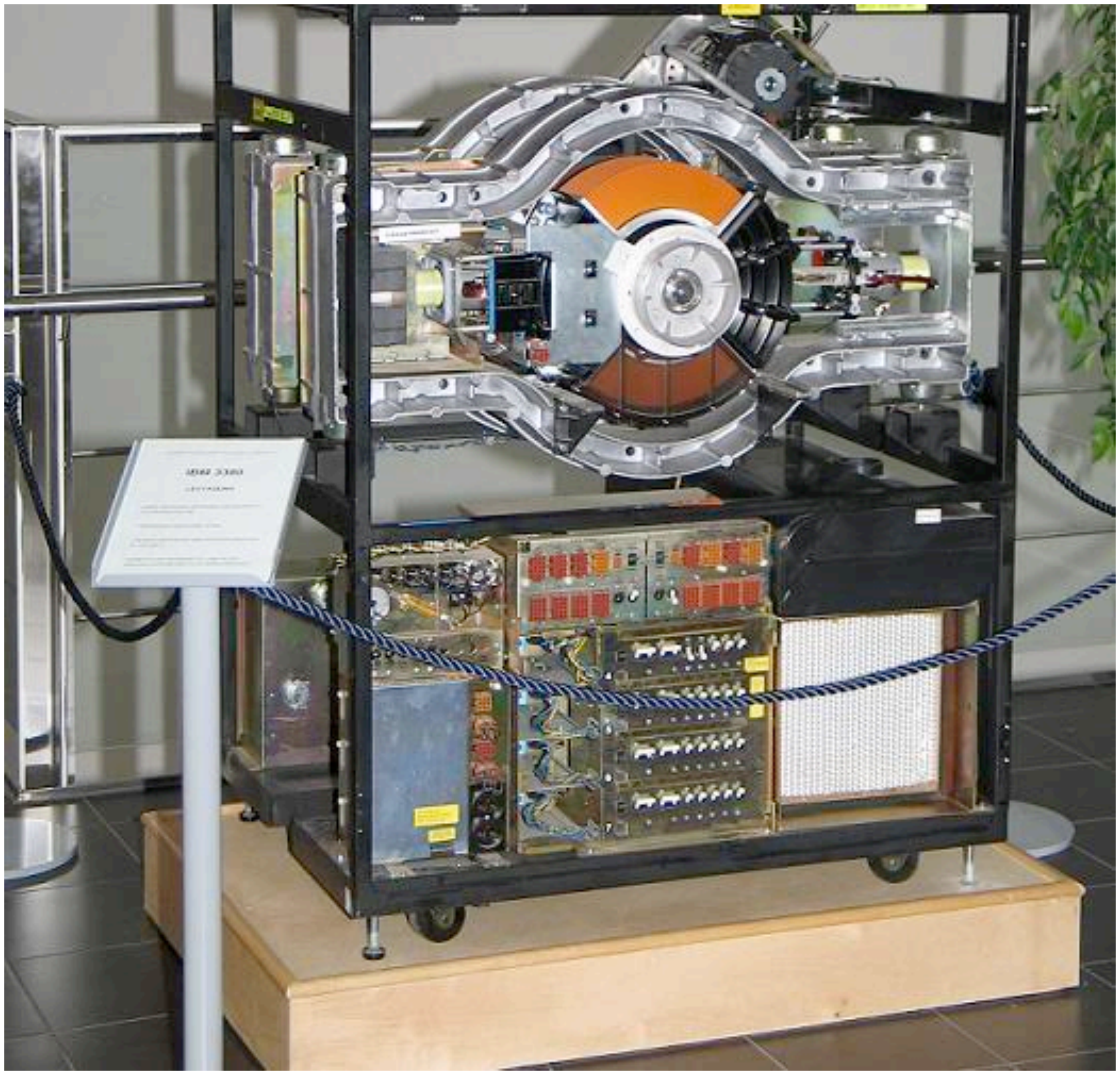
3350
300MB
1975



The “call history” memory on your cordless phone has more storage than this drive. It was manufactured (by IBM) in a clean room, and packaged against dust before shipment to the customer. It cost about $55,000 and looked like this when installed: http://www-03.ibm.com/ibm/history/exhibits/storage/storage_3350.html

Hard drive prehistory

3380
2.52GB
1981



The 3380 was a magnificent achievement by IBM at the very peak of its market dominance and technical excellence. The first time I saw one it took my breath away. This model cost $82,000.
http://www-03.ibm.com/ibm/history/exhibits/storage/storage_3380.html

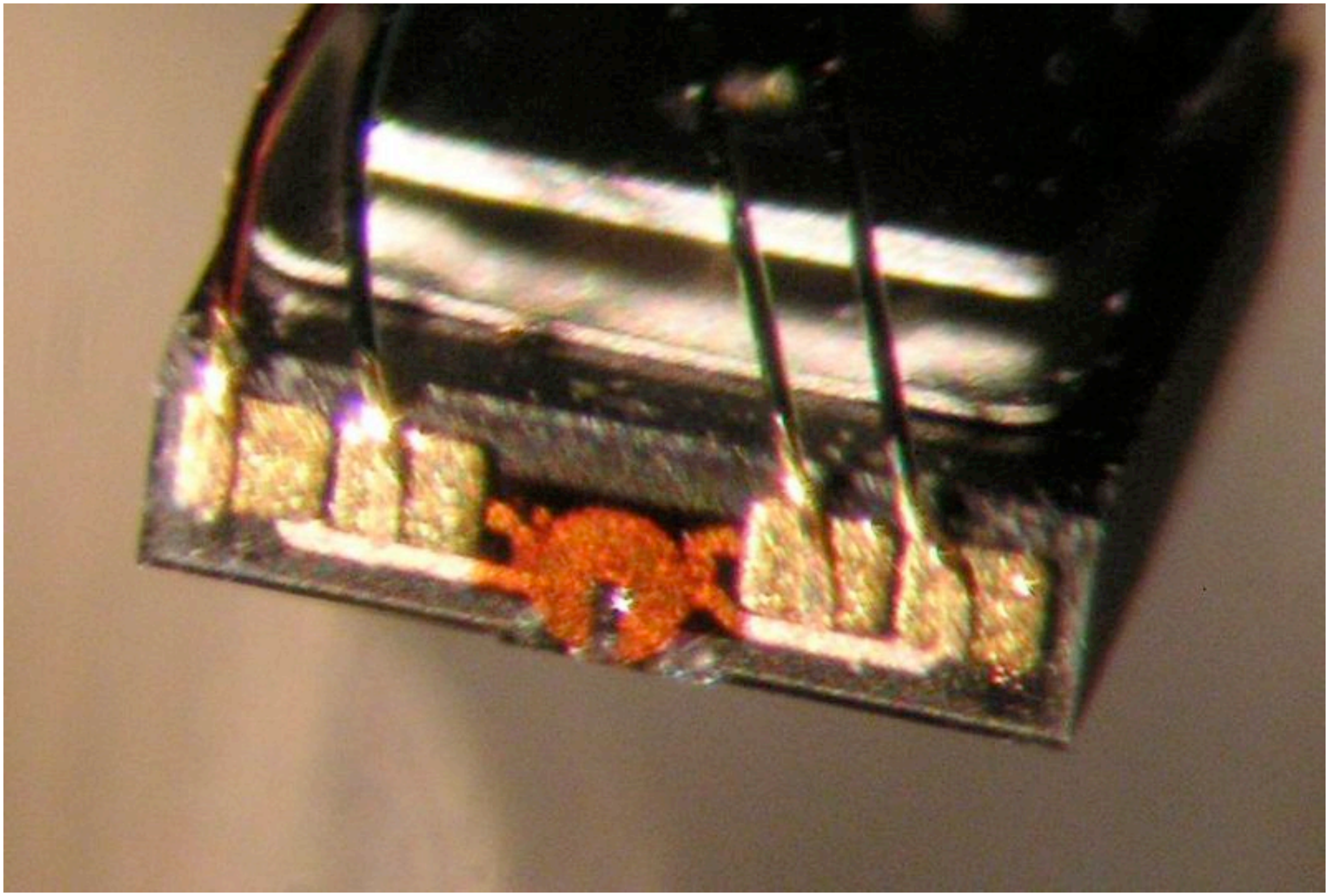
Hard drive, 2011

This is an IDE hard drive, 80GB, vintage about 2002. The platters spin horizontally. The boom with green and yellow parts on it moves in and out like a phonograph arm, positioning the read/write head over the needed track.

Close-up of head boom

This is a close-up of a different disk that has a similar read/write boom, showing the three platters of the disk (each has a head boom on each side). Look at the red spot at the very tip of the boom. That's the actual read/write head.

Close-up of R/W head

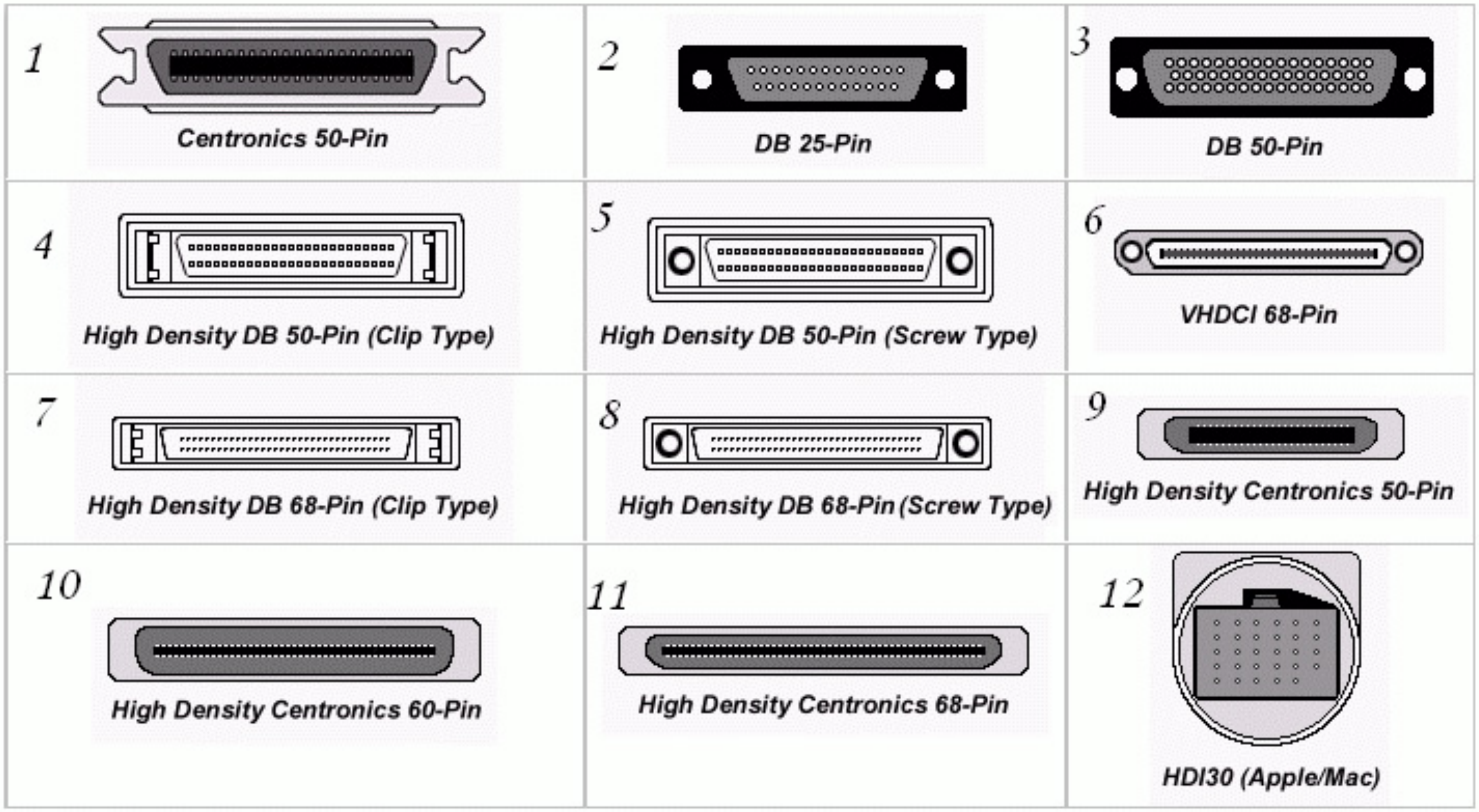


Here's an ultrahigh magnification (about 1000x) of a disk head. The actual sensing of the disk takes place in the gap at the bottom center of the red thing. Think about what would happen if a particle of dust got between this head and the platter spinning at 7200 RPM just beneath it.

What will it take to get the images?

- Physical: each medium must be connected to a device that can talk to it
- Software: each must have a file system format that your device and decode
- File format: you must have software that can deal with JPEG or RAW or whatever

Physical interconnects



Here are some connectors that have been used on commercial disks. To a certain extent you can identify a disk by the connector it uses, but these are all SCSI connectors. Connectors 7, 8, and 11 are the newest.

Quickly: what kind of connector is this?

This is an IDE disk, also called Parallel ATA. (It was just called ATA until Serial ATA came along). One major design flaw is that the bendable and breakable pins are on the precious disk and not on the cheap cable. This disk likely held 120GB.

What about these?

These are all SATA disks. The SATA connector itself is just the pair of similar-looking edge connectors, one long and one short. SATA is a technology for getting data to and from a disk a lot faster, but they fixed a lot of mechanical problems when they designed its connector.

Stumped?

- 100 years from now a person probably won’t even know what the connector is
- let alone actually have one
- let alone having a device that can talk to it
- Pictures trapped forever in a magnetic box
  - if it’s still magnetized

There were scenes in the Harry Potter stories in which he had objects with secrets locked inside and no way to get them out. Alas, in real life there is no magic wand to open old disks and get the pictures out.

OK, what about DVD?

- If we can’t get data from a 100-year-old hard drive, what about a DVD?
- DVDs were originally used to distribute movies, before the advent of online streaming.
- By 2016 DVDs were primarily used just to store computer data.

I think the real reason why DVD disks were used for so long to distribute motion pictures was that the technology made it easy to add the “FBI Warning” at the front and prevent you from skipping over it.

Um, what kind of DVD?

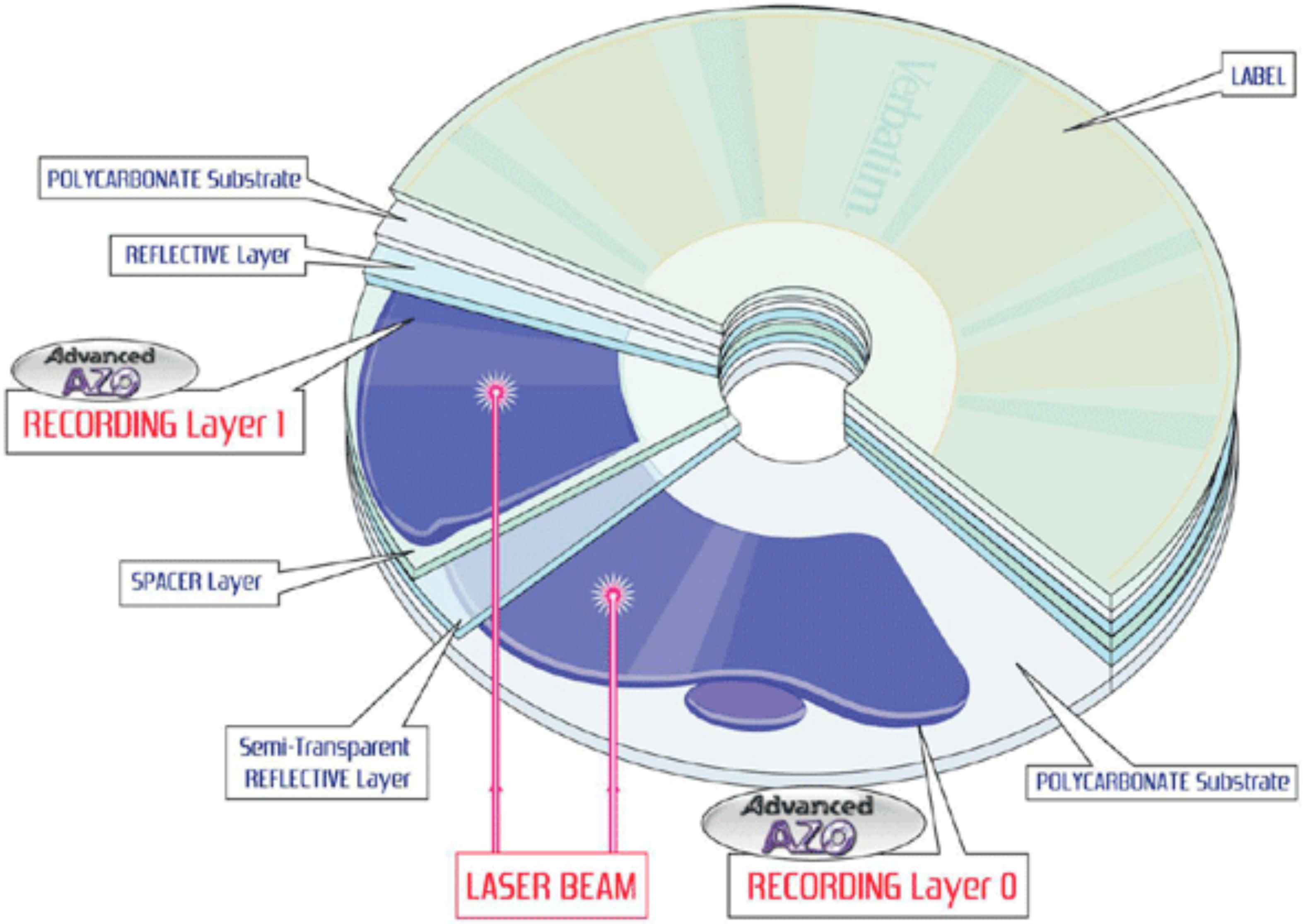
Capacity and nomenclature of (re)writable discs

Designation		Sides	Layers (total)	Diameter (cm)	Capacity	
					(GB)	(GiB)
DVD-R	SS SL (1.0)	1	1	12	3.95	3.68
DVD-R	SS SL (2.0)	1	1	12	4.70	4.37
DVD-RW	SS SL	1	1	12	4.70	4.37
DVD+R	SS SL	1	1	12	4.70	4.37
DVD+RW	SS SL	1	1	12	4.70	4.37
DVD-R	DS SL	2	2	12	9.40	8.75
DVD-RW	DS SL	2	2	12	9.40	8.75
DVD+R	DS SL	2	2	12	9.40	8.75
DVD+RW	DS SL	2	2	12	9.40	8.75
DVD-RAM	SS SL	1	1	8	1.46	1.36*
DVD-RAM	DS SL	2	2	8	2.65	2.47*
DVD-RAM	SS SL (1.0)	1	1	12	2.58	2.40
DVD-RAM	SS SL (2.0)	1	1	12	4.70	4.37
DVD-RAM	DS SL (1.0)	2	2	12	5.16	4.80
DVD-RAM	DS SL (2.0)	2	2	12	9.40	8.75*

From Wikipedia

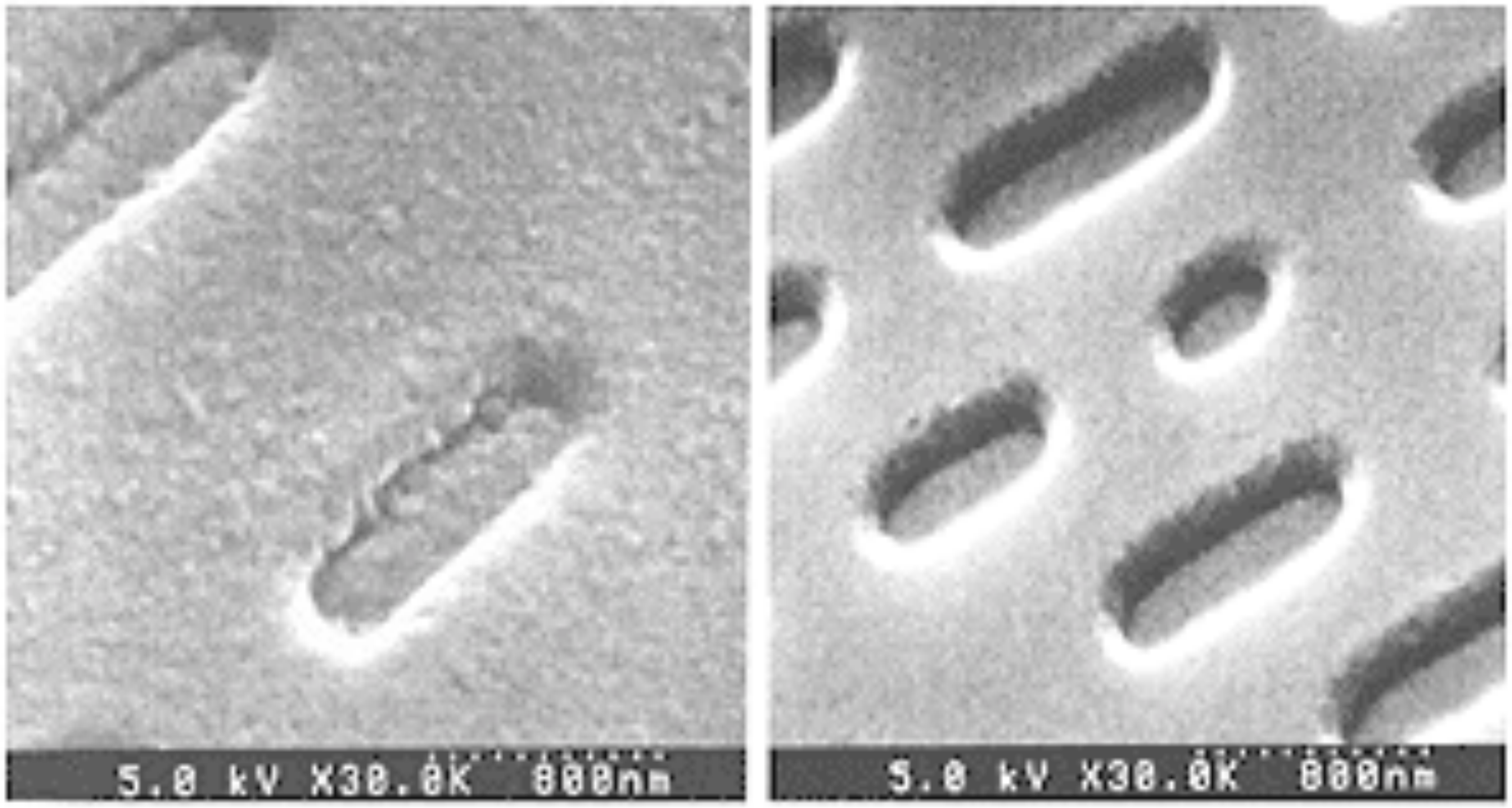
Patrick McFarland has a really good explanation of why you should care about the differences between types of DVD:
http://adterrasperaspera.com/blog/2006/10/30/how-to-choose-cddvd-archival-media/

DVDs are delicate



DVDs look simple, and teenagers toss them across the room. They are not simple. This diagram from Verbatim shows a single-sided double-layer disc, whose proper functioning depends on all of the layers being manufactured properly. But it sells for 59 cents, so there are limits.

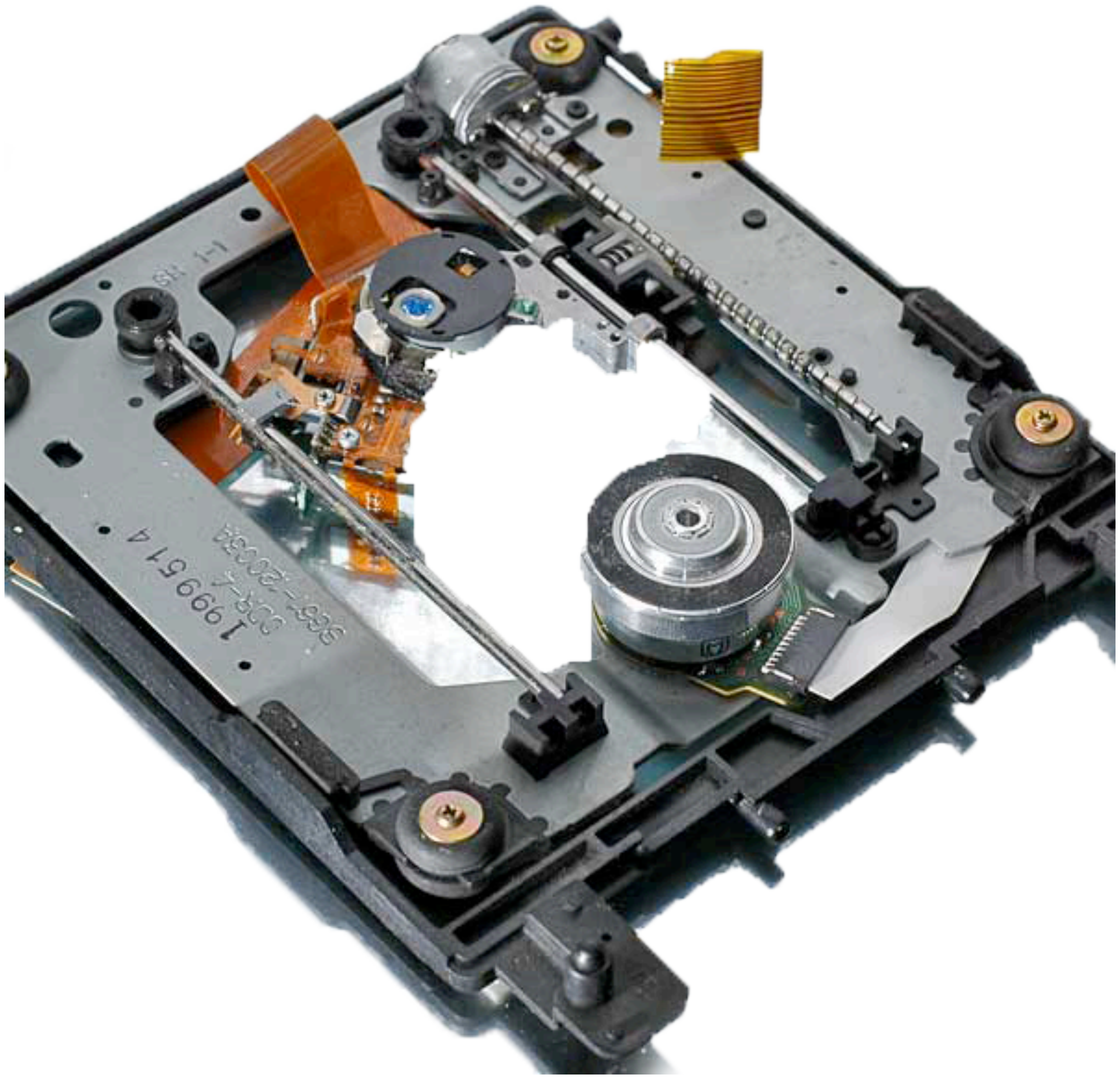
DVDs rely on precision



CD DVD

The primary difference between CD and DVD discs is the spacing between the circles of data. When reading a DVD it is much easier to get the wrong track by accident. A motor moves the laser head in and out over the tracks to choose which one is being processed.

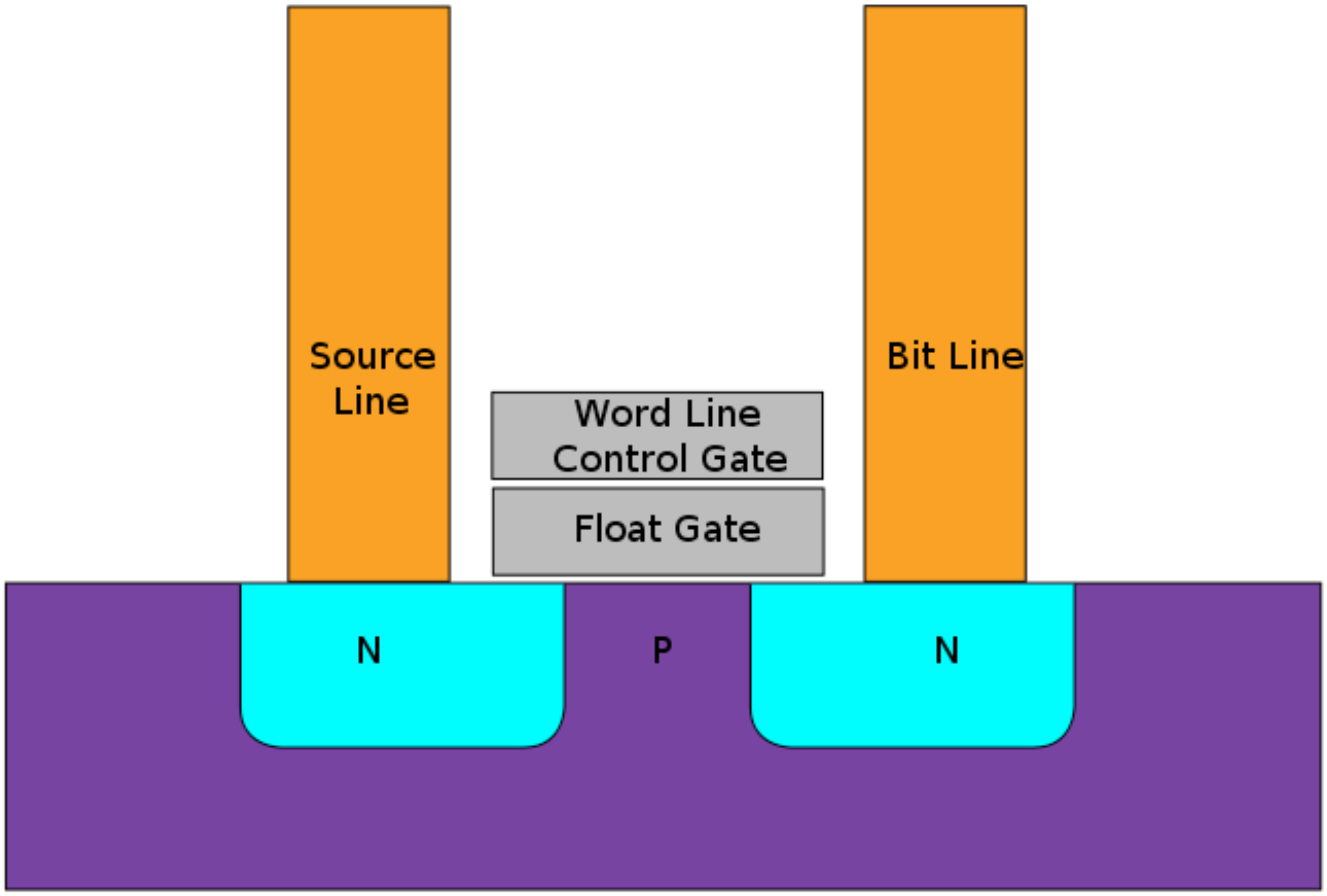
Has the DVD drive rusted away?



But the disc drive itself sells for about $5. It is a high-precision marvel.
This is a photograph of an ordinary commercial DVD drive, ready to be installed in an enclosure. The disc spindle sits on the round hub at bottom right. The laser is the blue dot; it moves in and out along the two rails, positioned by the motor at top center. Typical disc drives are not sealed and (like this one) all of their high-precision components are exposed the the environment. It is certainly possible to build a mil-spec DVD drive with its mechanisms protected and contained, but you won't find them in retail stores.

OK, maybe flash drive?

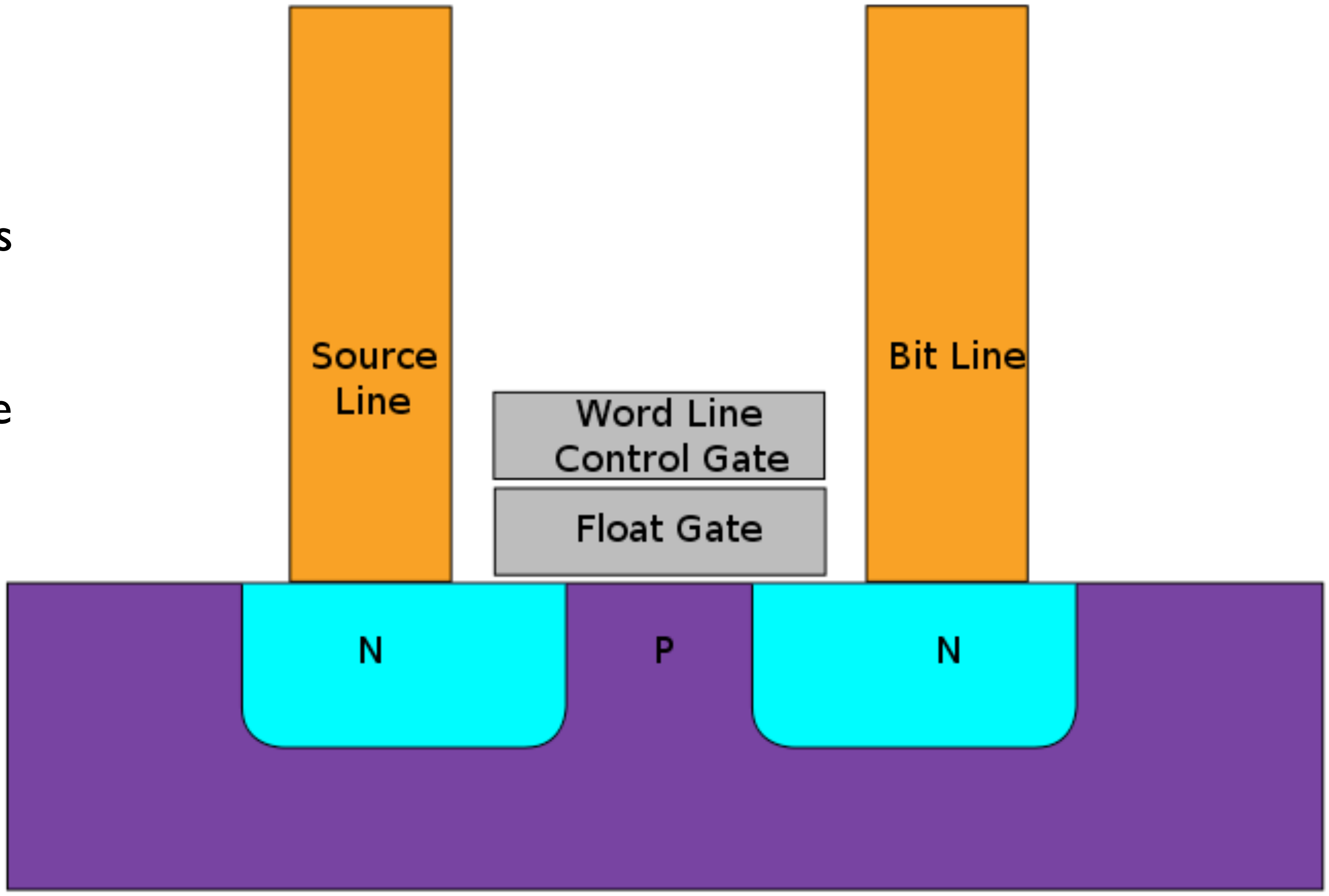
Let's look at how flash works



Flash memory uses floating-gate MOSFET transistors (Metal Oxide Semiconductor Field Effect Transistors). Never mind that they are no longer made with either metal or oxide, they are still called MOSFETS. A floating-gate transistor can store information by piling up electrons on the “Float Gate”. If enough electrons are there, then you can poke the “Source Line” and get a reply out of the “Bit Line”. If there are no electrons there, then you won't get a reply. One and zero. The number of surplus electrons that have to stay on that float gate in order that your flash memory doesn't lose data is shockingly small. 500 electrons mean “1”, and 50 electrons mean “0”. I do not trust 500 electrons to stay where I put them for 100 years.

Let's look at how flash works

Memory relies on electrons staying put in the Float Gate



But suppose you could

- Just for a moment, let's imagine that we can in 2111 connect to the old media.
- Now what? Hard drives have file systems. Which file system does it use?
- Wikipedia lists more than 200 file systems. Even the ones you've heard of have variations.

File systems usually have "FS" at the end of their names. But FAT and FAT32 don't. There's NTFS and HFS and ZFS and ReiserFS and UFS and FFS and hundreds more. Your computer probably uses exFAT or FAT32. If your software doesn't intimately understand the file system, it can't read the data.

If you get the image file

- Suppose that you get the image file.
- Is it JPEG, JPEG 2000, JPEG XR, TIFF, RAW, PNG, GIF, BMP, DNG, PNM, PSD, or something obscure.
- I have images in several Apple, Logitech, and Kodak formats that I can no longer read.
- Is any image file format future-proof?

For example, in 2011 it is surprisingly hard to read a Kodak Photo CD.

Best bets: image format

- Normally open formats better than proprietary formats;
- Normally you should make TIFF files.
- But it's beginning to look as though DNG (which is technically excellent) will distance from Adobe control.
- I make both. Storage space is not precious.

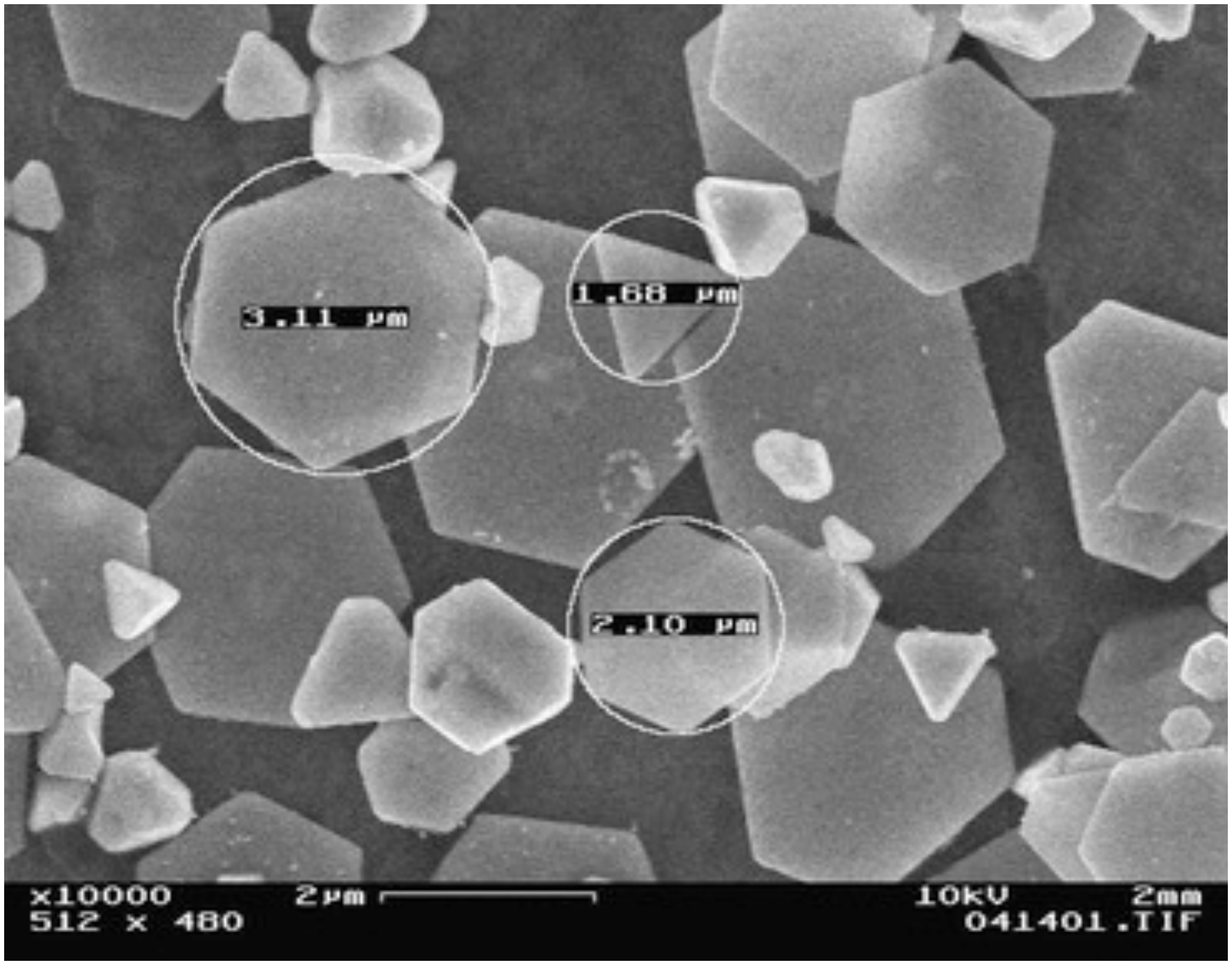
Adobe software won't let you make a DNG file out of something that is not a "negative", but most other companies' software will. Disk space today can be had for about 5 cents per gigabyte.

Best bets: file system

- FAT32
- But copy to something current every 5 years
- Don't use Apple file systems (HFS, HFS+, etc)
- Don't use NTFS

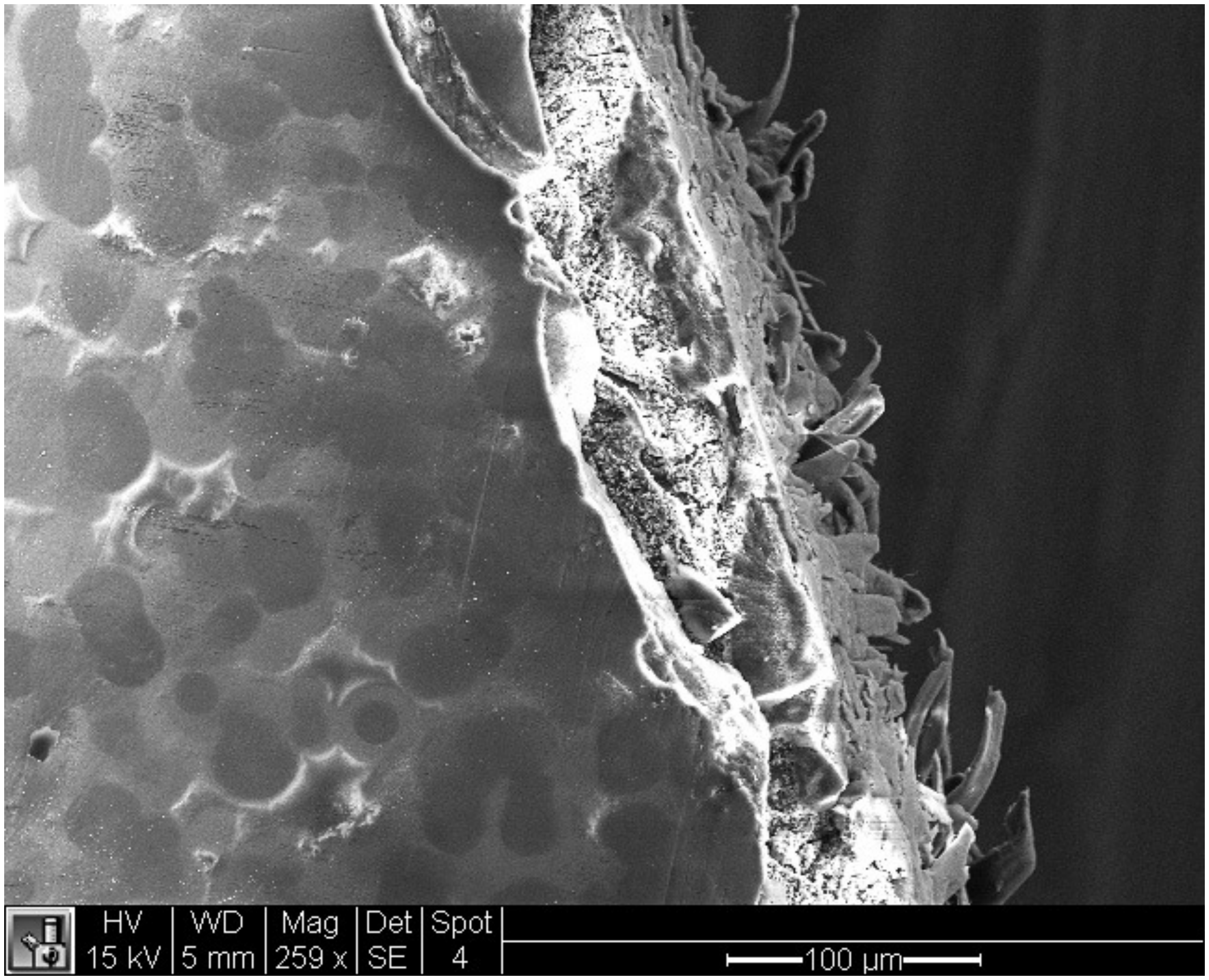
I believe that the file system is the part of this fragile chain that is most likely to become impenetrable. There is no one system that dominates the market, and vendors can't afford to make their devices understand multiple systems. Very few computers can decode a disk that wasn't written by its own kind.

More electron microscope pictures



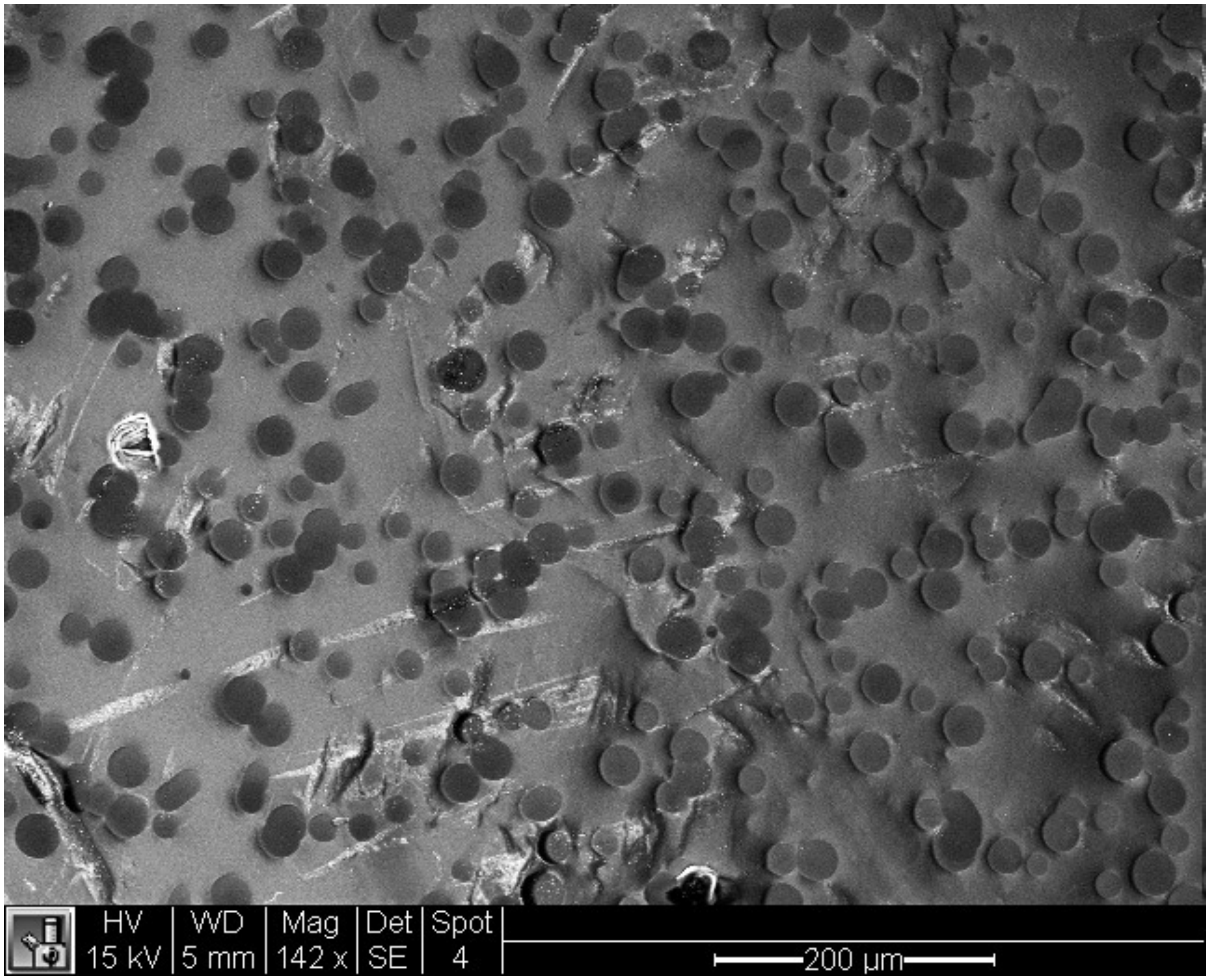
This is TMAX 400. It will still be TMAX 400 a century from now unless someone loses it or burns it or shreds it.

More electron microscope pictures



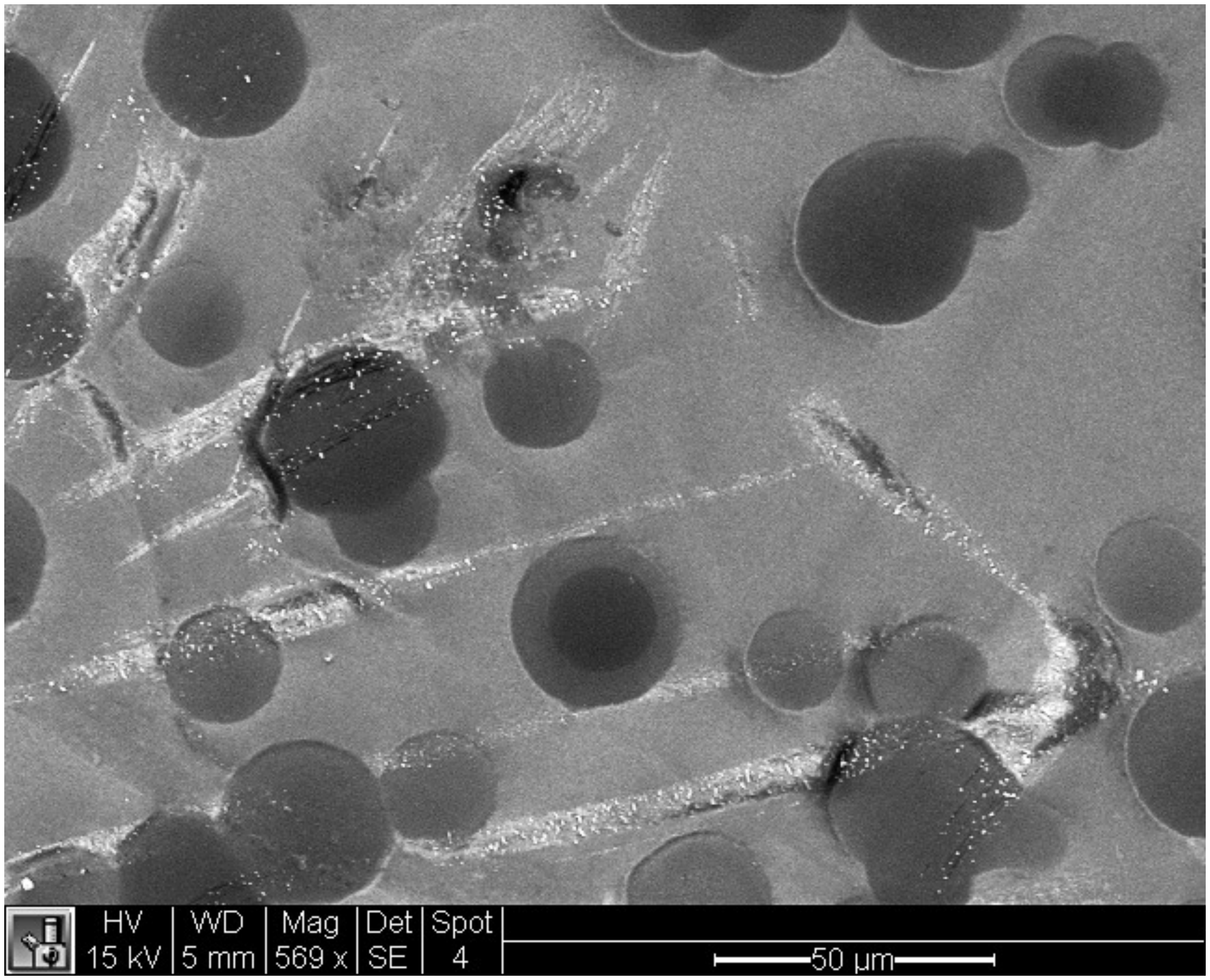
What is this? The white diagonal swatch is an edge cut. The label says it is 100 micrometers thick. Magnification 259x.

More electron microscope pictures



Here’s the same material, zoomed out to 142x.

More electron microscope pictures



Here's the same material, zoomed to 569x.
What is it? It's an inkjet print on glossy baryta paper. The white stripes are scratches from dropping it on the floor and picking it up carelessly.

Most permanent image medium is inkjet print

- Fooey on this digital stuff.
- Make inkjet prints on glossy baryta paper.
- Make them big. Use pigment ink. Epson K3 and HP Vivera seem to be the best choice.
- Store them according to library archive principles.

If you are a fanatic

- If you are worried about color fade, you can make RGB or CMYK separations and print each in black inkjet ink.
- Standard printer's alignment marks make it easy to reconstruct the original image.
- To get higher dynamic range you can make differently-exposed separation images.

Loss prevention

- Most loss of photos comes from human error:
  - Accidental deletion of files
  - Losing flash card
  - Losing your camera before you get the pictures out of it

What do I do?

- Make prints and keep them in a box.
- Keep files on my computer. No RAID.
- Back up every day, automatically, using 3 different systems (Retrospect, Time Machine, and /sbin/dump)
- TM copies to RAID; others to online disks
- Every few weeks, make external copy; send it far away

Boeing used to run advertisements noting that “Tomorrow never comes”, aimed at people who thought they might buy airplane tickets tomorrow. If you can find a way to automate your backups, they will actually happen.

Making external copies

I keep these near my computers. You put a SATA disk into it, let your backup software copy pictures to it, then take it out...
(It’s a Plugable USB 3.0/2.0 SATA Hard Drive Docking Station)

Storing external copies

... slip it into one of these, label it, and move it to a safe place.
(It's an IO Crest 3.5-Inch IDE/SATA HDD Storage Box)

What will I do?

- I have re-engineered my backup system about every 5 years for 30 years
- I leave external backup copies alone, even in ancient formats
- Live files and online backup get copied to new technology every 5 years or so

My external backups

- 1980: ½-inch magtape
- 1985: TK50 cassettes
- 1990: DDS-1 cassette tapes
- 1996: DDS-3 cassette tapes
- 2000: Exabyte M2
- 2002: PATA drives
- 2008: SATA drives

You’ve probably never heard of any of these. I list them to emphasize that making backups is not something I do casually, because it’s so important to me. I have about 200 DDS-3 cassette tapes in a box in my basement, packaged with a SONY drive that can read them. Well, it once could read them.

My backup hierarchy

- Live copy: online on ordinary disks
- Primary backups: online on my LAN
- Secondary backups: stored offline
- Tertiary backups: stored offsite

I really do this. I sleep better at night. And if a meteor falls on me, friends and family will already have the offsite copies. I don’t encrypt them.

Thank you

This presentation is currently online at

http://leica-users.org/NYLUG-2011.pdf