What can we tell about sector error rates from all those papers. They don't seem to measure quite the same things, so we have to normalize the data as much as possible.

RBLandau 20161009

Take the various numbers given in a dozen or so papers that we read as background material. Do they seem similar to the numbers that we have been looking at? We are interested in *sector* lifetime, not specifically in drive lifetime, so many of the papers were not very relevant.

Method:

- Find whatever seemingly relevant numbers were published in the papers.
- Try to re-scale them to the units that we have been dealing with, namely lifetime and half-life of our "metric sectors" expressed mostly in mega-hours.
- See if they relate to the range of numbers that we have been using. (Be "we" I mean 99% me as the guilty party, since I chose the ranges and units and such based on "feel" from early simulation runs. Dr. Altman has been encouraging me for months to justify my choices. This is a start.)

Assumptions about scale differences:

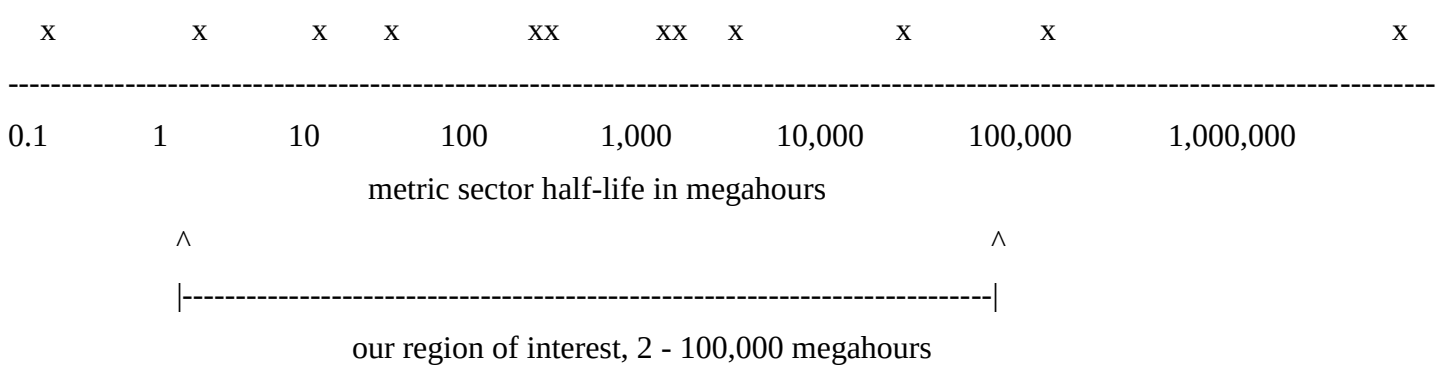
- Their year is 8,760 hours. Our metric year is 10,000 hours.
- Their sector size is 512 bytes. Our metric sector size is 1MB (2,000 vanilla sectors).
- Their drive size, for the old data, at least, like the Internet Archive data, is 200MB. Our shelf size is 1TB. For slightly more recent data, I have assumed 2TB drives.
- Some numbers are quoted for enterprise SCSI/FC drives, and some for consumer grade SATA drives. We prefer the SATA numbers, since the huge disk JBOD and RAID arrays tend to be inexpensive drives.
- Some figures quoted are like 1.5E-9 sector errors per two years or 1.2E-9 corruptions in six months. We translate those rates into mean sector lifetime in hours, and then into half-life, also in hours (or kilohours or megahours), and then scale up to metric sectors. (Isn't this fun?)

Original figure (and date published)	Rescaling required	Output SWAG Mean Metric Sector Lifetime (ML) and Half-Life (HL)

Original figure (and date published)	Rescaling required	Output SWAG Mean Metric Sector Lifetime (ML) and Half-Life (HL)
ASER 1.5E-9 per 2 yrs (2008)	0.75 bad sector / 1E9 sectors / yr 9E-6 bad sector / 1E9 sectors / hr 9E-15 err / sector / hr 175E-12 err / metric sector / hr	ML 6E9 hr HL 4E9 hr
errors on 1.5 to 10% of drives per year (2008) (Internet Archive 200MB drives)	1.5-10 errors / 100x200 MB / yr pick 4%: 4 err / 20E3 MB / yr 1 err / 10E6 sectors / yr 1 err / 88E9 sectors / hr 1 err / 44E6 metric sectors / hr	ML 44E6 hr HL 30E6 hr
sector half-life 200E6 years (2010)	200E6 yrs 1.8E12 hrs	ML 1.8E12 hr HL 1.2E12 hr [Way too high, in the immortal disk range.]
1.2E-9 corruptions in 6 months (2010)	2.4E-9 err / yr 300E-15 err / hr 600E-12 metric sector err / hr	ML 1.7E9 hr HL 1.2E9 hr
1.5E6 drives had 400E3 silent errors in 41 months (2008)	120E3 errs / yr / 600E9 sectors 200E-9 errs / sector / yr 1 sector err / 5E6 yrs 1 sector err / 40E9 hrs 1 metric sector err / 20E6 hrs	ML 20E6 hr HL 14E6 hr
4E-3 errors / GB in 18 months (2008)	2.7E-3 errs / GB / yr 1 err / 375 GB / yr 1 err / 3.3E6 GB / yr 1 err / 6.6E12 sectors / yr 1 err / 750E6 sectors / hr 2E3 err / 750E6 metric sectors / hr 1 err / 370E3 metric sectors / hr	ML 370E3 hrs HL 260E3 hrs [Way too low, this is outside the range of even rusty garbage can lids.]

Original figure (and date published)	Rescaling required	Output SWAG Mean Metric Sector Lifetime (ML) and Half-Life (HL)
sector failures 5x drive failures (2006)	drive AFR 2-4% / yr, pick 3% 15% of drives have 1 error / yr assume 4TB drives 1 error / 15% x 4TB / yr 1 err / 600 GB / yr 1 err / 1.2E9 sectors / yr 1 err / 10E12 sectors / hr 2E3 err / 10E12 metric sectors / hr 1 err / 50E9 metric sectors / hr	ML 50E9 hours HL 35E9 hours
1 error in 1E15 bits (SCSI) or 1E14 bits (SATA) 2006		
sector mean lifetime 2.7E6 hours for 200MB drive (2006)		ML 2.7E6 hrs HL 1.9E6 hrs
1700x200MB drives: audit 4 months => MTTDL 30E3 hours; audit 2 weeks => MTTDL 100E3 hours (2006)	[tricky to translate to ML]	
3.5% on drives > 3 yrs old (2007)	assume 2TB drives 1 err / 3.5% of 2 TB / yr 1 err / 70 GB / yr 1 err / 140E6 sectors / yr 1 err / 1.2E12 sectors / hr 1 err / 600E6 metric sectors / hr	ML 600E6 hr HL 400E4 hr
3% nearline drives in first yr; 1.5% enterprise drives in first yr (2007)	assume 2TB drives 1 err / 3.5% of 2 TB / yr 1 err / 60 GB / yr 1 err / 120E6 sectors / yr 1 err / 1E12 sectors / hr 1 err / 530E6 metric sectors / hr	ML 530E6 hr HL 360E6 hr

Where do all these SWAGs fall in or near our assumed range of interest? (Apologies for the ASCII art.)



Note that almost all the SWAGs from the various articles fall somewhere in the "region of interest." Recall that the "region of interest" was chosen by these criteria:

- There are *some* errors rather than none. This is consistent with experience. If there were no errors, we would not need to take any special actions to protect our collections; they would be immortal.
- There are not so many errors as to make the system unusable. Also consistent with experience. Disks do lose some data now and then; it happens with disconcerting frequency on *my* computers, at least.

So the range runs from sectors with very short lifetimes -- so short that you would never consider such storage "archival" for longer than a few minutes or hours -- to very long lifetimes where you might not suffer any bad blocks for years at a time even with large collections. Still seems reasonable to me.

RBL