In Search of Lost Data: A Study of Flash Sanitization Practices

By:
Janine Schneider, Immanuel Lautner, Denise Moussa, Julian Wolf, Nicole Scheler, Felix Freiling, Jaap Haasnoot, Hans Henseler, Simon Malik, Holger Morgenstern and Martin Westman

From the proceedings of

The Digital Forensic Research Conference

DFRWS EU 2021

March 29 - April 1, 2021

In Search of Lost Data: A Study of Flash Sanitization Practices

Janine Schneider

March 30, 2021

IT Security Infrastructures Lab
Department of Computer Science
Friedrich-Alexander University Erlangen-Nürnberg (FAU)

Joint work with:

Immanuel Lautner, Denise Moussa, Julian Wolf, Nicole Scheler, Felix Freiling, Jaap Haasnoot, Hans Henseler, Simon Malik, Holger Morgenstern and Martin Westman

Data Sanitization

- Data can be recovered from disk drives unless effort is spent on its deletion
- Poor sanitization practices for second-hand market have been confirmed
- Sanitization practices are relevant if incriminating data is found on a storage device

Data Sanitization

But what about new devices?

Data Sanitization

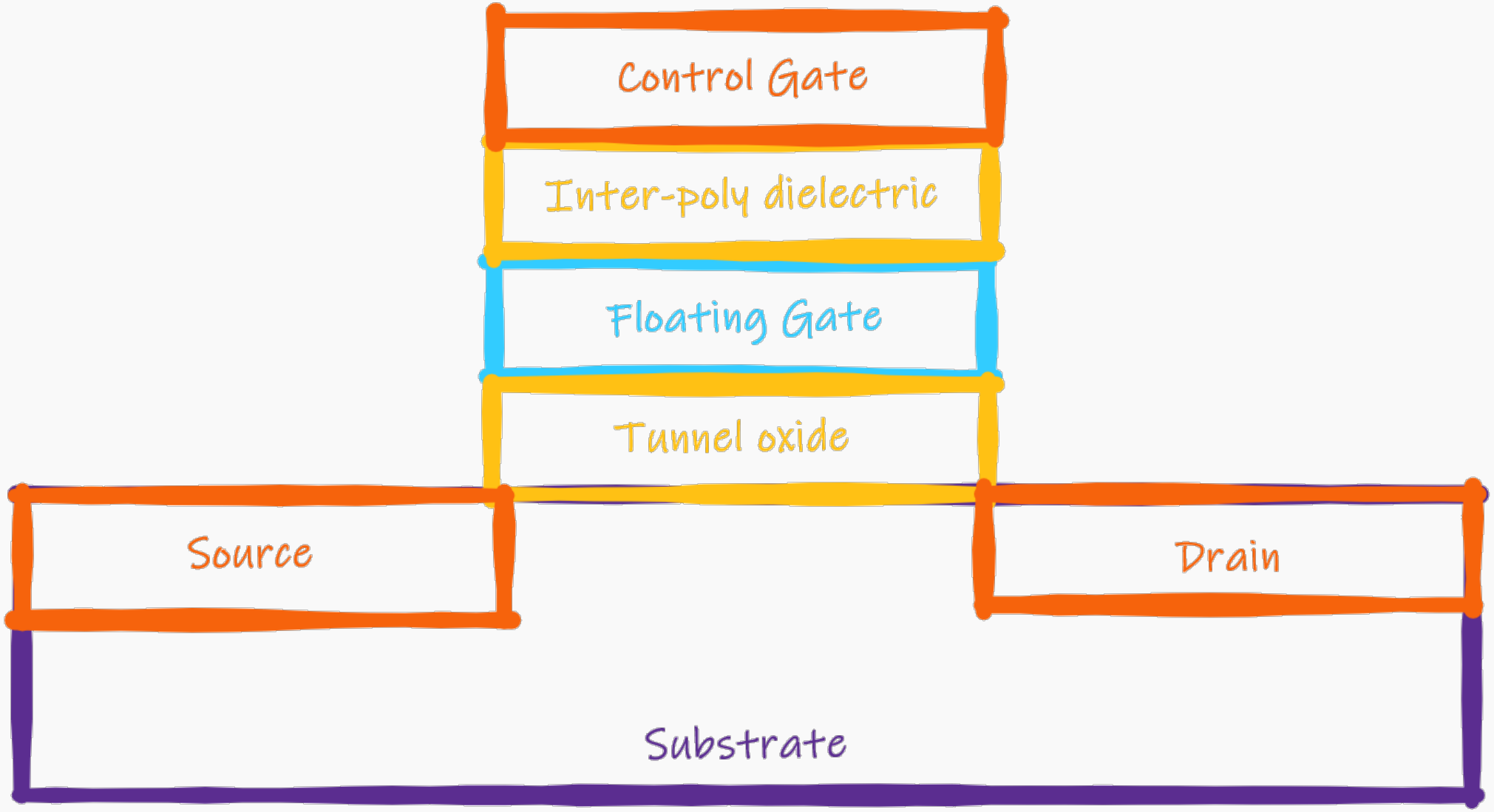
- 2017 Martin Westman reported that he had found non-trivial data on new USB drives
- Data may come from the reuse of memory chips in USB devices
- Reused memory chips are much cheaper than new ones

Research questions

- What is the risk of acquiring evidence on new USB drives that are due to former usage (of components) of the drive, i.e., usage before it was bought?
- What factors influence the probability for the existence of non-trivial user data on new USB drives?

Background

Flash Memory



Flash Memory

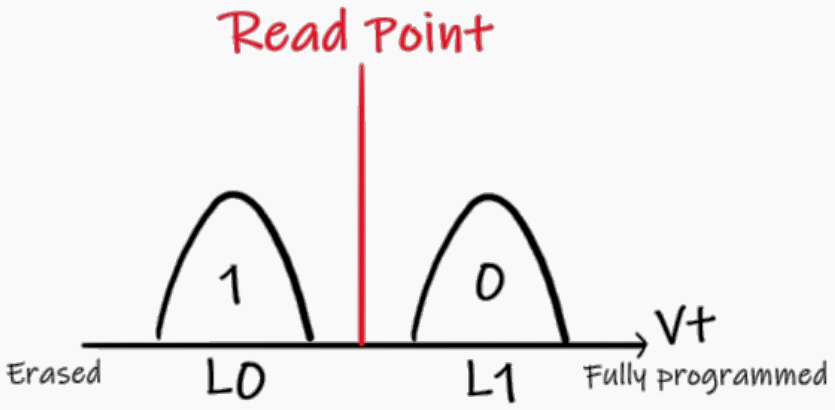
SLC
1 Bit

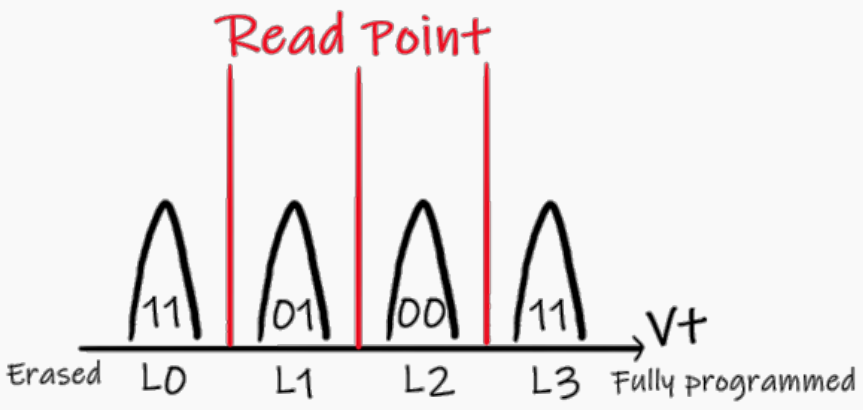
MLC
2 Bit

TLC
3 Bit

QLC
4 Bit

PLC
5 Bit

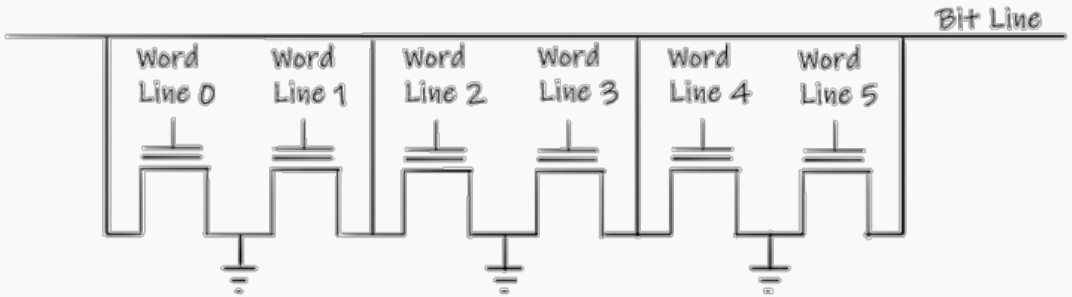




NOR and NAND Flash

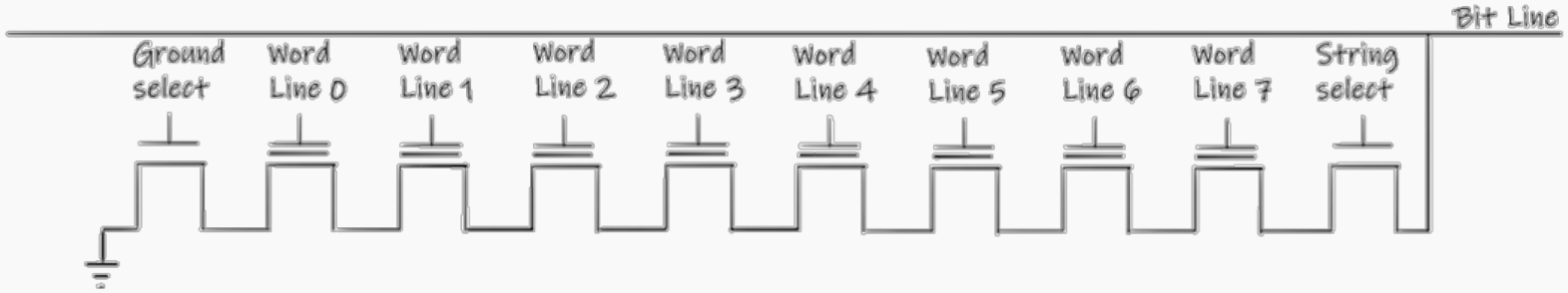
Input1	Input2	Output
0	0	1
0	1	0
1	0	0
1	1	0

NOR

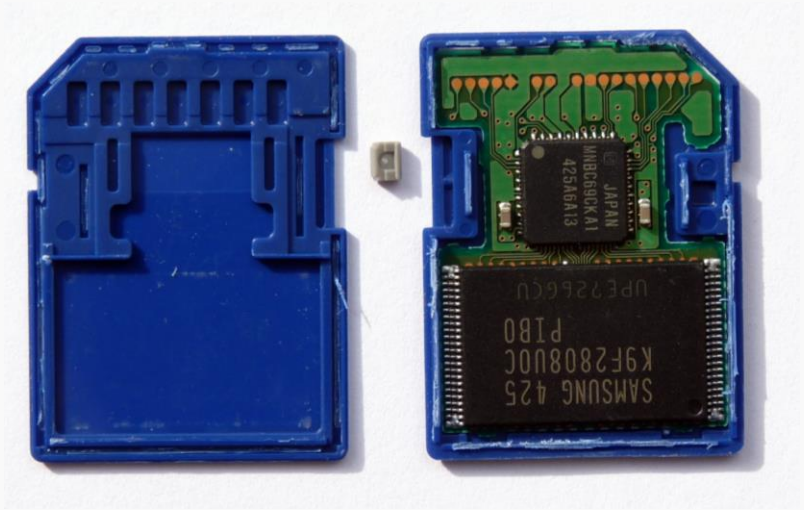


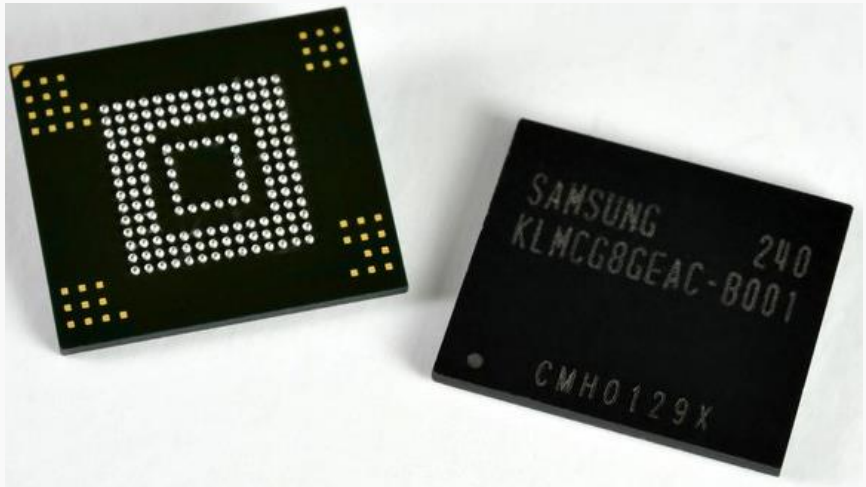
Input1	Input2	Output
0	0	1
0	1	1
1	0	1
1	1	0

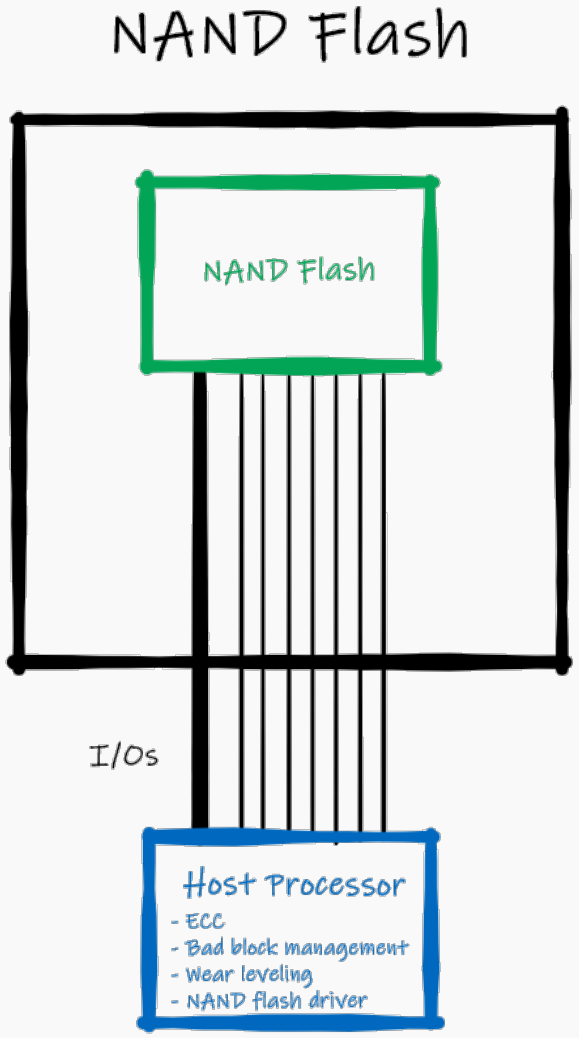
NAND

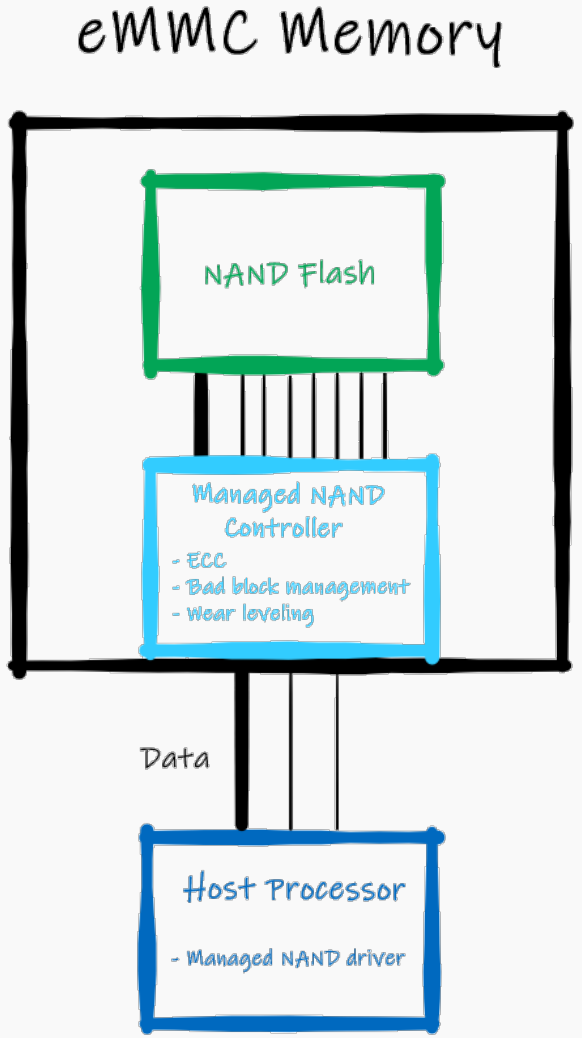


MMC, eMMC and ONFI



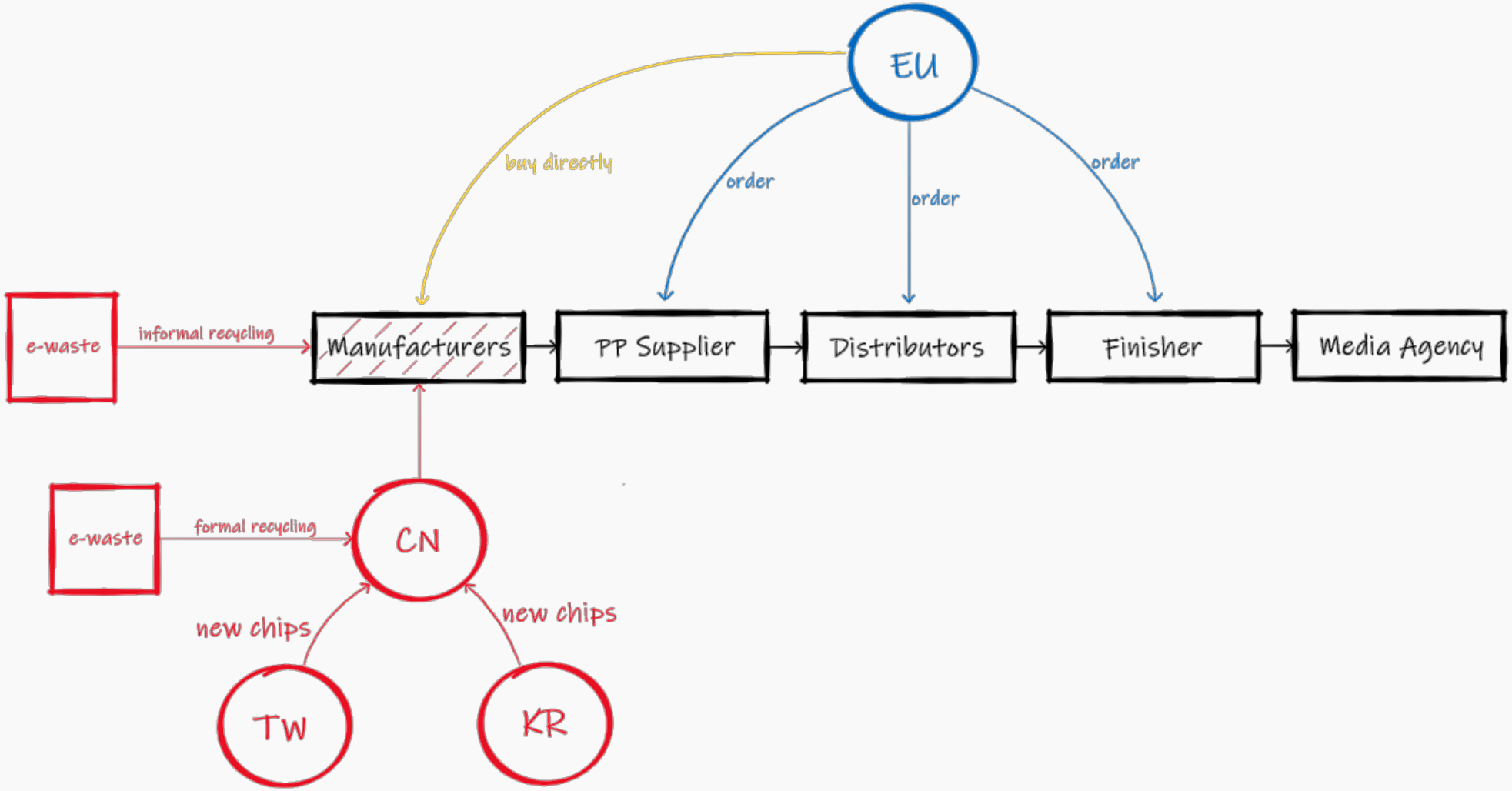






Top: Fritz Jörn, CC BY 3.0 <https://creativecommons.org/licenses/by/3.0>, via Wikimedia Commons, SDKarteoffen.jpg
Bottom: "File:EMMC.jpg" by Toniperis is licensed with CC BY-SA 4.0. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/4.0

The promotional USB flash drive market



Study Design

Recap

- What is the risk of acquiring evidence on new USB drives that are due to former usage (of components) of the drive, i.e., usage before it was bought?
- What factors influence the probability for the existence of non-trivial user data on new USB drives?

Recap

- Crucial point:
  - Chip Recycling
- Thus:
  - What factors influence the probability of getting a new USB drive including a recycled chip?

USB drive acquisition

- Hypothesis:
  - Cheap and low-quality USB drives have a higher chance of including recycled chips
- Thus:
  - Very cheap drives directly from manufacturers via Alibaba
  - Primary factors are cost and capacity

USB drive acquisition

- Group FAU:
  - 500 drives from 10 different suppliers, 50 drives per supplier
  - Five 4 GB batches, five 2 GB batches
  - Additional 16 GB batch with 16 drives from MSAB
- Group Leiden:
  - Three 4 GB batches of different sizes
  - Additional 16 GB batch with 14 drives from MSAB
- Group Albstadt:
  - Three 2GB batches, two 200 size batches and one 100 size batch

Measurements

- Non-trivial user data:
  - Data is non-trivial user data if it is clearly distinguishable from random data, e.g., viewable as an identifiable photographic image
- Drive ID
- Supplier ID
- Analyst pseudonym
- Date imaged
- Drive size in GB
- SHA256 and MD5 hash
- NAND technology
- eMMC / ONFI standard
- Chip manufacturer
- Visual features
- Data found (yes/no)
- Category of data found

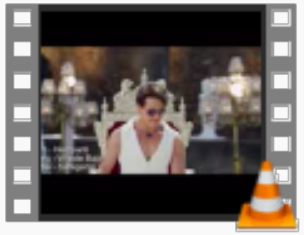
Results

Overview

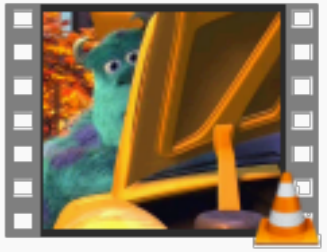
	Group FAU	Group Leiden	Total
Total	516	134	650
Analyzed	484	130	614
Data found	61	14	75
Probably Recycled	59	14	73

Types of found data

- Gifs, icons, emojis and logos
- Photos, pictures, wallpapers, maps
- Music, film and series covers and posters
- Ringtones
- RPM, TAR and ZIP archives
- Music, Videos and Movies
- Speech recordings
- Documents
- Source code

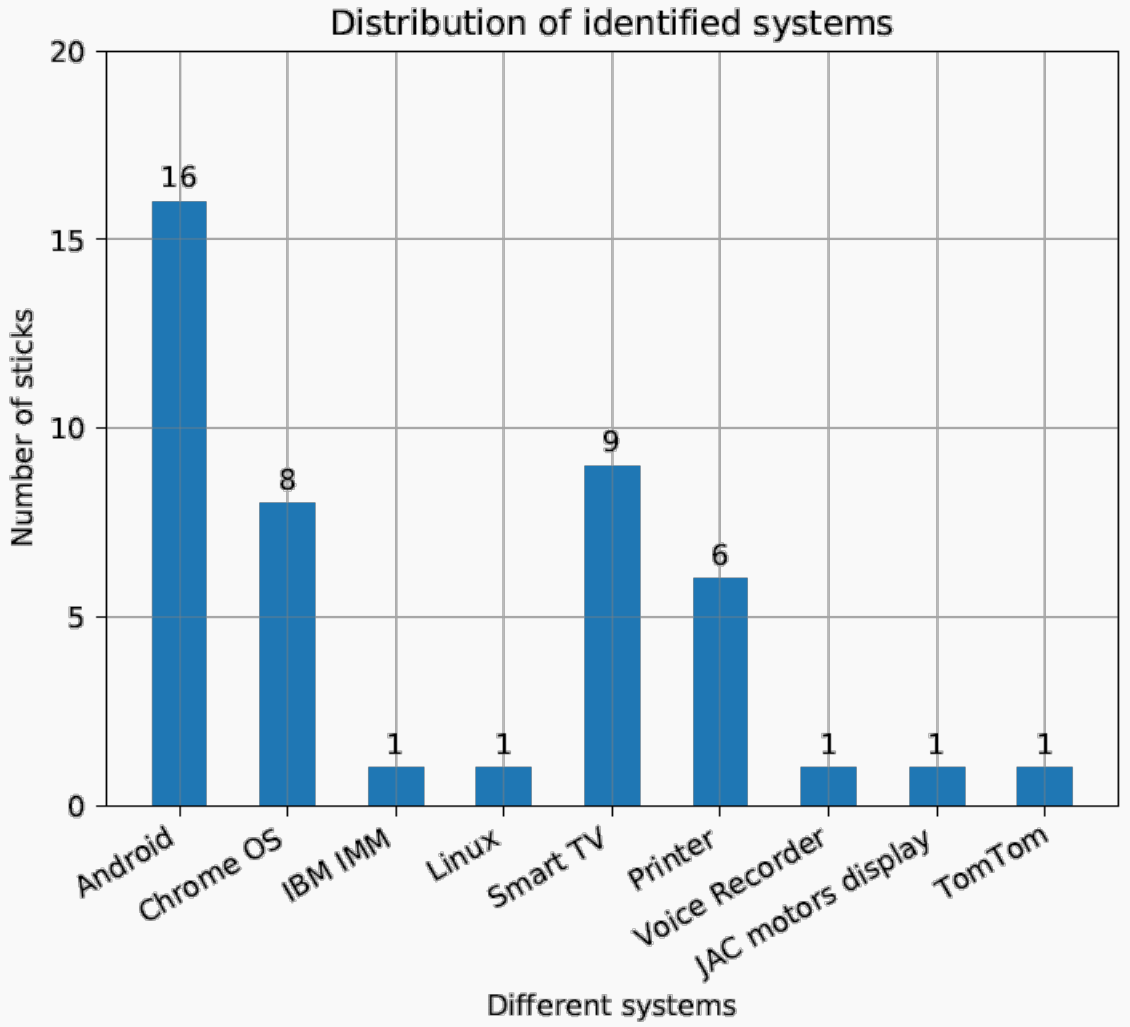


01213496.mp4

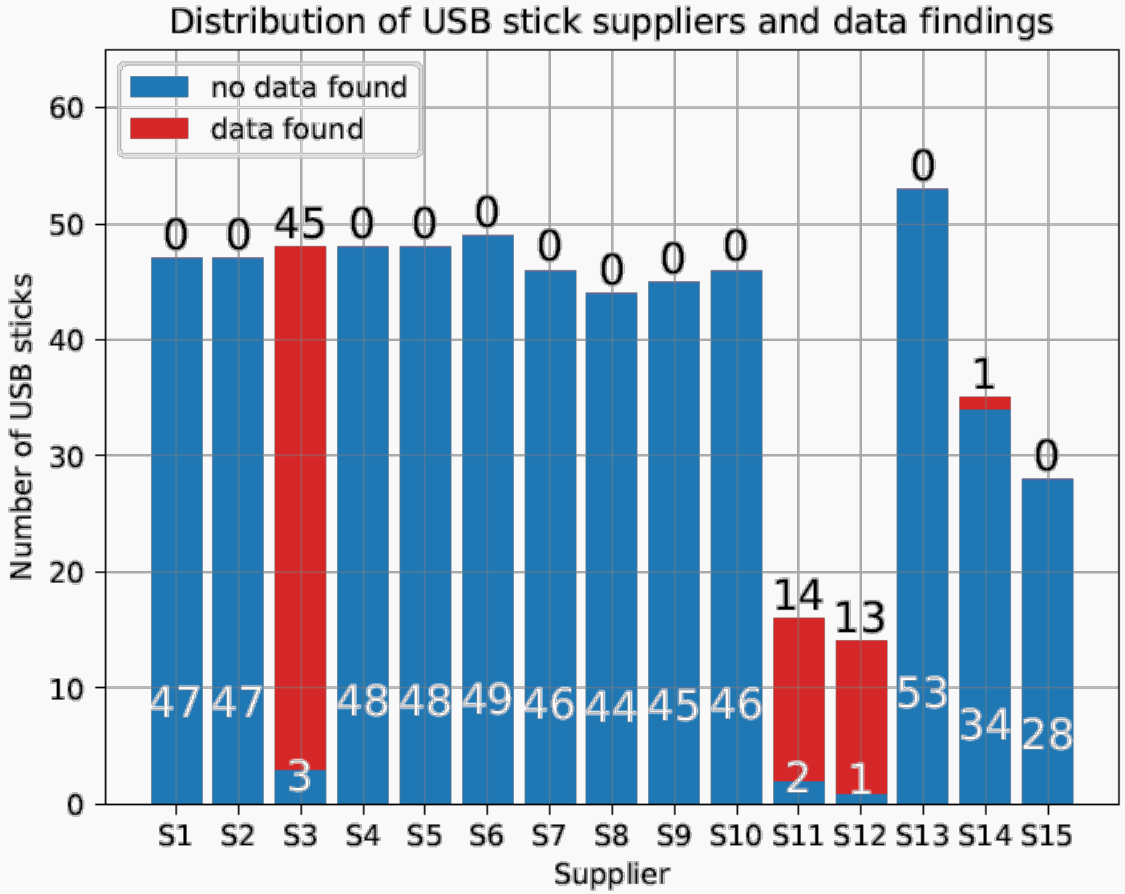


00002614.wmv

Types of found data

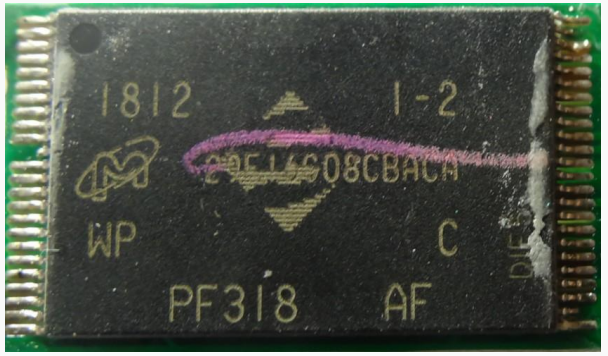


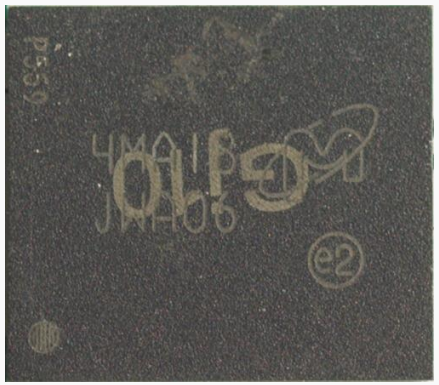
Distribution of data findings

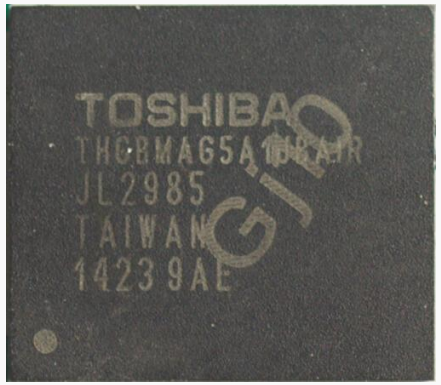


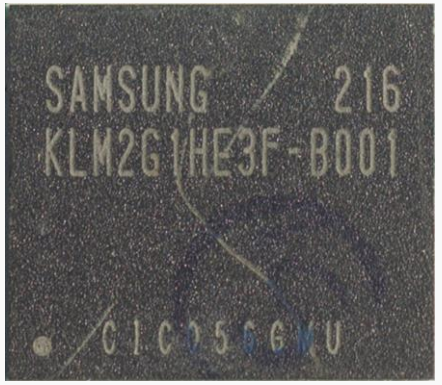
Visual Inspection

- Regular serial and manufacturer specific numbers and inscriptions
- Regular manufacturer logos
- Glue and Epoxy
- Scratches
- Dirt
- Flux
- Paint
- Irregular stamps
- Handwritten notes







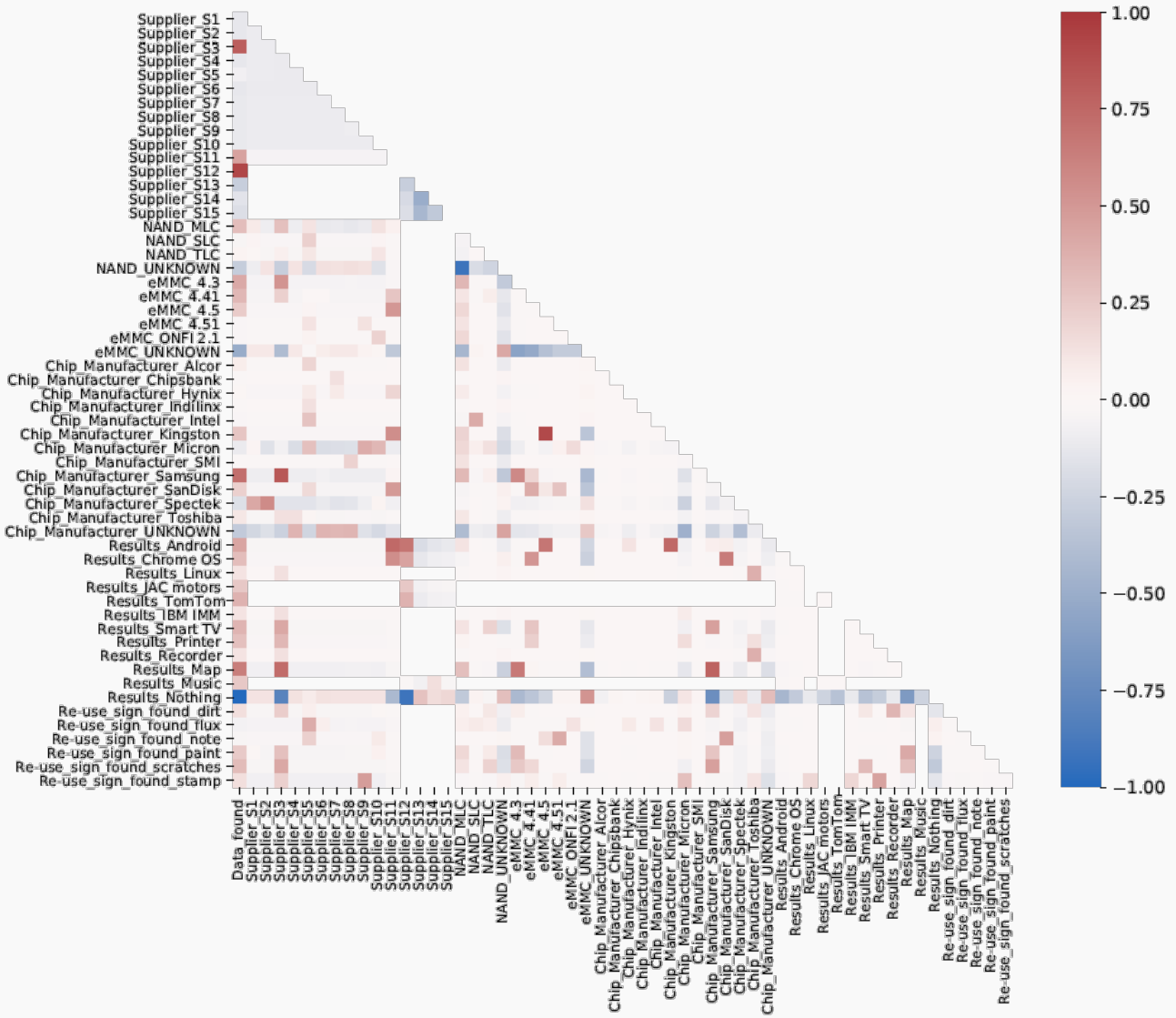


Visual Inspection

	Irregular stamps	Paint	Notes	Dirt	Scratches
Data found	9	7	0	4	8
No data found	19	3	6	4	0
Total	28	10	6	8	8

Correlation

- No correlation between:
  - Data findings and NAND technology, standard or signs of re-use
- Correlation between:
  - Data findings and Samsung (0.716)
  - Samsung and finding of maps (0.777)
  - Kingston and Android (0.742)
  - SanDisk and Chrome OS (0.641)



Conclusion

- Non-negligible probability (12 %) of finding data on cheap but new USB drives ordered for promotional products
- Clear indication of weak sanitization practices in this market sector
- No clear correlation between external factors and the existence of user data
- Focus only on cheap drives
- Follow-up study with higher priced USB drives and more quality-oriented brands
  - This requires a substantially higher research budget, thus we are still **looking for possible sponsors/donors** for this experiment
  - **If interested please contact me :)**

Thank you!