Monolith

Migrationless long term digital storage for archives

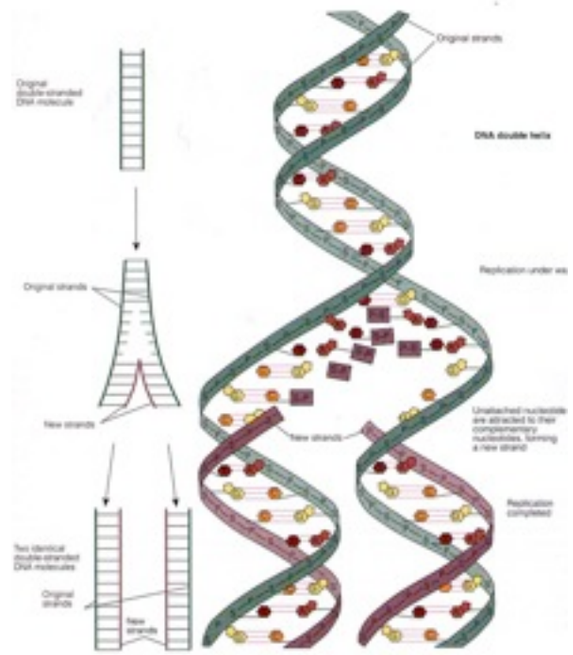
Peter FORNARO and Rudolf Gschwind
Imaging & Media Lab
University of Basel

550 JAHRE WISSEN BEWEGT UNS

UNI BASEL

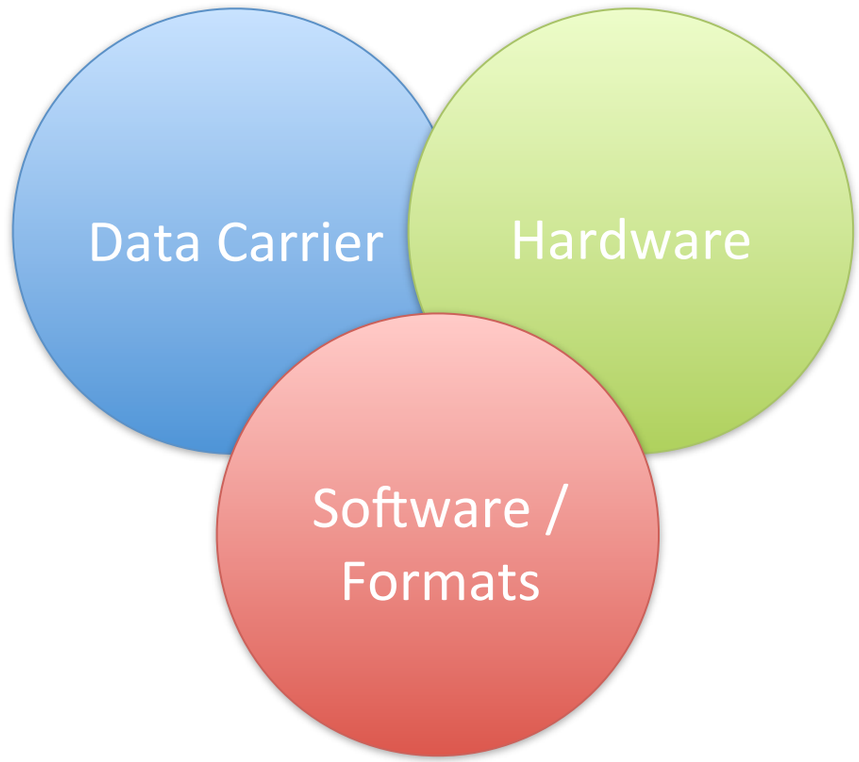
Digital Data is ideal for archival purposes:

- Digital Data can be **copied without any loss**
- **Error correction mechanisms** can be applied
- Digital data can be stored on very **different media**
- **Nature shows it works**: cell reproduction based on DNA (two bit code)



Challenge:

- Data carrier age with time
- Hardware is getting obsolete (short live cycle of IT components and systems)
- Data formats change



...this has prompted my observation that digital information lasts forever - or five years, whichever comes first.

Jeff Rothenberg

3‘892‘179‘868‘480‘350‘000‘000 Bits

~ 4‘865‘220‘000 Terabytes

IDC Digital Universe White Paper, 2008

“Every minute, 13 hours of video are uploaded on YouTube.”

Doug Garland, Google, 2008

"By late 2008 or early 2009, the cumulative number of camera phones shipped will surpass the cumulative number of both conventional and digital cameras shipped in the entire history of photography."

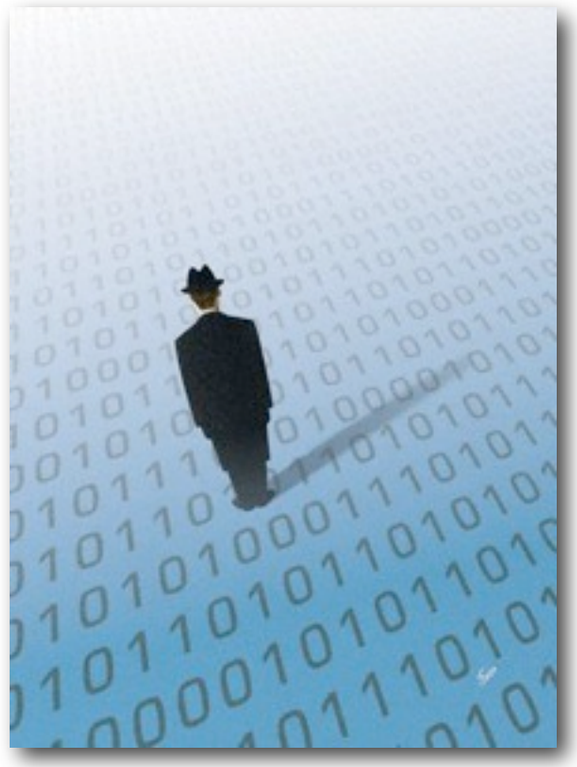
Lyra Industry Reports

The job of archiving digital motion picture has been successful **after we have watch it somewhen in future**.

What is “Digital Data”?

- Any information recorded in a code based on a limited set of symbols
  $S = \{s_1, s_2, s_3, \ldots, s_n\}; n \geq 2$

- Most of it converted from an analog signal as a function of location and/or time:
  $F(x,y,z,t) \rightarrow s_i, s_k, \ldots \in S$

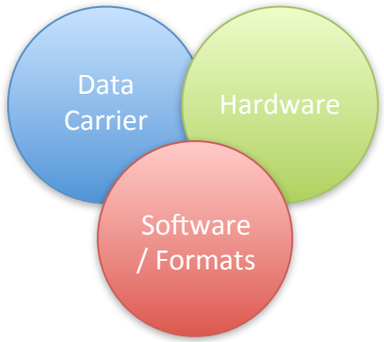


“Binary” Data?

- A special case of “digital data”
  - S={0,1}
  - S={TRUE, FALSE}
  - S={+5V,-5V}
  - S={↑, ↓}
- Most simple set of symbols
- Easy to implement with semiconductors
  => easy to process with computers

Concept of the a Digital Storage without Migration

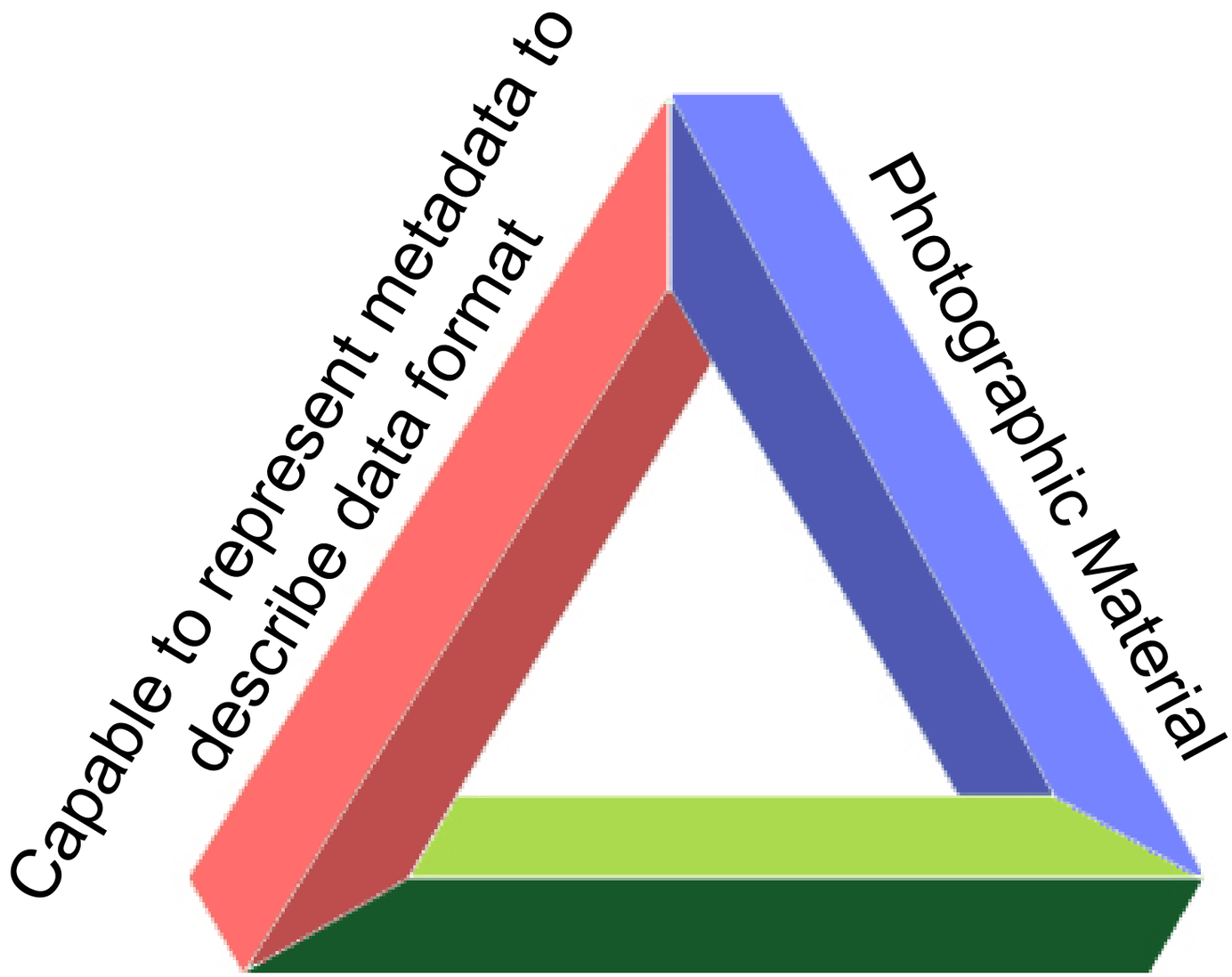
- Example: The book
- Written in a well known and distributed format (language)
- Stored on a permanent data carrier
- Simple access to read data due to visuality
  => human readable!



IT Analogy

	Analogue	Computer Technology
Data Carrier	Glassine, Papyrus, Paper	Punched Tape, Photograf. Material, and many others
Format	Latin Greek English	TIFF, AIFF RTF, PDF/A ASCII, XML
Hardware	Eye (visual)	Camera Scanner

Choose the best possible combination

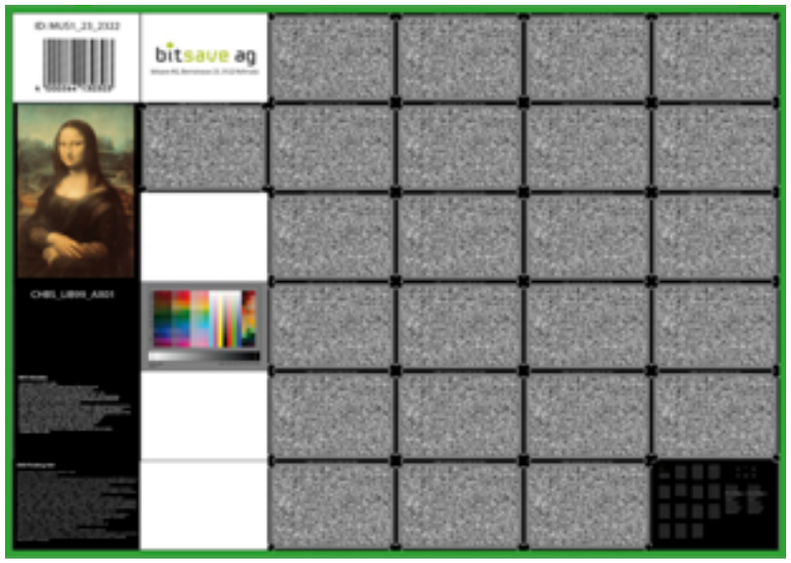


.PDF

.WAV

.J2K

Encode
and
Record



Monolith™ Datafilm (eg microfiche)

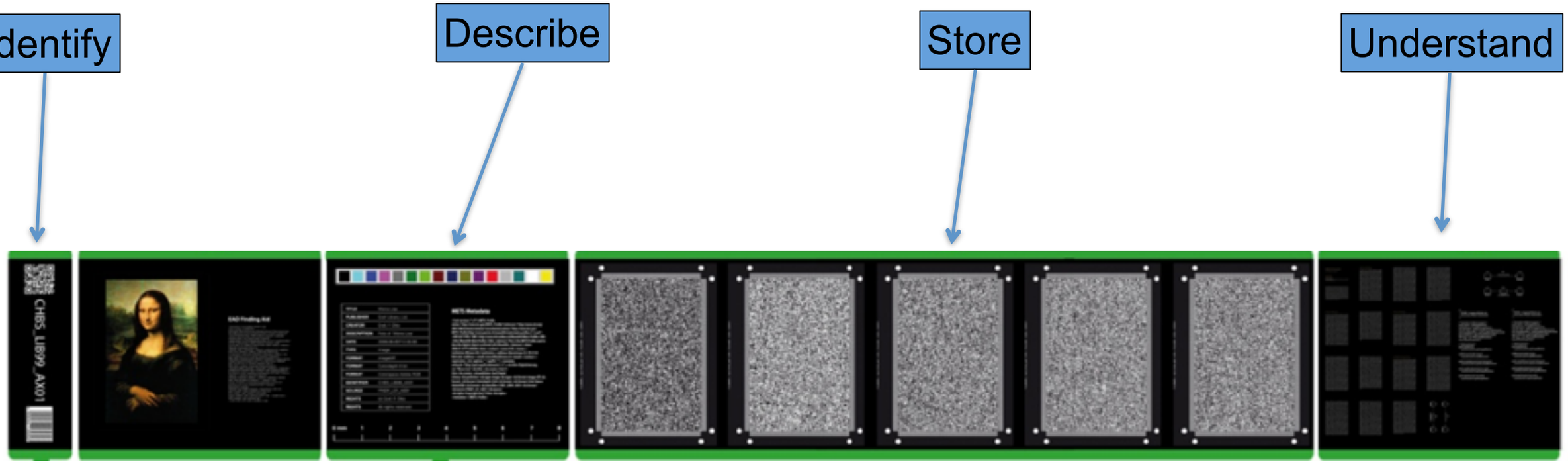
0010010010011001110101010010


CineSafe 4K

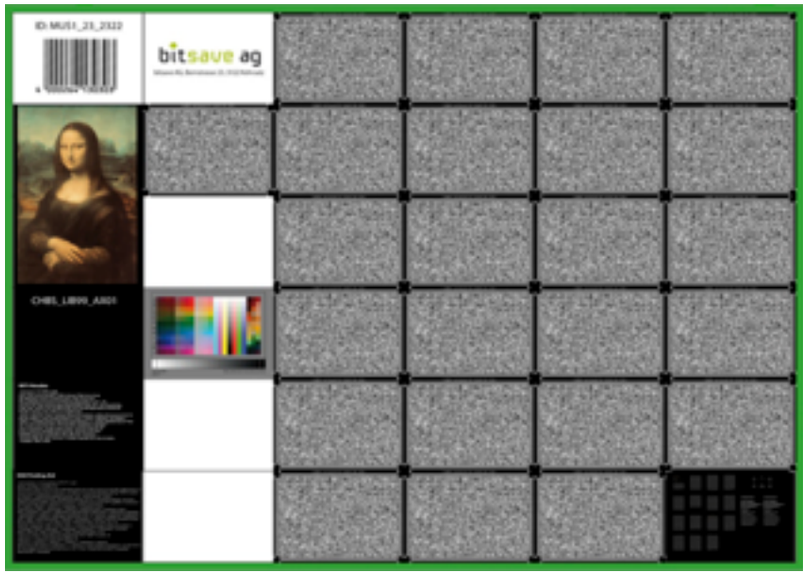

ARRILASER HD


PROARCHIVE
Eternity 105


OP500
ZEUTSCHEL



Monolith: A **hybride** archival storage medium



Monolith™ Datafilm

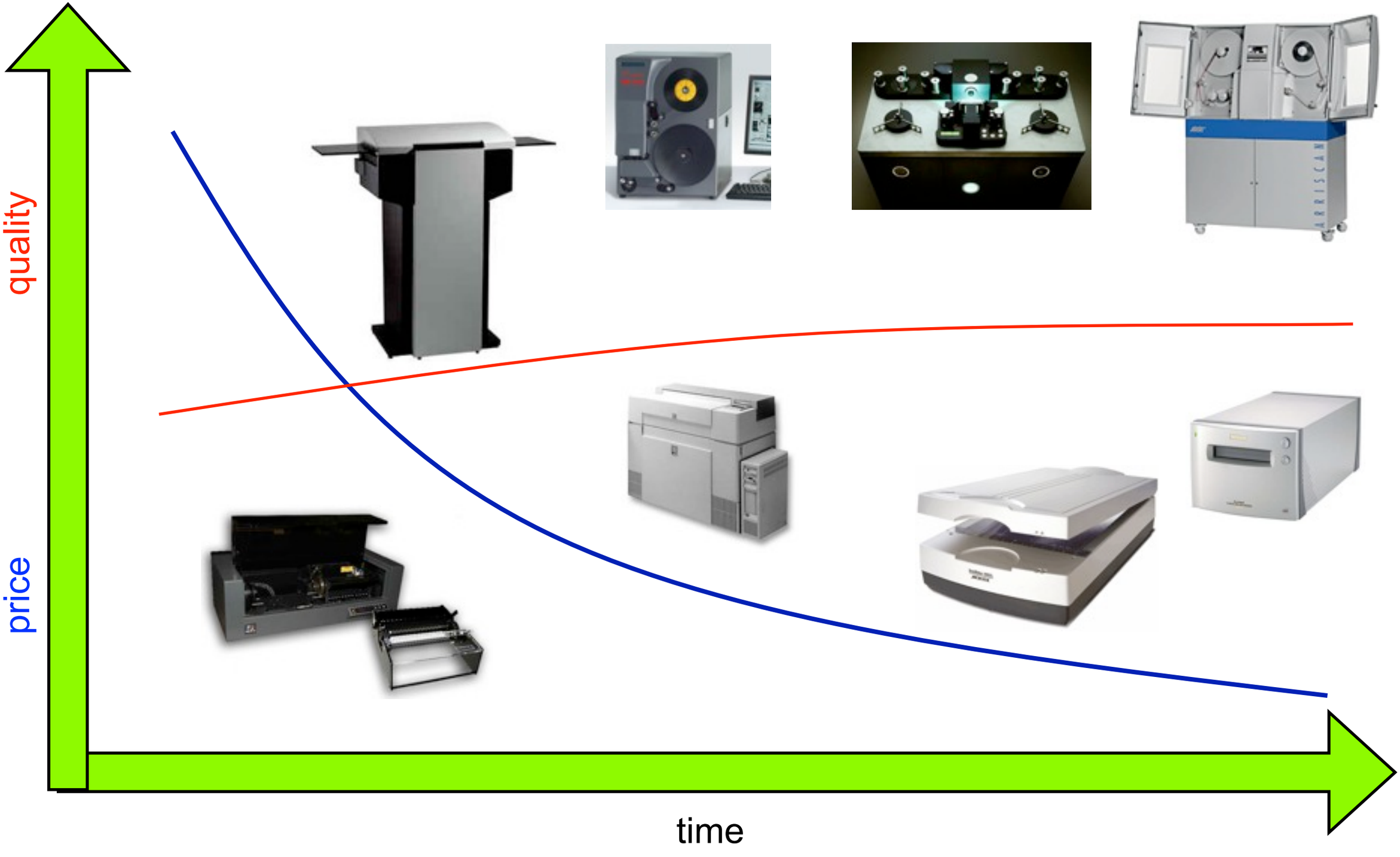
Scan
and
decode

.PDF

.WAV

.J2K

Scanning images is a fundamental technology.

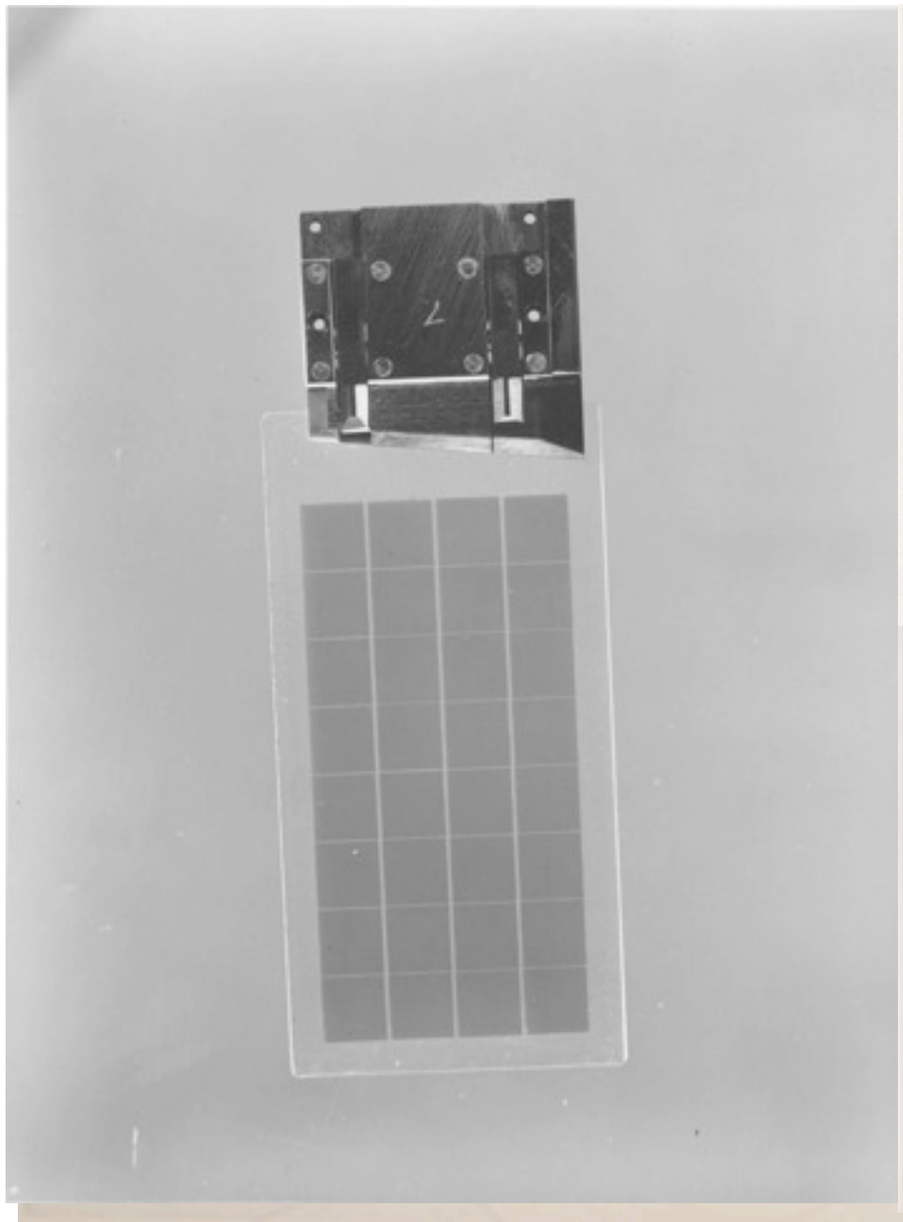


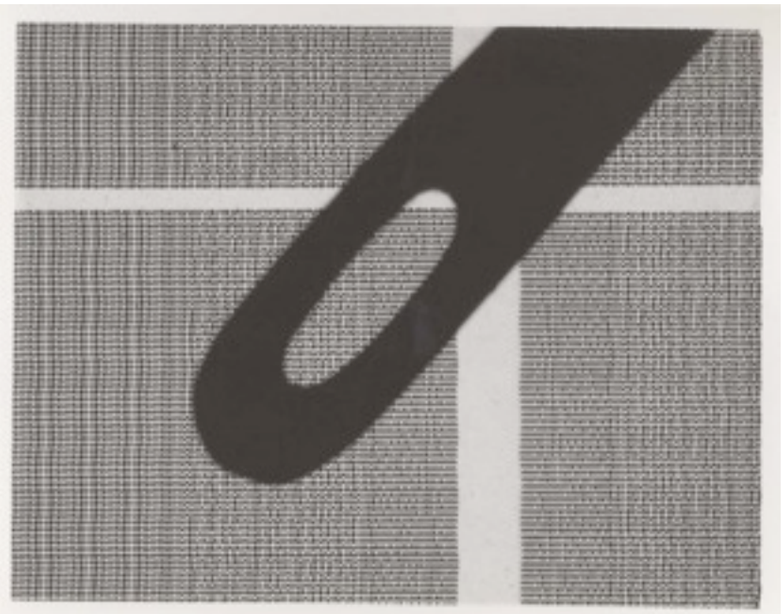
Custom Made Solution

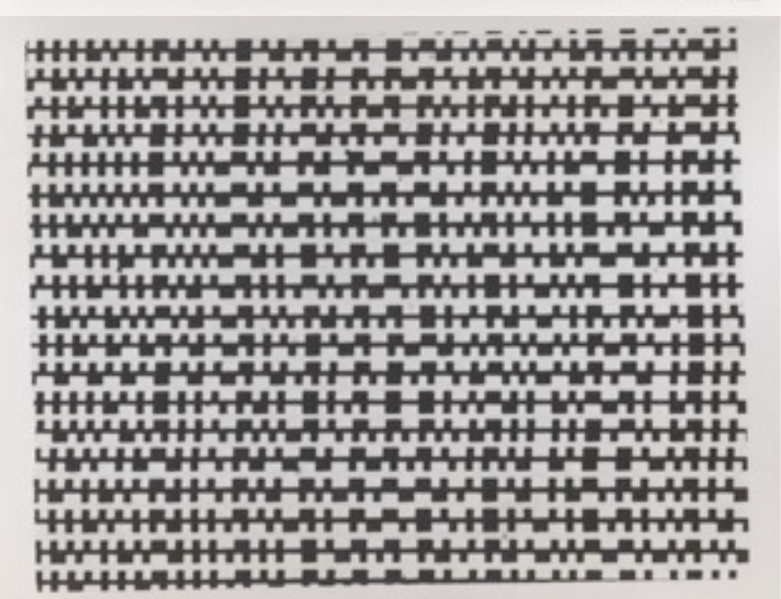
- Custom made scanner based on standard components

Nothing new...

- **1967: IBM Photo-Digital Storage System**

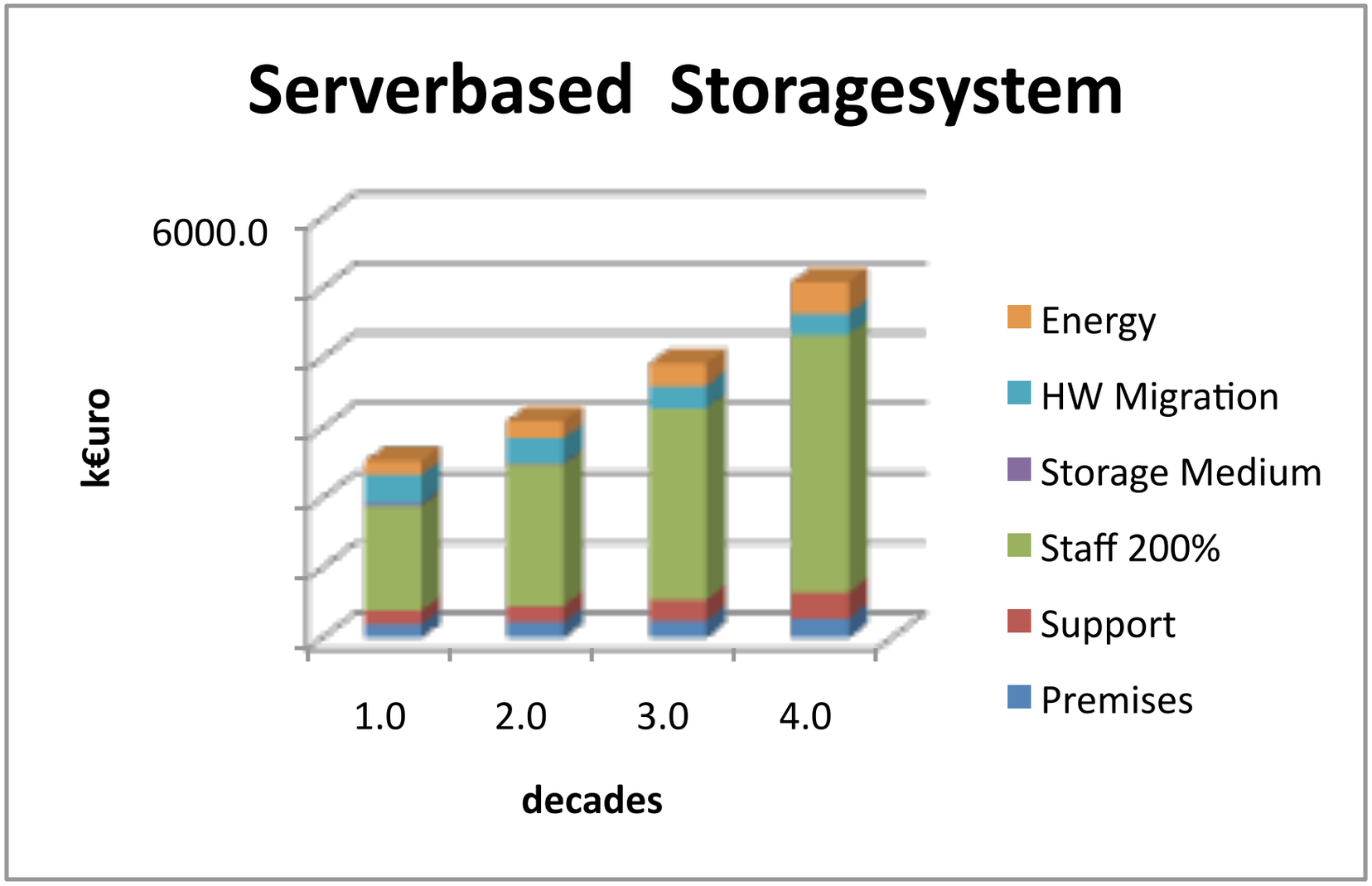






Digital Data in Film Archives

- **Digital data are inevitable** and allow easy access, search and distribution
- The **benchmark for permanence** of assets is given by photographic material => digital data needs to be save for a long time
- With time **migration** is causing **high costs without any value added**



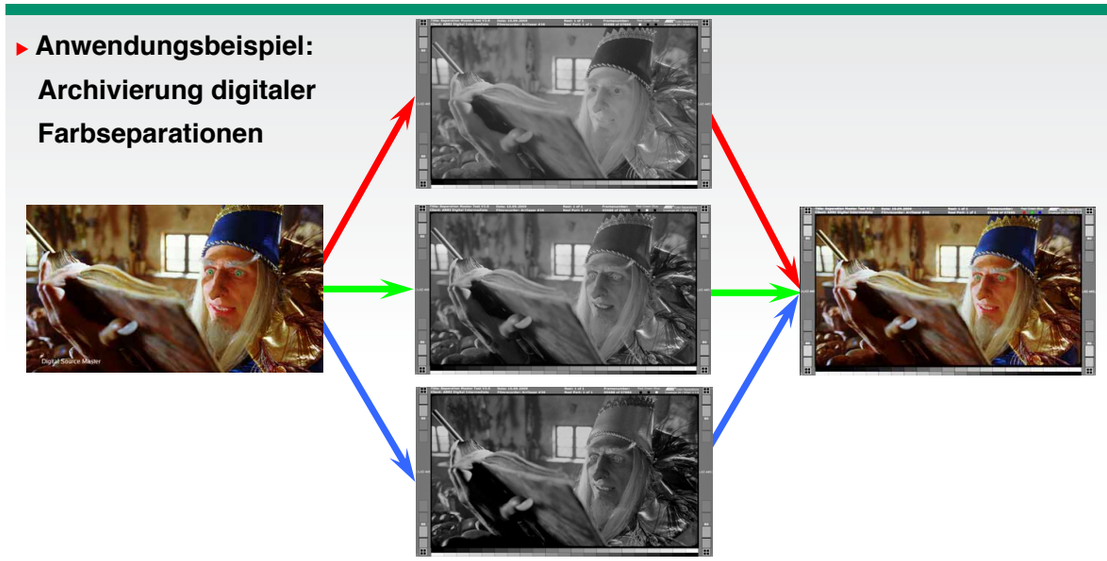
The estimated total costs of a digital archive vary in a large range: Price assumption for 1 GByte of archival data between €0.20 and €3.00 per month can be found.
Fact is: **They are of periodic**!

Possible Solutions

Color Separation

FUJIFILM

- **Anwendungsbeispiel:**
  Archivierung digitaler Farbseparationen



Digital Source Master

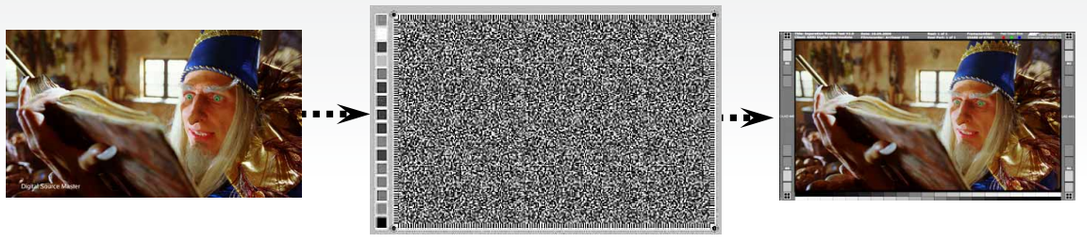
RGB-Farbauszüge auf s/w-Film

Aus Scans der Farbauszüge rekombiniertes Farbbild

Monolith™

FUJIFILM

- **Anwendungsbeispiel: “Bits on Film”**



Digital Source Master

Barcode auf s/w-Film

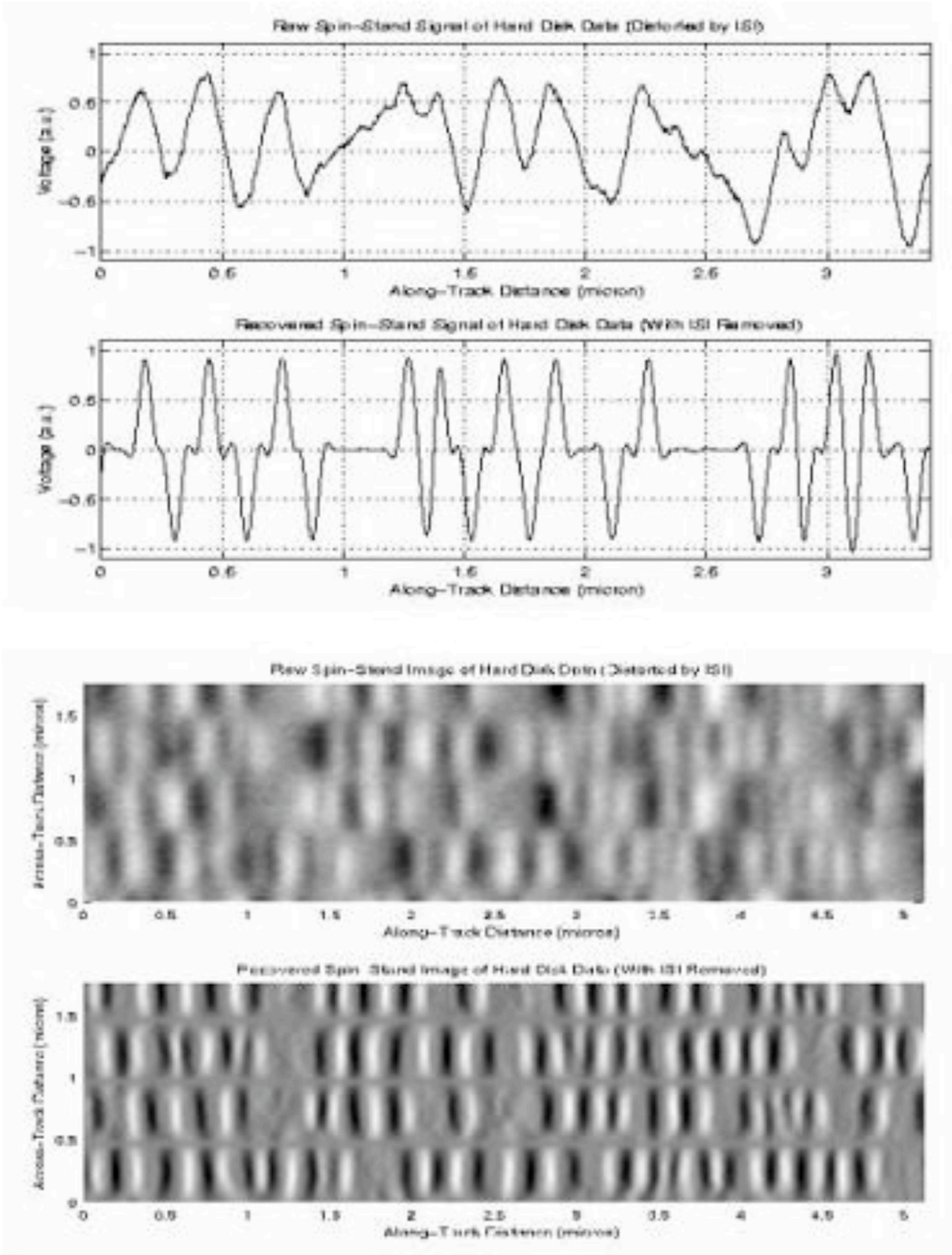
Aus Scans der Barcodes rekonstruiertes Farbbild

Jens Ruppert, Fujifilm RDS Product Presentation 2010

Which approach is better regarding image quality?

Comparison? Digital data do not exist!

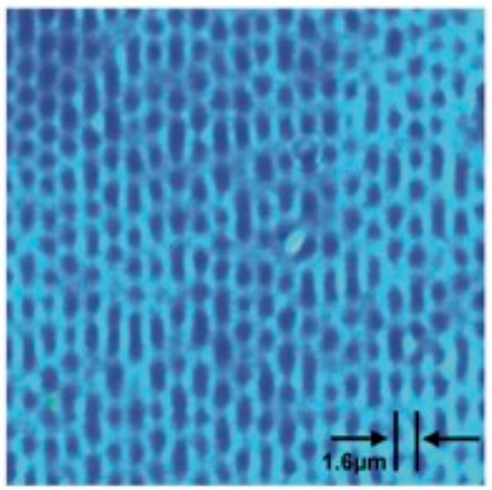
- “digital data” is an ”immaterial, logical concept”
- all material physical recordings are always analogue
- to go to “digital data”, a decision process is required



Analog vs. Digital

- There is nothing like a „digital“ film
- We always deal with an analogue image that may represent digital information

→ We can use the same theory



CD recording pattern

HD recording pattern

Film as Information Carrier: The Theory

PHOTOGRAPHIC SCIENCE AND ENGINEERING
Volume 6, Number 5, September–October 1962
pages 281-286

The Application of Fourier Techniques and Information Theory to the Assessment of Photographic Image Quality

R. SHAW, *Research Laboratories, Ilford Limited, Brentwood, Essex, England*

In the more advanced photographic systems of the type used in satellites, it is essential that the highest possible efficiency is obtained. By using the powerful tools of Shannon's Information Theory, an over-all figure of merit may be obtained for the capacity of a film to receive and store information. This involves joint consideration of the speed, contrast-transfer function, and noise power spectrum of the film. The significance and practical evaluation of this figure of merit are discussed, and examples are given for various film–developer combinations.

C. E. ***Shannon****, Bell System Tech.J ., 27: 379 623 (1948).*

MTF of film

MTF of machinery

$$I = 2\pi \int_C^H \log_2 \left(1 + \frac{T^2(\omega)\, p_0}{n(\omega)}\right) \omega \, d\omega \text{ bits/unit area}$$

Amount of information

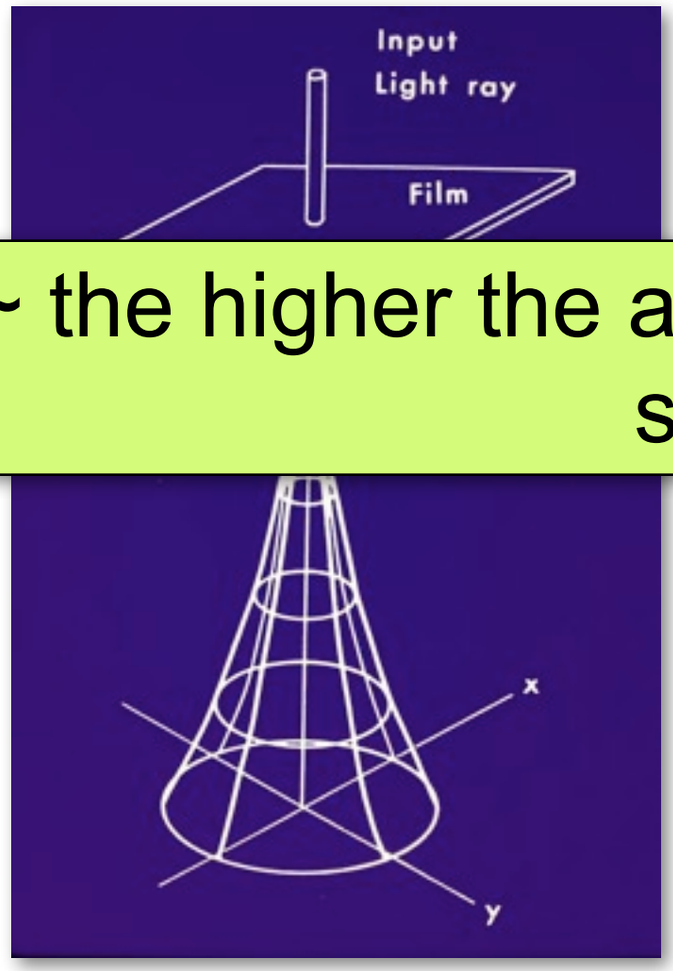
noise (grain)

Information on Film

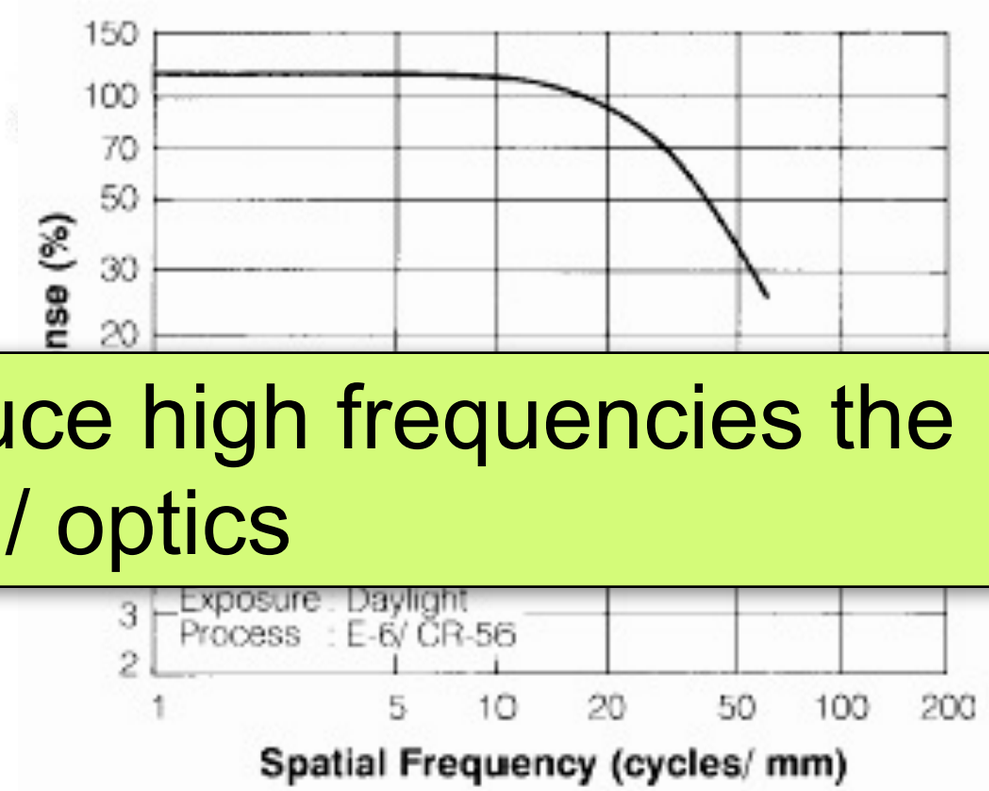
- Reproduction of signals of different frequency
  - **Modulation Transfer Function** MTF
    Decrease the distance of periodically repeated black/white transition
  - **Spectral Frequency Response**
  - **Point Spread Function**
    Describes the response of an imaging system to a point source or point object. A more general term for the PSF is a system's impulse response

Finally all describe the same => Ability of material / optical system to reproduce excitation of different frequency

Frequency Response

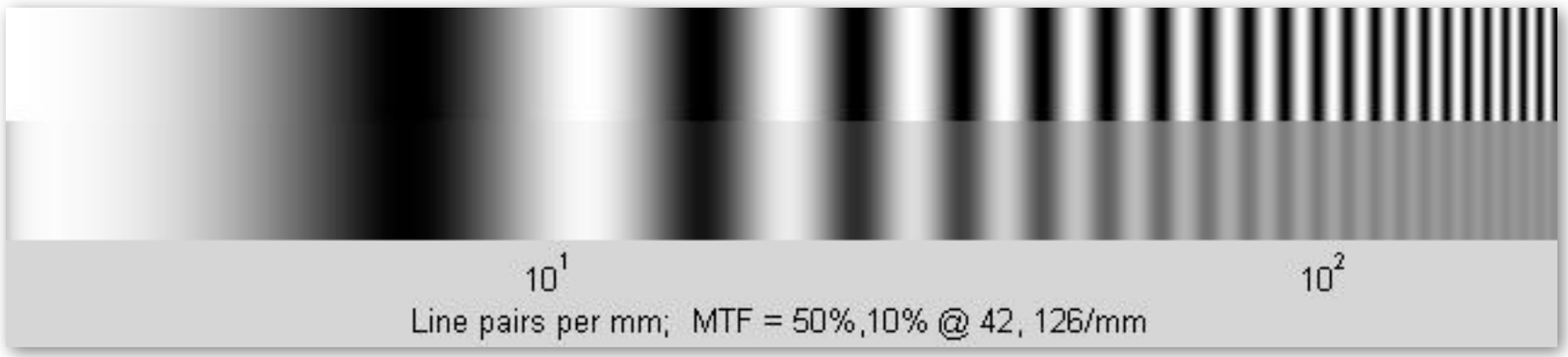


PSF



MTF

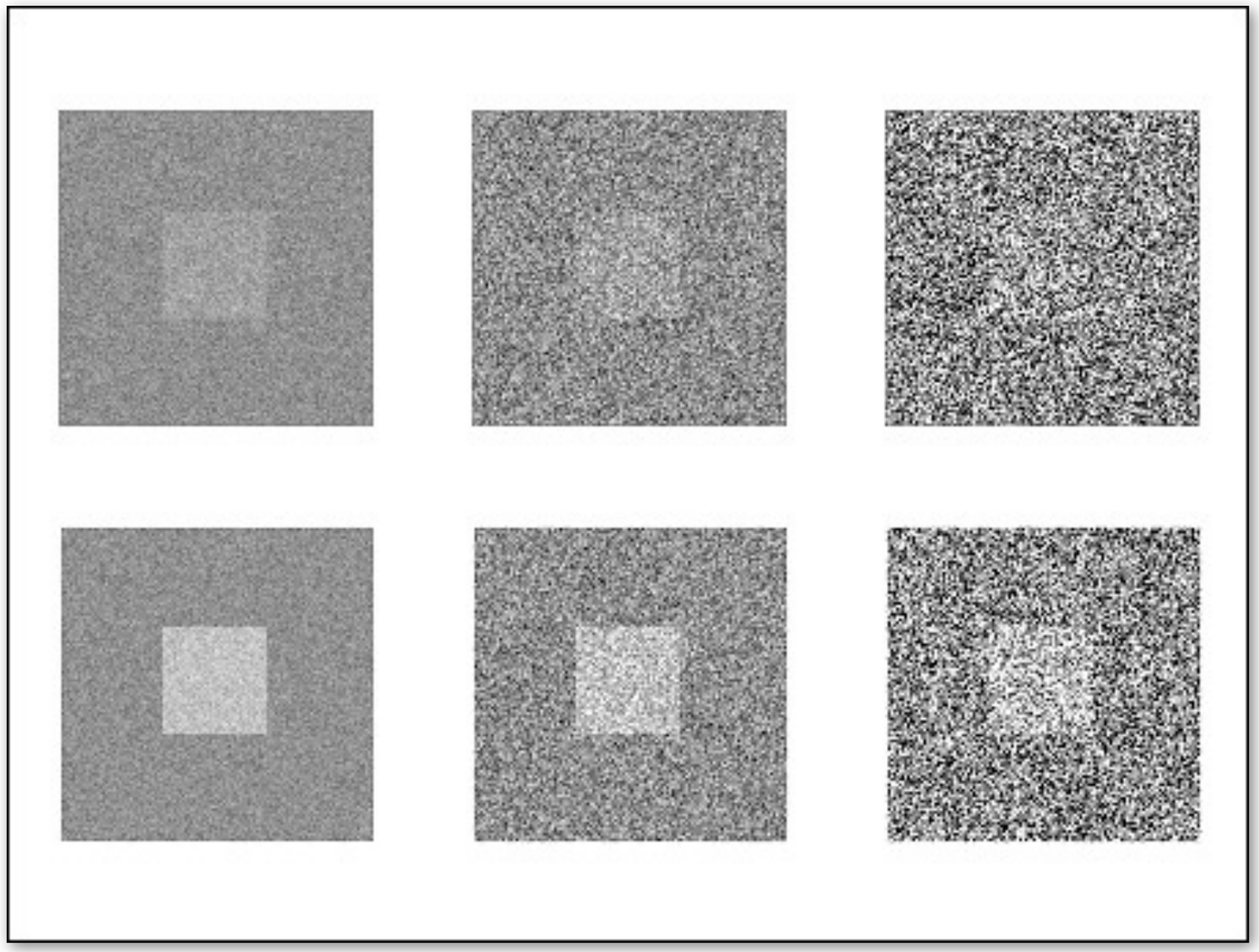
~ the higher the ability to reproduce high frequencies the sharper the film / optics



Information on Film

- Signal to Noise Ratio
  - Quality of the applied signal in relation to the average noise level
  - Noise in photographic film
    Film grain or granularity is the random optical texture of processed photographic film due to the presence of small grains of a metallic silver or color dyes.

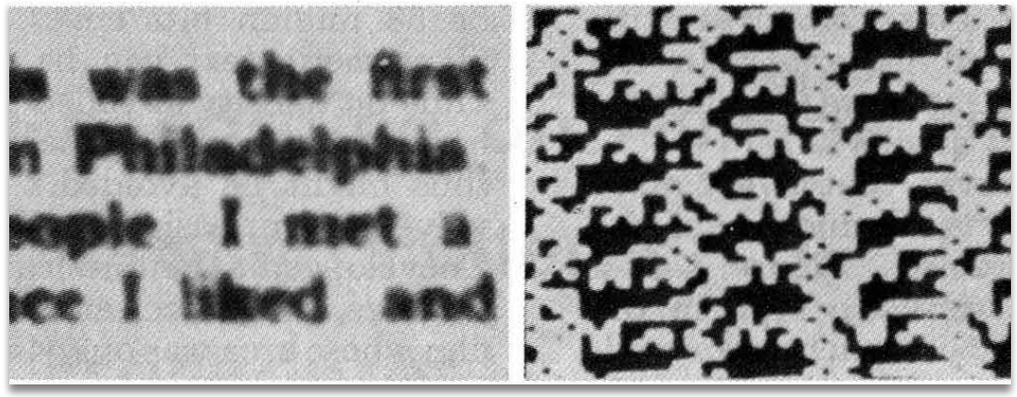
Signal to Noise Ratio in Film



Information Packing

- Alphanumeric text (actually a digital code of an alphabet higher than binary) can be packed at essentially the same bit density as binary coded text

- Maximum information packing on the photographic emulsion occurs for the case of binary recording



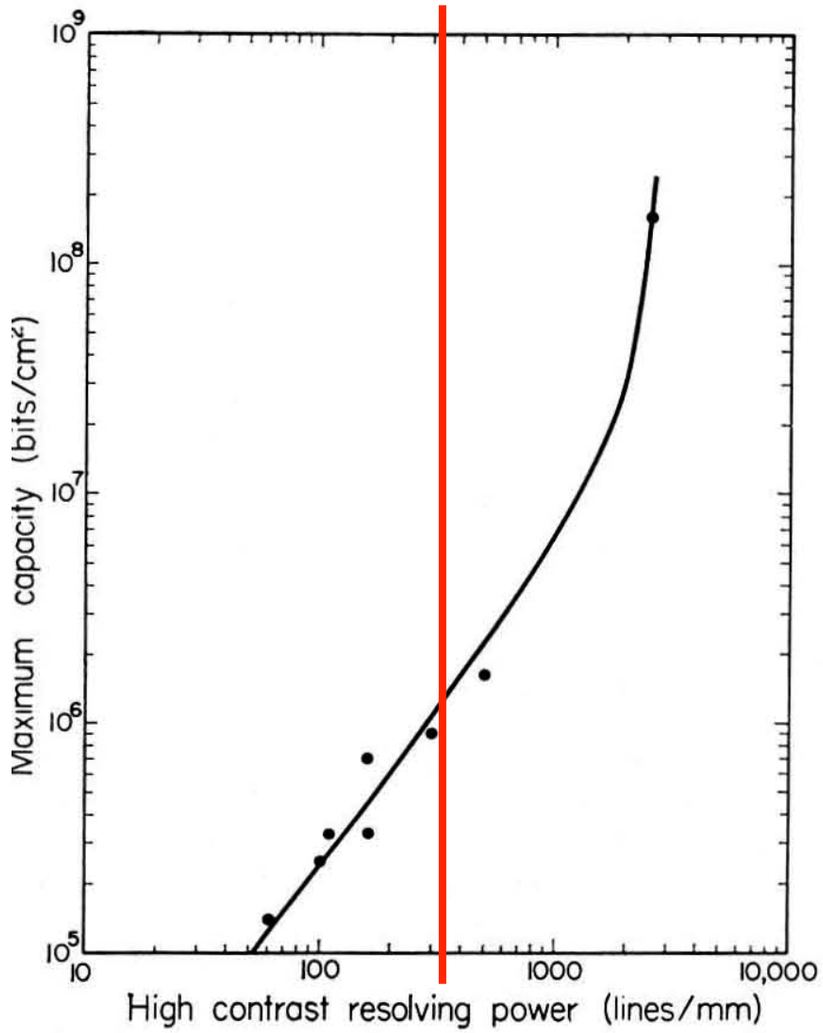
Effect of Spread Function on the Storage of Information on Photographic Emulsions, J. H. ***ALTMAN*** *and H. J.* ***ZWEIG****, PHOTOGRAPHIC SCIENCE AND ENGINEERING Volume 7, Number 3, May-June* ***1963****.*

MTF vs max. Capacity

- Maximum calculated capacity of films as a function of resolving power

- ILFORD CMM
  240 lp/mm

- Fujifilm RDS
  75 lp/mm



Film Stock

- Typical color film material will allow to store
  < 1 MByte per frame (full app) netto information!

Max. Capacity	Film A	Film B	Film C	Film D	Film E
bit/cm2	5.00E+05	7.50E+05	1.00E+06	1.50E+06	2.00E+06
byte/cm2	62'500	93'750	125'000	187'500	250'000
Byte/FullAPP Filmframe	270'000	405'000	540'000	810'000	1'080'000
MByte/FullAPP Filmframe	**0.27**	**0.405**	**0.54**	**0.81**	**1.08**

Conclusion of Comparison

- To record images on film (color or separation) has to be regarded as **analogue lossy data compression** resulting in an image quality reduction comparable to mathematical compression methods

- To store compressed (eg JPG2K) moving images on a specific amount of film stock by the bits-on-film method results in equal or better results than analogue image recording on the same length of material.

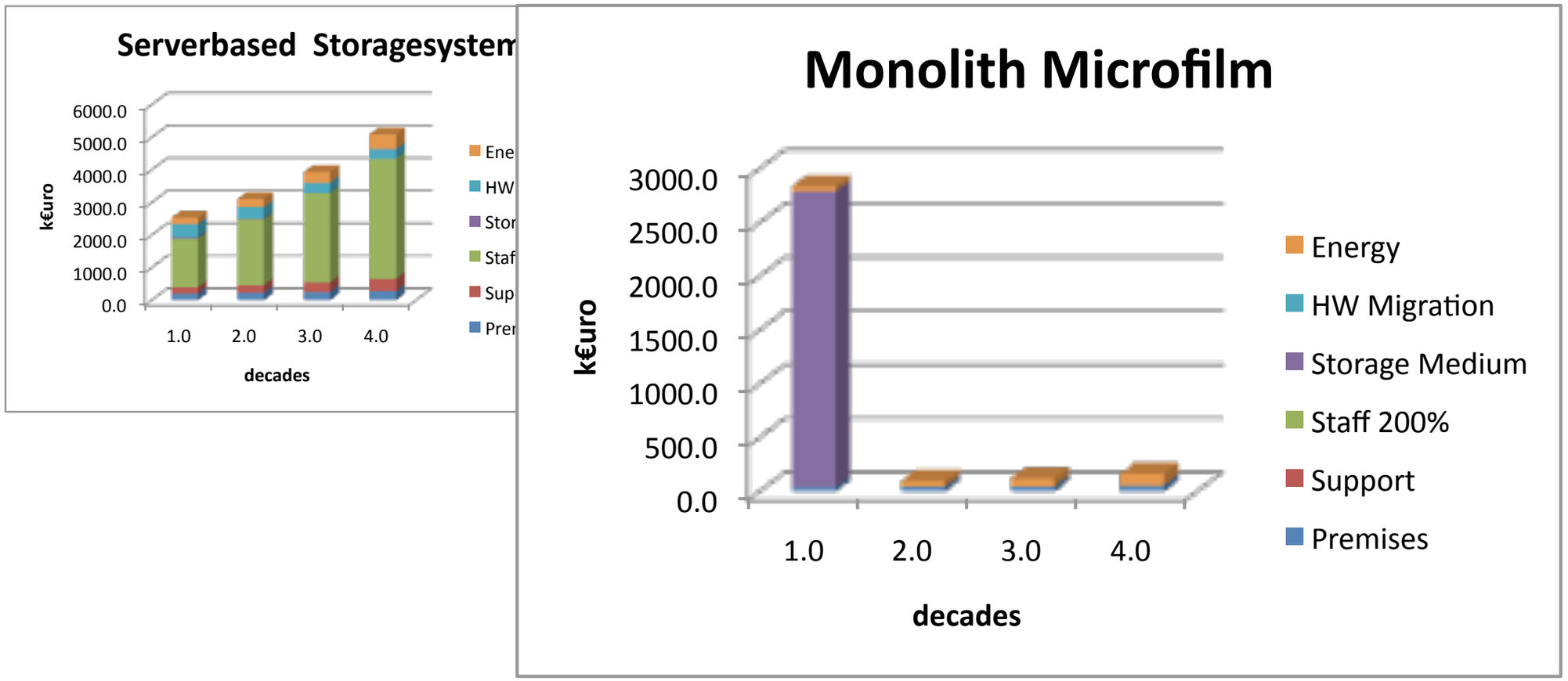
- Quality of the Monolith method can be arbitrary selected (eg. take 1.5 times the stock needed for 90 min of film)

The fact that you create for example a 43 Mbyte 4K image file out of an 35 mm scan does not mean that you get as much information out of it.

Other Advantages

- **Color reproduction** is no longer a problem (no additional color-managment needed, fading of film dyes has no effect)
- **Error correction** can be applied: Dust, scratches are up to an selectable level no problem
- **Fully reproducible**
- **Sound** and **metadata** are included (Digital Cinema Package)

Comparison of Costs



Periodic costs of a film based archival solution are much lower.

Generation

Certified Service Provider

Data of Customers → Data Preparation → Data Transfer → Monolith Encoding → Recording → Validation → Archive

Recovery

Certified Service Provider or Customer

Desaster Occurrence → Determination of Digital Object Digital Film → Scanning → Monolith Decoding → Validation → Data Transfer → Application

Infrastructure

- Cinema will become fully film-less in near future
- Scanners and Filmrecorders will have free production capacity
- Monolith can be easy applied to motion picture post production equipment as well

Robust und Durable, Cost-Transparent

Independent of Proprietary Technology

Free of Migration

Error-Correction, Digital Signature

Hybrid Media (Human and Machine Readable)

Unchangeable, Valid, Authentic

Compliance

- Storage over a predefined time (eg 10 years)
- Probability of access is low!
- Migration is necessary evil!
- Authenticity and validity must assured

Longterm Archiving

- Cost transparent alternative to a n-th server solution

Objects of High Value

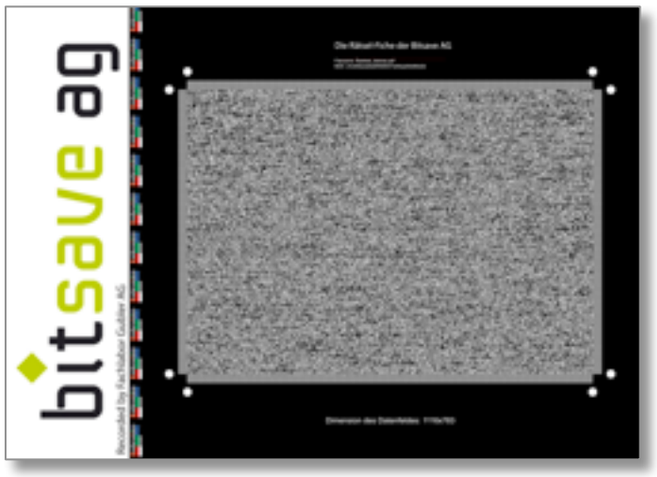
- Contracts, Records
- Digitised Audio, Paintings, Art

Consumer Market

- Your Image Libraries
- Your Audio Libraries

thank you!

... and why don‘t you test it by yourself?



Download the Decoder and the Manual at: www.bitsave.ch/

peter.fornaro@unibas.ch