# An Immutable Solution to the AI Threat to Data Provenance

A DOTS White Paper

GROUP 47

February 6, 2025

## Executive Summary

The age of Artificial Intelligence (AI) has already begun to impact every facet of our collective experience. This affects sources of information, news reporting, knowledge, government, intelligence, military, finance, creative arts, and more.

There is a valid concern by those charged with preserving the integrity of data, that a bad actor, using the power of AI, could alter the data of what should be an uncompromised file representing an original piece of information. That could be a critical document, agreement, judicial evidence, a photograph, video, audio file, basically anything that we trust represents the immutable truth. There is additional concern that an AI bot, either unintentionally or maliciously, could manipulate the data of an original file, further compromising the integrity of one's ability to rely on a file truthfully representing the original. When we refer to "AI data provenance" in this paper, we are referring to the risks posed by AI *against* data provenance and the concerns being brought to bear by the Coalition for Content Provenance and Authenticity (C2PA) and at c2pa.org/faq/.

To be clear, visual and audio media can also be altered by a number of different means in addition to AI. Such as editing, visual effects, audio manipulation, among others.

Yet, in our current environment of reliance on digital storage media that is easily erased, deleted, rewritten, or short lived, we are faced with a world where the integrity of a digital file's provenance cannot be assured. C2PA ensures that we will know if the file has been altered from its original state, but once the original primary source is changed, there is no way to resurrect that original.

Software solutions for proving data provenance with metadata tags are often remarkably creative but, unfortunately, anything that is rewritable leaves open the opportunity for malicious malware or cyber attack AI manipulation of those original files.  C2PA ensures we'll know the original has been changed, but it doesn't prevent that original from being altered.

This white paper will posit that proof of data provenance has always been a hallmark of **DOTS**. Then provide an in-depth analysis and explanation of how **DOTS** exceeds what is needed for protection from AI compromising the provenance of original files.

# Requirements for Immutable Data Provenance from AI

**To ensure that original files can be relied upon as unchanged and represent the original, the storage media must have these characteristics:**

- **The data *must* be saved to a physical media. Software and metadata solutions are all at risk of being tampered with**
- **The data must be saved to WORM media (Write Once Read Many)**
- **The data must NOT be subject to migration in order to preserve 100% of digital files – this includes cloud data maintenance**
- **The mere act of migration puts provenance into question**
- **The data must be saved to an inert, stable archival media**
- **It must be non-magnetic and immune from magnetic fields & EMP**
- **Future proof storage format without operating system dependencies**

## Integrity of Data Provenance from AI Requires an Immutable Physical Media

# Why cloud and data storage options aren't meeting the metrics for data provenance

In the past three decades, the amount of digital data produced has grown exponentially. This data includes both native digital files and legacy analog data that has been digitized.

As a result, the focus has been on ***data density*** rather than preserving and protecting that data from harm.

Most digital data that is stored on magnetic media has an archival life of a few years, unlike historical storage methods (film, paper, etc.) with lifespans of decades or even centuries. Magnetic media is susceptible to hostile electromagnetic environments, mechanical and electronic breakdowns, and, in the case of magnetic tape, delamination, oxidation, and chemical degradation.

It's crucial to emphasize that AI data provenance can only be assured with an immutable and permanent physical solution.

Group 47 has developed the Digital Optical Technology System (**DOTS**), which ensures data provenance using a simple, human-readable media with an archival life exceeding two hundred years. **DOTS** is non-magnetic, immune to electromagnetic fields (including EMP), chemically inert, won't oxidize, and contains no chemical binders that can degrade due to humidity, resulting in data loss.

These unique qualities of simplicity and archival data permanence set **DOTS** apart from all other proposed assurances of data provenance protection from AI.



## Why cloud and data storage options aren't meeting the metrics for data provenance

Data is stored on **DOTS** as a series of dots and blanks, representing ones and zeros. These dots and blanks can be read using low-power white light optics and an imaging system, including the human eye if necessary. Data can also be stored as written text, similar to microfiche, or as a **DOTS** Bit Plane Image which ensures the provenance of an image or sound file can be verified in the distant future without any dependencies on operating systems or file format.

**DOTS** can also store digital data as the binary data of the actual file and the Bit Plane Image on the same piece of media.

**DOTS** ensures data provenance from AI because, at its core, it is designed for long term preservation of data which, due to its value and the inability to recreate it, must not be lost.

**DOTS** has also been designed with the philosophy that data must be retrievable without sophisticated technology. To support this, **DOTS** provides visual text instructions written onto the leader of each tape, describing how the data is written and how to build an automated reader. It is compatible with LTFS and offers a reasonable data density of 20TB per 12" open reel format.

All these attributes make **DOTS** a simple, reasonably priced way to ensure guaranteed data provenance from AI for over 200+ years, along with archival preservation of an organization's most valuable data.

# How Does DOTS Meet Those Metrics?

- **Non-magnetic WORM recording of any digital file**
- **Carnegie Mellon tests conclude well over 200 years stability**
- **Data is recorded *visually* - Literally an image of the data**
- **A dot = One, no dot = Zero**
- **Future proof Bit Plane Image format ensures the provenance of video, photographs, and audio for *centuries***
- **Noncomplex - Data is retrieved with a camera**
- **Immune from hacking, malware & EMP**
- **Store anywhere from -9° to 66° C – Immune from water & petrochemicals**
- **20 TBytes capacity per reel**

**Ensuring provenance of digital files from AI requires preservation in a future proof format**

# The Archivist’s Perspective of Data Provenance

Archivists will tell you that history has proven the ONLY successful way to ensure the provenance of an asset is to use long-lasting media where you can SEE the information.

With visual technologies such as photographic prints or negatives or paper text documents, one can look directly at the medium to access the information and ensure its integrity as the original.

With all magnetic media, complex optical, biologic, or holographic storage, a machine and software are required to read and translate the data into a human-observable and comprehensible form.  If the machine or software is obsolete or lost, the data is likely to be lost as well.

It is critical that the method employed to protect the provenance of data is unencumbered by complicated technology.



From the moment the first magnetic solutions for digital storage were introduced, the lessons of the archivist were not considered. Nor was there any concern about ensuring the integrity and provenance of digital files. Storage companies never ventured to solve the problem, but instead promoted the concept that increasing data density was more important.

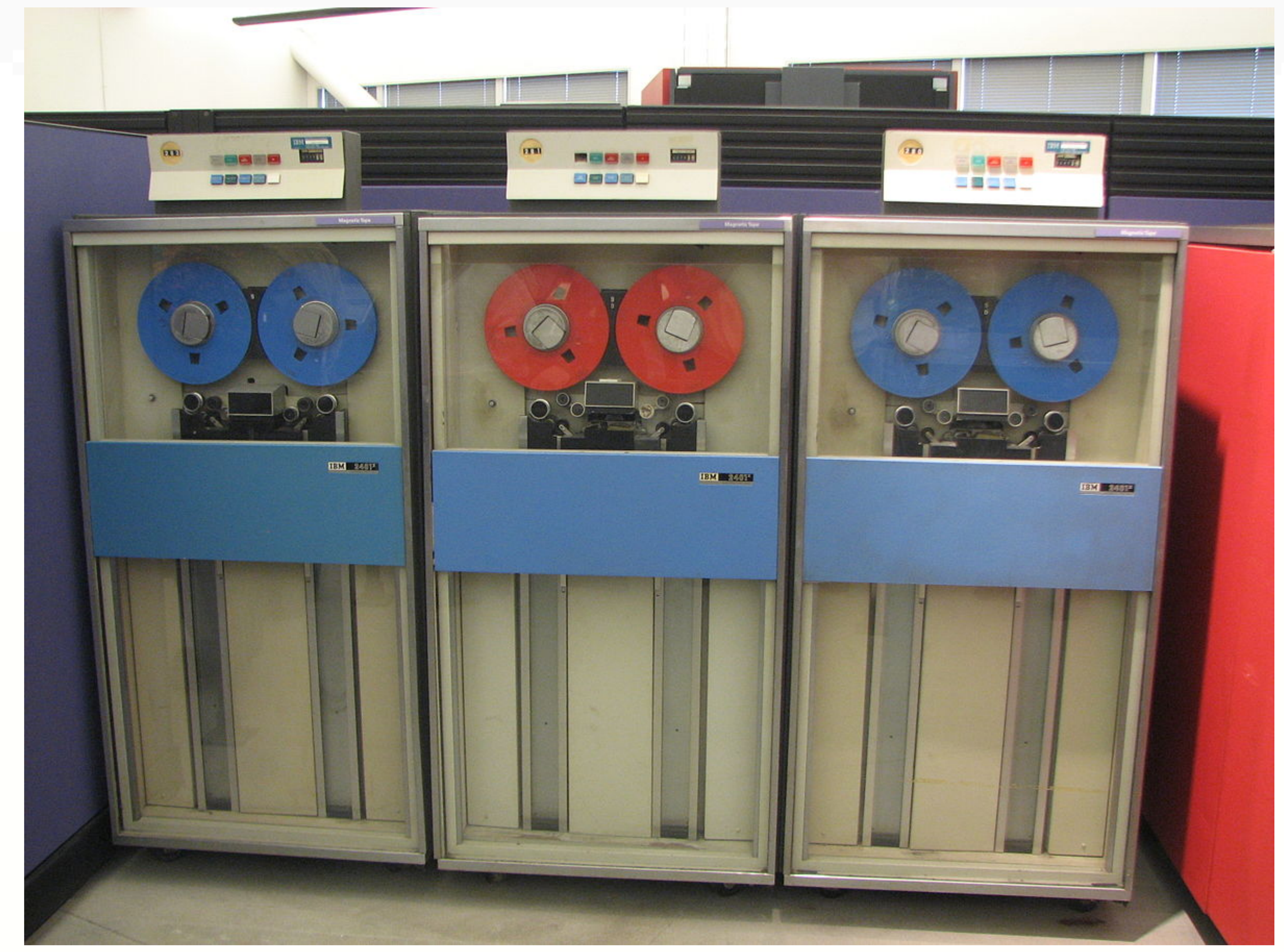
Data density is important and has its place, but it should not be at the expense of a solution that guarantees the provenance of a critical asset and preserving that information which cannot afford to be lost.

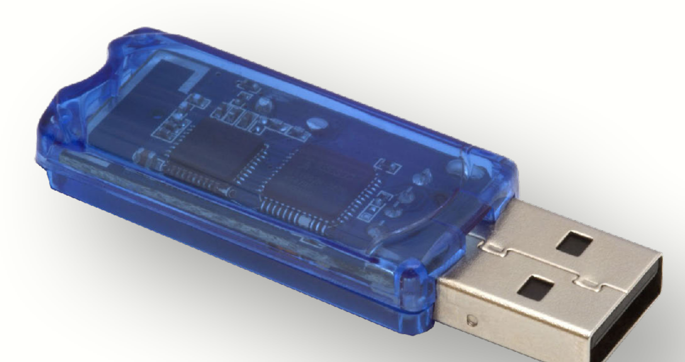
Instead of solving this endemic preservation problem, the imposed solutions from IT professionals and storage providers promoted a belief in something that had never been done previously in history.

That is, forever migration and making copies of copies of copies.

By requiring constant migration and maintenance of data that needed to be preserved, ensuring provenance from AI becomes an impossibility. The associated costs of power and HVAC environmental demands placed additional burdens on preservation and curation of asset provenance that were unprecedented.

There is little impetus to mitigate these costs on the part of the storage and cloud providers because this approach generates significant revenue.





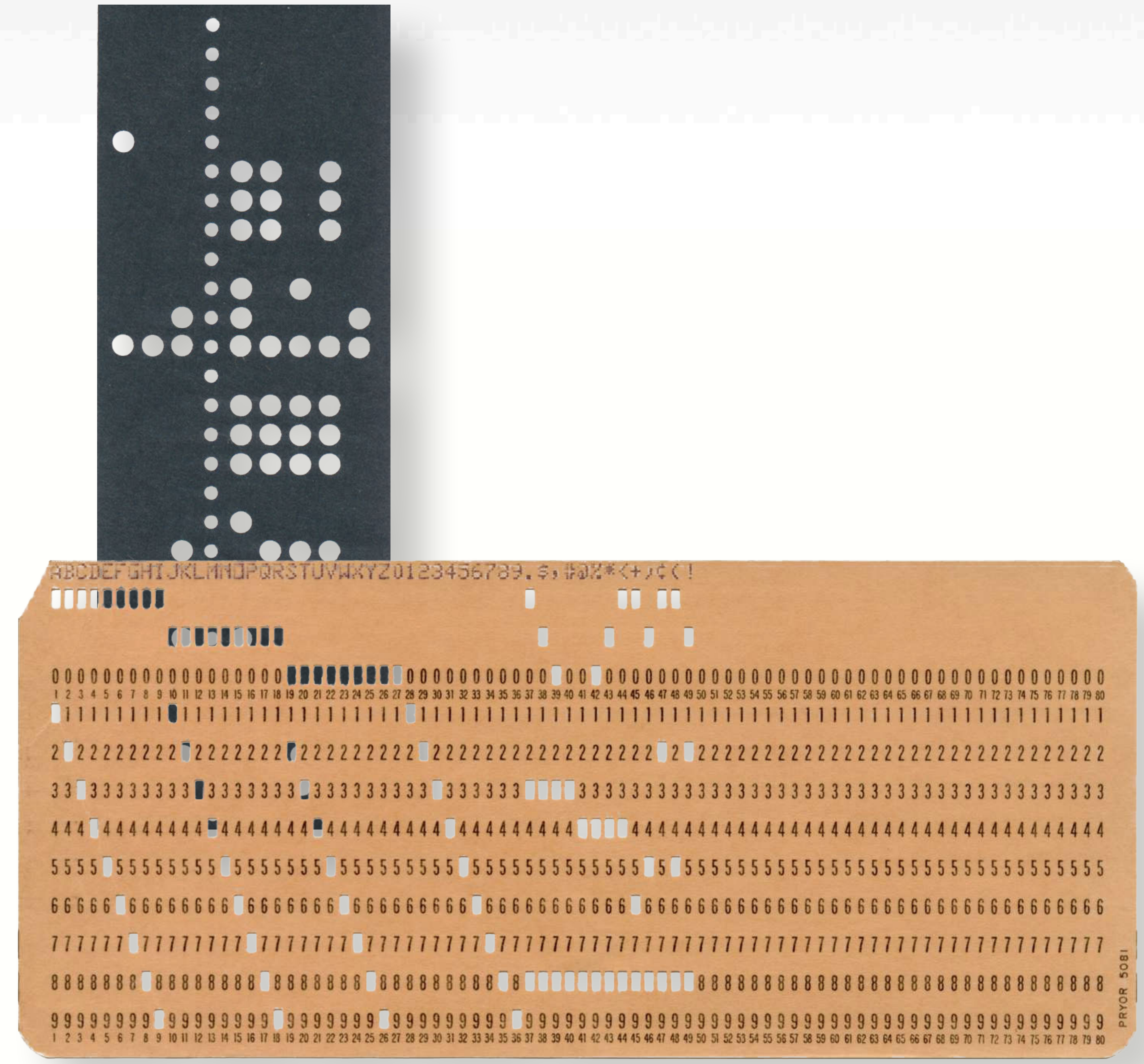
 

Interestingly enough, the dawn of practical electronic computers saw the use of key punch cards and paper tape as a method of storing programs and information. Because they were on paper, they can still be read 70 years and more after their creation, and because they are physical, also offer their own unique ability to assure provenance of any file.

Completely impractical from a data access and data density standpoint, but they provided a glimpse into the characteristics that were needed for proof of provenance and long term preservation of data.

However, the immense demand for data density and speed of access precipitated solutions that, understandably, had no long-term preservation characteristics. Combined with a world where hacking, malware, and AI weren't even imagined, this is understandable. Because concerns of AI compromising data provenance weren't even on the radar, an ecosystem evolved where preservation wasn't deemed necessary, since it was so easy to make copies of digital files (migrate), and the amount of data being created through the 1980s requiring preservation was so small.

Since then, the explosive growth of unstructured data from geo-spatial, audio, healthcare, media and entertainment, surveillance data, photos and more, created an environment where concerns for protection from AI attacks on data provenance became unaddressable. Constant migration of Exabyte data sets thwarts any opportunity to protect data provenance from AI. In addition, there is the risk of data loss due to threats from magnetic fields, hostile electromagnetic environment such as EMP, obsolescence, and data corruption from migration itself.





# Current proposals for data provenance protection from AI rely on software and migration that are *Not* foolproof and not good for the planet

- Whether it's Cloud, enterprise data center, hard drive, data tape, or solid state

- All current proposals rely on storage media that is rewritable, leaving a back door for AI bad actors to compromise originals

- Since migration is also involved, expired media all end up in landfills and are ***NOT*** recycled

- Current practices create massive waste from forced migration, and sending countless hard drives and data tapes into landfills

# A Sustainable Solution for AI Data Provenance

## DOTS is Green Technology & Sustainable

**DOTS** eliminates the constant flow of discarded storage media into landfills. It eliminates media & energy waste from forced migration, costly power requirements, and rigid environmental control demands

As mentioned earlier, it is critical that the method employed to protect data provenance must be unencumbered by complicated technology. Yet it also cannot be so simple as legacy paper tape or keypunch cards as to be impractical. The solution must be designed to ensure provenance, preservation and comprehension demands can be met, yet at the same time deliver write and read speeds that are compatible with today's data access demands.

The technology that successfully meets these metrics of preservation and data access is **DOTS**.

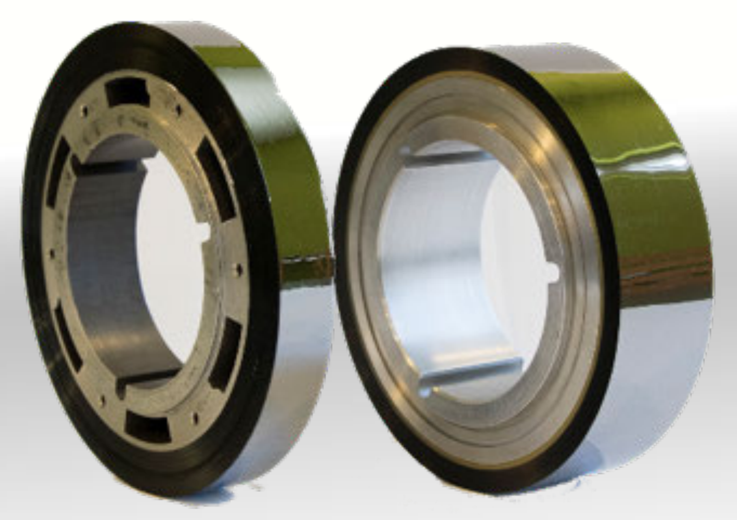
The foundational tech behind **DOTS** was invented by Eastman Kodak over 20 years ago, and those core characteristics have stood the test of time and are still compelling today.

Group 47 has further improved upon the design that already ensure the needs of data provenance and the archivist are met:

- Immutable write once medium, eliminating any chance of AI compromising provenance of originals
- Immune to magnetism, hostile electromagnetic environment such as EMP, and able to withstand environmental stress
- Long lived medium meeting the requirement of data provenance requiring a file be unchanged
- WORM and Archival for 200 to 2000 years*
- Can be open reel or the same form factor as existing LTO data tape systems
- Compatible with current robotics, commands and LTFS
- Hardware devices are backwardly compatible for all previous generations
- Total cost of ownership (TCO) superior to endless migration

*Results of Carnegie Mellon Batelle Class II accelerated life testing

# What is it?



- **DOTS** is a phase change media composed of a patented metallic alloy sputtered on an archival polyester base (e.g.: Aramid, Mylar™, Estar™)
- Standard manufacturing techniques are used to create **DOTS** tape
- **DOTS** media and prototype recorder/readers were successfully built by early 2001
- Group 47 has dramatically improved and simplified the hardware design with an engineering upgrade/ refresh and strengthened the patent portfolio
- Components for **DOTS** recorder/readers employ off-the-shelf imaging and laser technologies
- Carnegie Mellon University tested **DOTS** and concluded an archival life of 200 to 2000 years.
- Group 47's new design was proven in a contract with the CIA



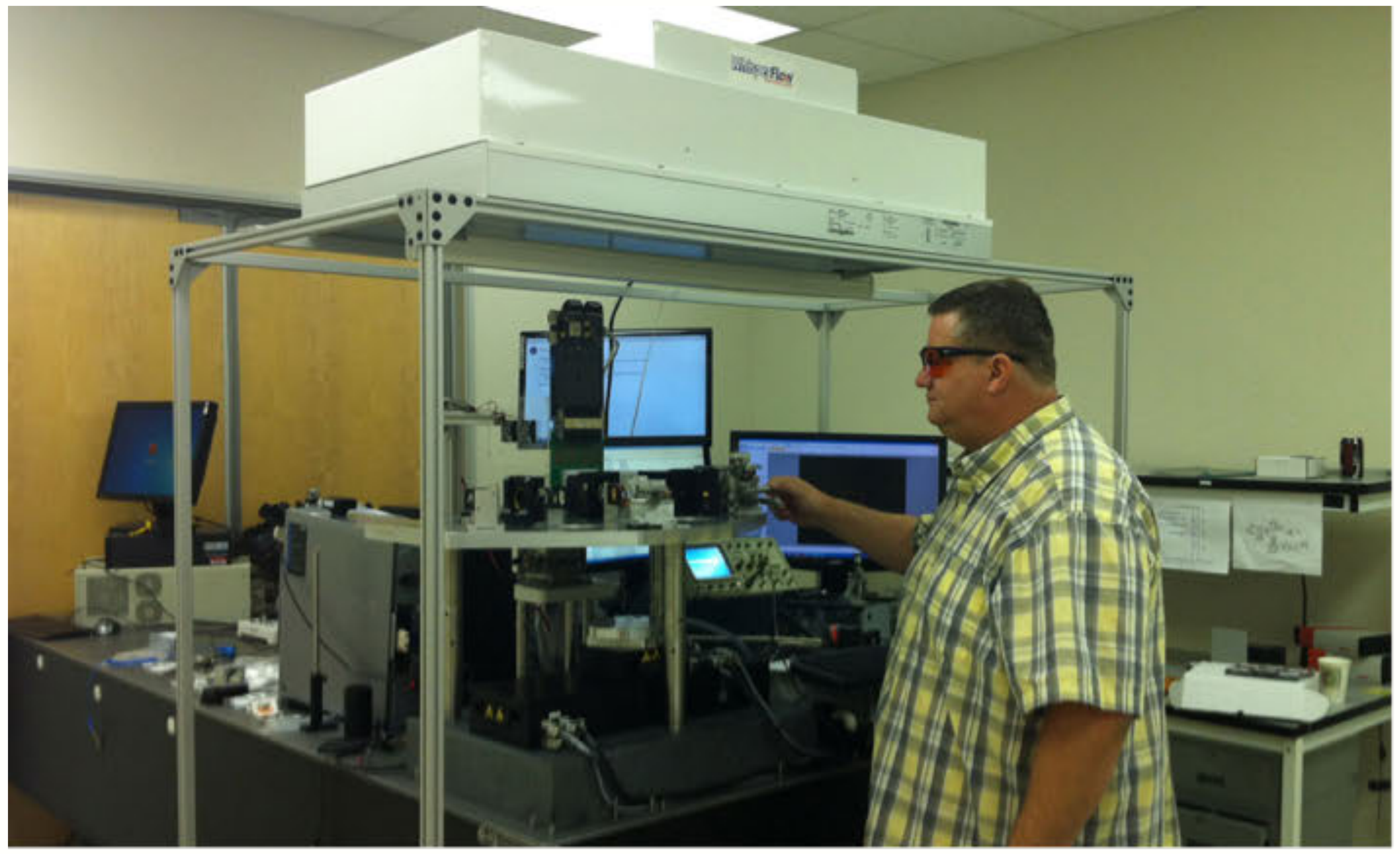
 

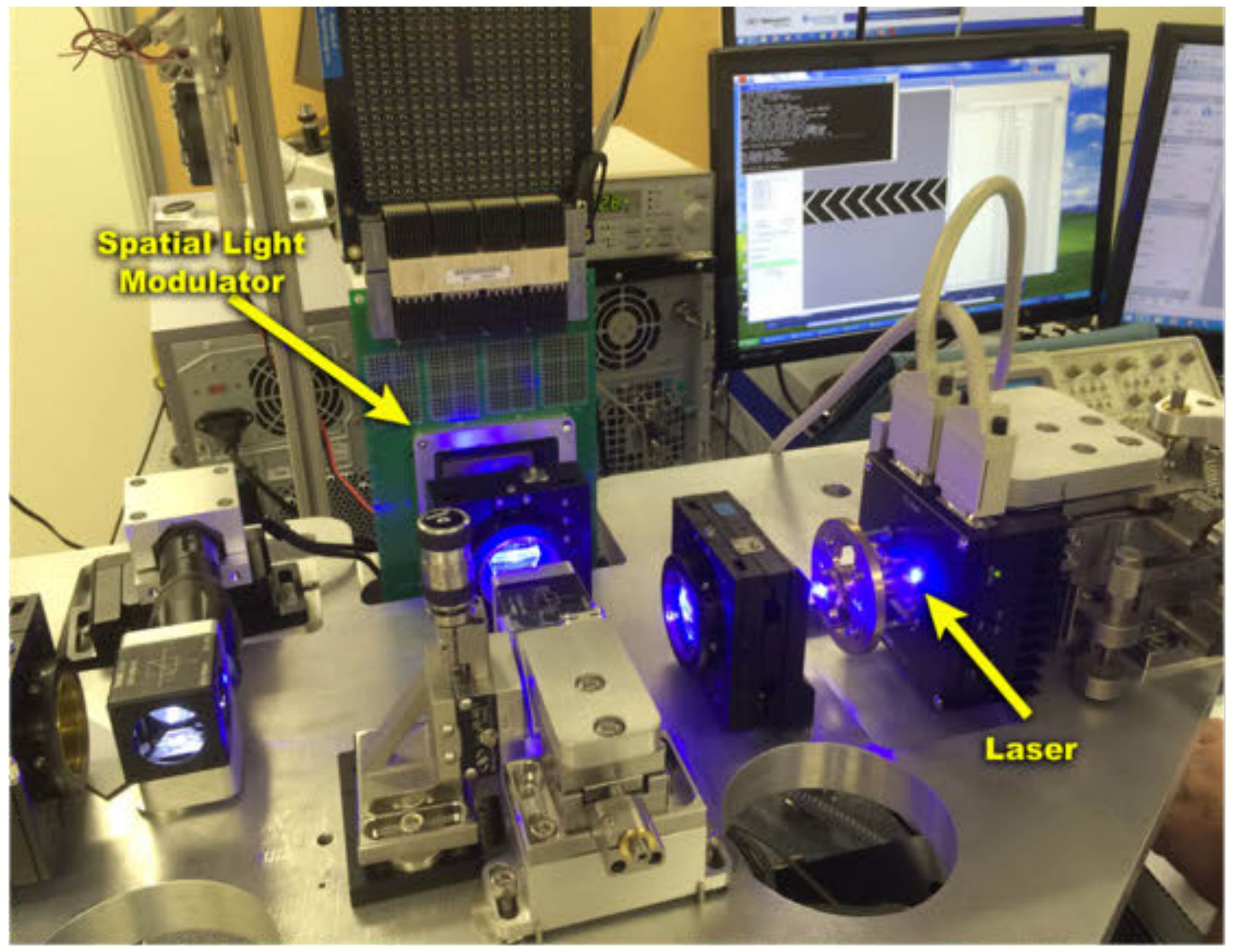
# The Product

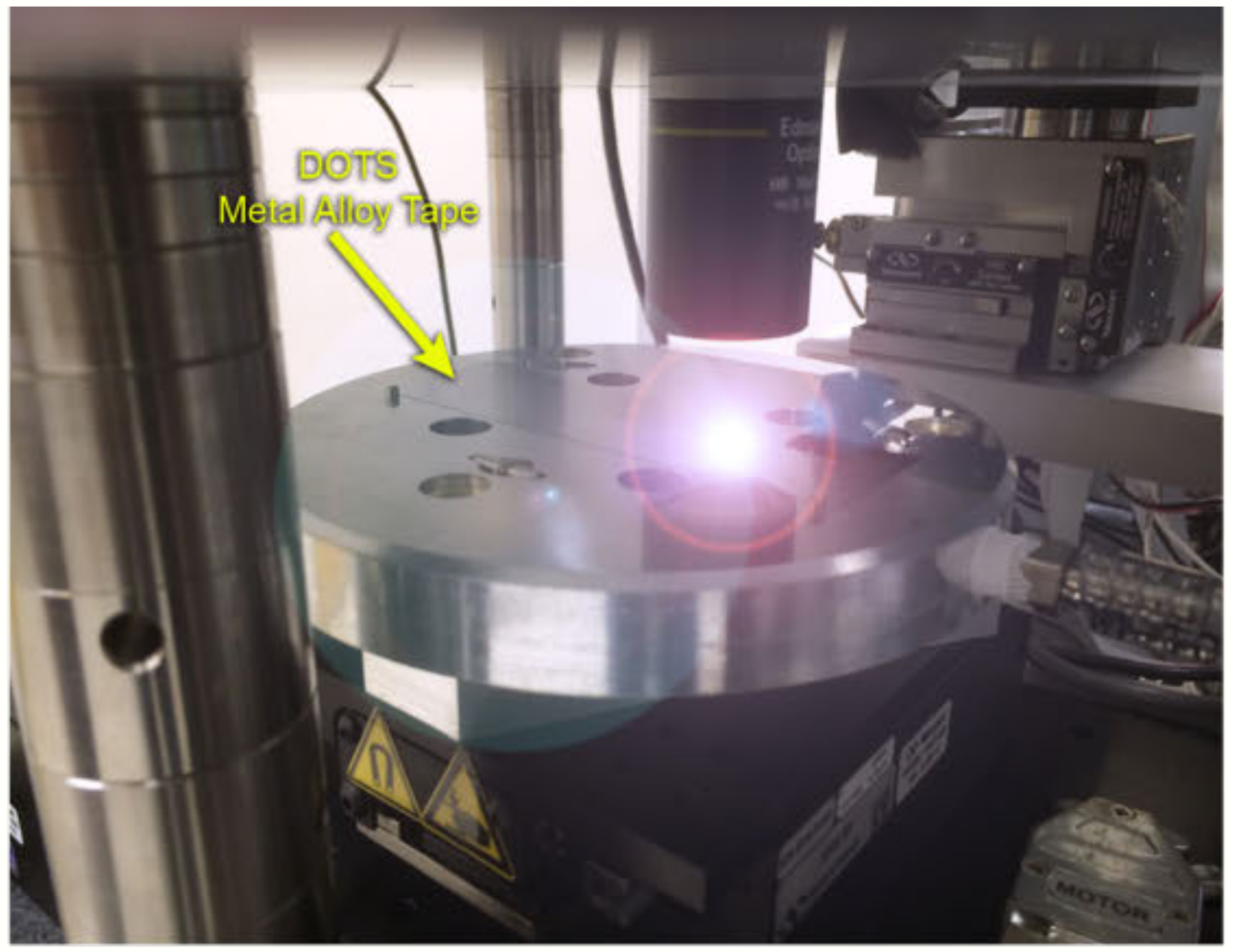
- **DOTS** can store any file or application you currently save to a hard drive, data tape, flash memory, or an SSD
- Because the data is represented visually, as long as we have access to cameras, the data will always be recoverable, and natively provides proof of provenance of any file saved to **DOTS**
- A visual representation means all hardware will be backwardly compatible to the 1st Generation
- New, efficient design eliminates “serpentine” recording techniques, ensuring record & read speeds in excess of 1 GByte/sec upon commercial availability
- Each unit of **DOTS** media will have factory-written, human-readable instructions for building a reader on the first few meters – Rosetta Leader
- Group 47 will “open source” the **DOTS** read only units, licensing them to be manufactured royalty free by qualified manufacturers
- Bit Plane Image preservation format ensures proof of data provenance for image and sound file and their recovery for centuries

# CIA Proof of Concept

- CIA awarded contract to build a laboratory prototype proving the **DOTS** technology
- Purpose was to assist in funding process by eliminating need for building another prototype – CIA certifies that the **DOTS** technology works
- Contract milestones included demonstration of writing and reading applications and document data in the **DOTS** visual format, and successfully writing and reading to **DOTS** metal alloy tape in the [Bit Plane Image preservation]() format







# CIA Proof of Concept

First successful write to DOTS media using G47's new approach

Psalm 121:1-2
A Song of Ascents

I lift up my eyes to the hills.
From where does my help come?

My help comes from the Lord,
who made heaven and earth.

The data image as sent to the G47 Engineering Model

The data image as written to the DOTS media, and recaptured with a camera

Later test writes succeeded with substantially greater clarity

# **DOTS** Detailed Technology Brief

## The **DOTS** Metal Alloy

**DOTS** media is a digital optical tape media based on a very stable phase-change recording layer created from a proprietary alloy of three metals that is DC magnetron sputtered onto a PET (Mylar™, Estar™), with a wavelength matched overcoat.

In a study conducted by Mitsubishi Polyester Film, Inc., it was found that when a metalized PET film is created using PVD, a physical interlocking of the metal atoms with the PET occurs, yielding very good adhesion. Further strengthening of the adhesion occurs due to a covalent bond being formed between the metal atoms and the PET polymeric chains through chelation.

The PVD coating technique, having no chemical binders, yields strong adhesion and ensures no mechanical failures such as delamination, chemical creep, fading, etc., suffered by magnetic media which is produced using a continuous wet coating process, where a magnetic dispersion is applied to the base film. This magnetic dispersion consists of binders, magnetic pigments, dispersants and lubricants which are dissolved in organic solvents to form a slurry. These binders (along with residual solvent, dispersants, etc.) will over time undergo chemical processes which may weaken their ability to hold the magnetic pigments to the base media or may cause breakdown of the pigments themselves due to chemical reactions.

The **DOTS** alloy, when sputtered, forms a phase-change media which undergoes a change from amorphous to crystalline state when heated to a temperature above 160° C. The phase-change crystallization rate is < 1ns and has been observed to be a threshold process consisting of extremely sharp transitions. In addition, the media supports a wide wavelength write response having been demonstrated at wavelengths from 830nm to 405nm.

The crystalline (written) state possesses a reflectance of 8.3% at 780nm vs a reflectance of 32.7% at 780nm for the amorphous (unwritten) state. There is a contrast of 59% between the crystalline (written) state and the amorphous (unwritten) state.

The DOTS media is stable for 100 years at 100°C or 212°F and over 200 years at room temperature (25°C or 77°F). The DOTS media is immune to chemical, water, and other environmental damage and is impervious to corrosion.

The DOTS alloy may also be sputtered onto other materials and could be used in the production of ID badges, key cards, etc.

# Data Encoding Using DOTS

Data is recorded on to the media by a laser that effects a phase change in the alloy. This phase change significantly changes the index of refraction of the alloy creating a substantial decrease in reflectivity. Laser writing of the data coupled with optical imaging for data retrieval ensures no physical contact with the media and therefore, no degradation due to friction.

## Data Encoding on Current Magnetic and Optical Media

All current optical media and in point of fact, all magnetic media which are used to store digital data, do so by storing a physical means of modulating a base signal source. This time varying modulated signal is a reproduction of the original electrical signal, which represents the binary bits of the data being stored.

Optical media has a series of pits and lands that are used to modulate a laser beam. This modulation is accomplished using a scheme wherein the base signal has a value of constant “zero” over time (NRZI). When the laser light reflecting off of the optical media to a detector changes in intensity, the change is registered as a “one” at that clock tick. At the very next clock tick, the signal is again read as a “zero”. Thus, the only time the signal is read as a “one” is when there is a large enough change in the signal level (i.e. it exceeds a specified threshold), regardless of whether the change is from high to low or low to high. In the case of CD, DVD, BD, etc., this signal change is created as the laser beam traverses the edges of the pits. That is, when the laser transitions from a land into a pit or vice versa, the amount of light reflected back to the detector changes drastically; from a large to a small amount at the leading pit edge (transitioning from a land into a pit) and from a small amount to a large amount at the trailing pit edge (transitioning from a pit onto a land).



Now, because only the signal change event is read as a “one”, it is clear that there is no way to have two ones in a row, i.e. “11”. It is also easy to see that because we are dealing with a time-varying signal, the accuracy of the data readout clock is extremely important. In addition because of the use of edges to create signal change, the tracking of the laser must be very, very accurate so as to avoid wandering into adjacent tracks or only partially detecting transitions so that they do not register high enough of a signal change.

All these considerations taken together mean that in order to recreate the original data with high accuracy the original data needed to be re-encoded in such a way that the new encoding provided a robust method of unambiguously recreating this re-encoded data signal with the physical, transition based “pits and lands” storage mechanism. This encoding in the optical world is EFM+ (Eight-to-Fourteen Modulation Plus). Simply put, each of the 256 possible 8 bit values of a byte is converted into a 16-bit code word utilizing a 256x3 entry table. Each code word for each byte values has three possible representations depending on the 3 possible values of the state machine. Which of the three states is selected is established by a set of algorithmic rules.

By using this tri-state, 256-entry code word table, the original 8 bit data signal is converted into a 16 bit code word which is then used to modulate a writing signal (laser) which creates a physical representation of this code word in the media. When read back, the physical media modulates the reading signal in such a way that it represents the original 16-bit code word. This code word is then translated back into the original 8-bit data signal.

Figure 1 shows how the two ASCII bytes of the string “47” are represented on DVD. The ASCII bytes representing “47” are 52 and 55. The 8 bit binary representations of these bytes are 00110100 and 00110111. The EFM+ code words for these bytes are 0010000100100010 and 0010000001000010.

Data is stored on magnetic media in basically the same fashion except that the base signal is an electrical signal created by a magnetic field and the physical signal modulation is accomplished by magnetized particles on the media, which perturb the base magnetic field.

Storage companies have always viewed this “signal modulation mechanism” storage scheme as the best way to store data.



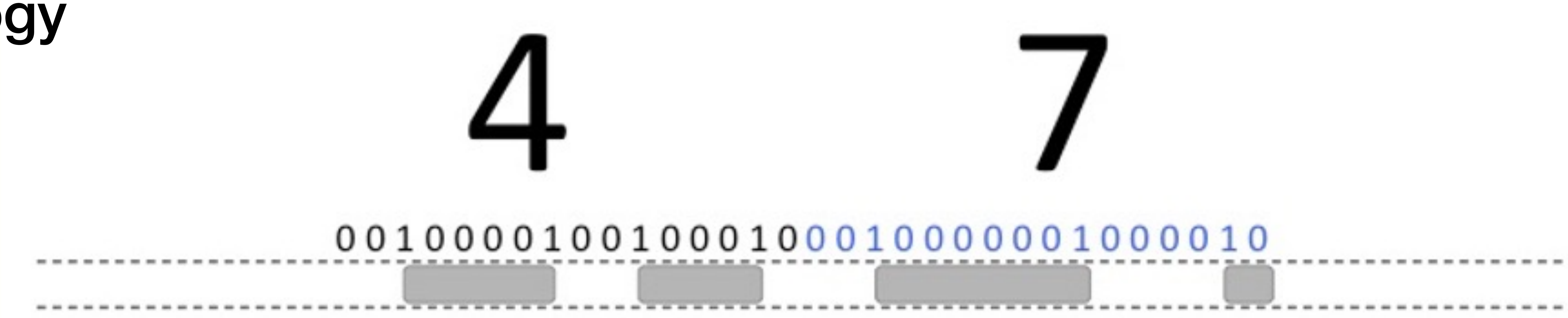



Fig. 1

## Data Encoding with DOTS

Group 47, has decided, that given the archival and optical nature of the DOTS media, it is time to view the storage of the data as an image processing function. That is, the actual data itself should be physically drawn on the media and a camera along with image processing should be used to extract the data from its own image.

There are many ways in which data can be stored as a physical image. Figure 2 illustrates how binary data can be represented as a series of low reflective dots representing ones and highly reflective spaces representing zeros.

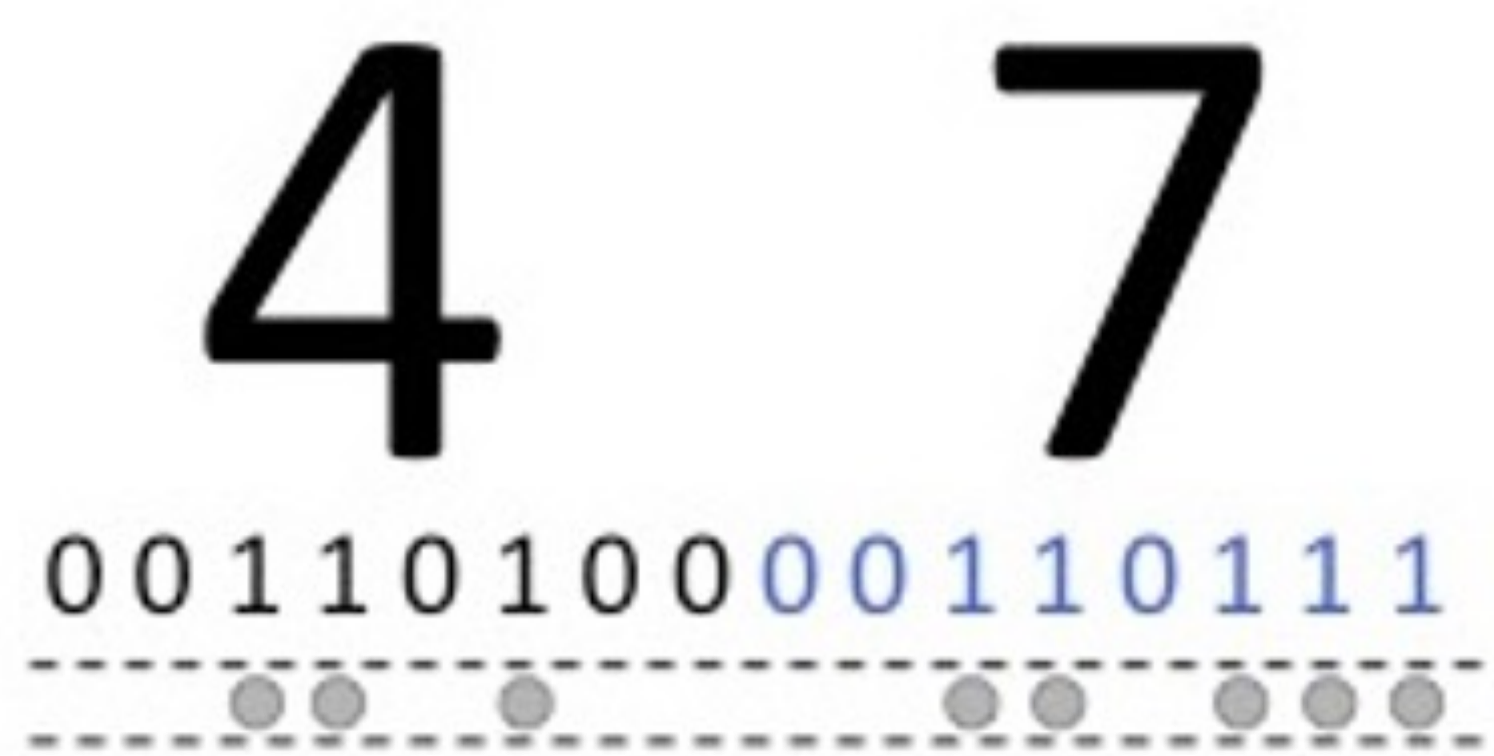



Fig. 2

As can be seen, if the size of each dot is known, then it is a simple image processing problem to determine exactly how many spaces exist in between each dot.



The basic physical design for the use of **DOTS** to store data optically is shown in Figure 3 (*not to scale*).

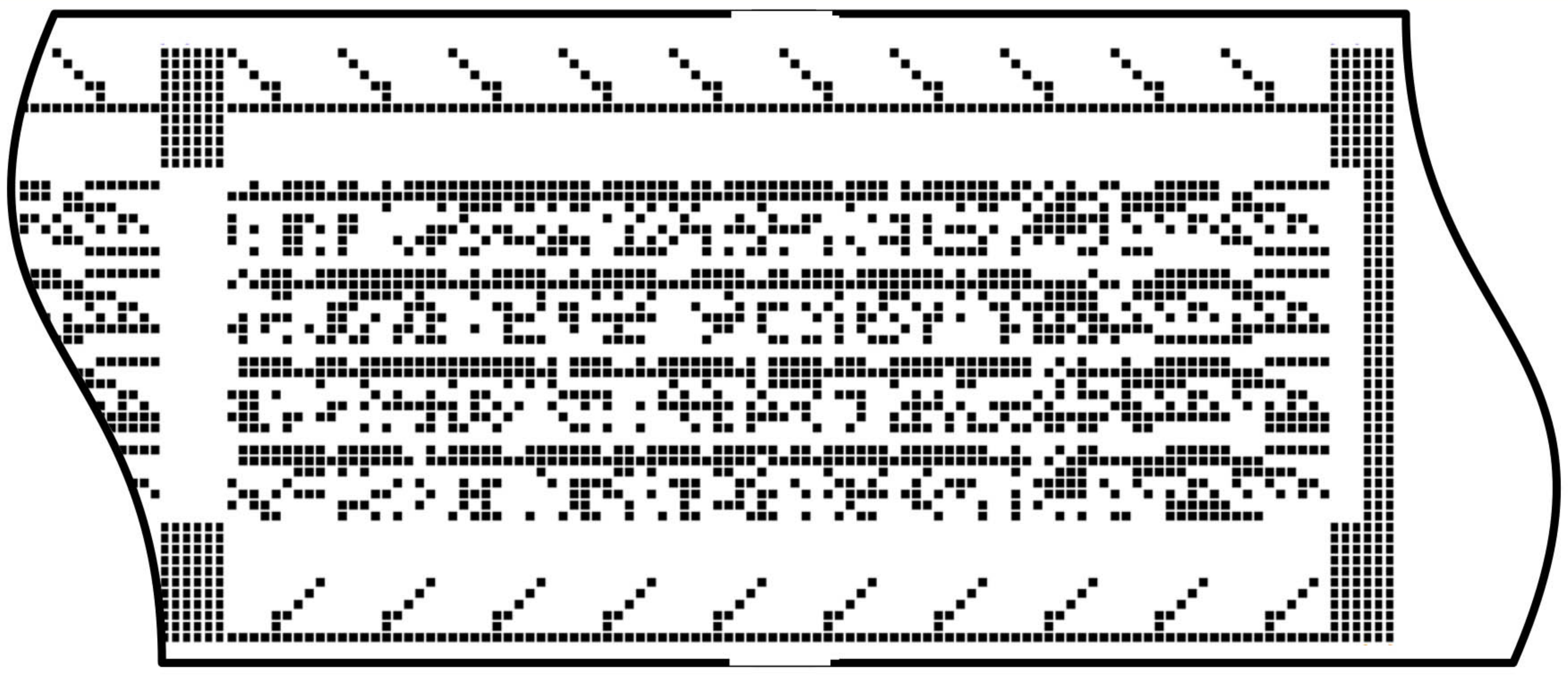


Fig. 3

Inset .5 to 1mm from the physical edge of the tape, as shown in Figure 3, are two optical guide tracks. These tracks are used as fiducial marks for the image processing of the data, which is written between two “dead bands”. These dead bands provide a separation zone between the actual data and the guide tracks.

The guide tracks themselves are written on the media simultaneously with the data. Each column of written data has a guide spot written above and below the column, exactly in line with the data. These are known as guide spots. To provide for a more robust guide track detection, a second set of guide spots is written to the outside of the “top” guide track and the inside of the “bottom” guide track. These are known as stagger spots. The stagger spots are oriented so that they provide a visual reference of tape travel direction as well as “top” and “bottom” of data columns.

Three columns of pixels written the entire width of the tape, represents the beginning of tape (BOT), while two sets of three columns opposing each other and written across the guide track and dead band areas (e.g. excluding the active data area), represents a file mark; both beginning of file (BOF) and end of file (EOF).



Figure 4 shows a notional diagram of how the tape would be written by a laser array.

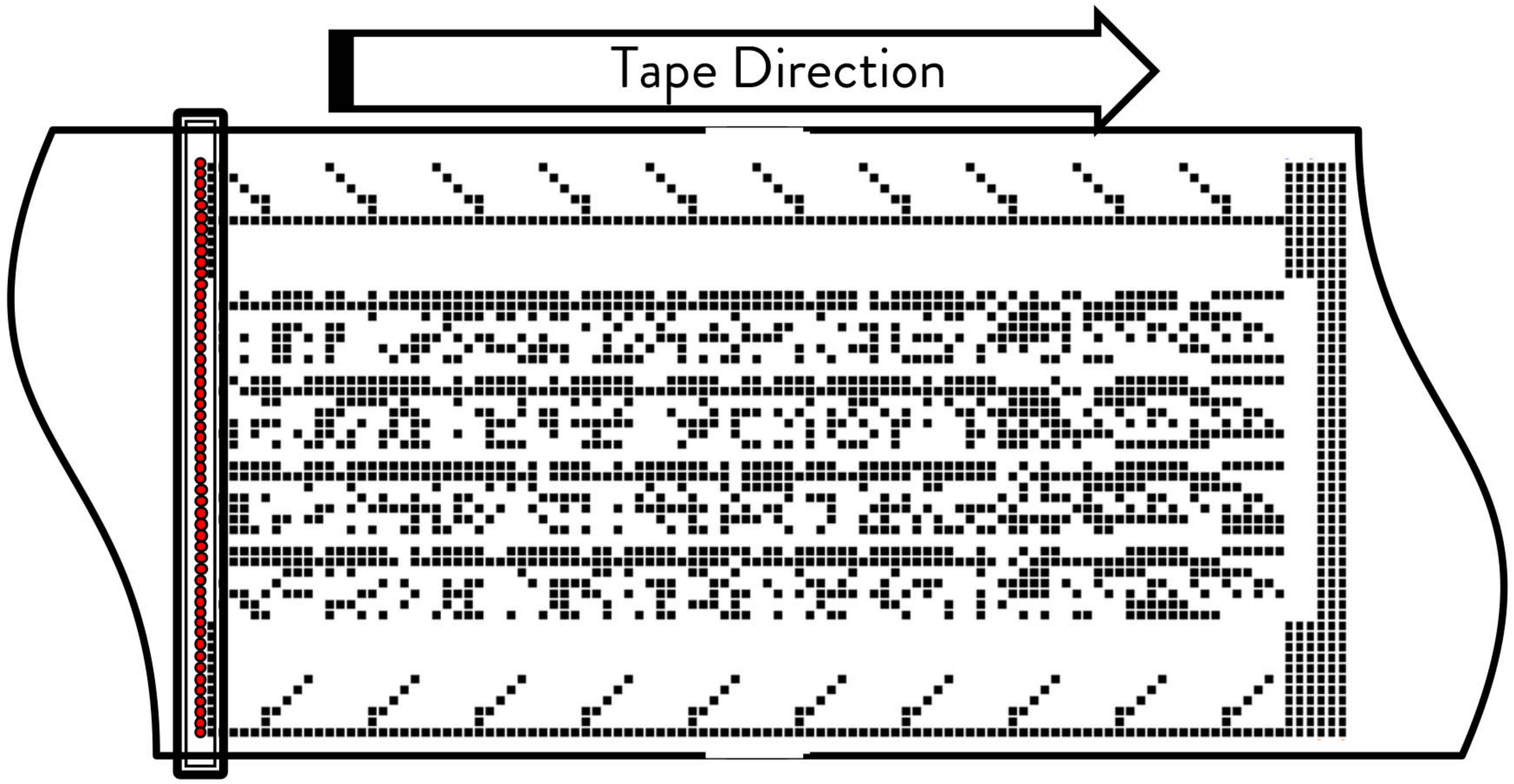



Fig. 4

Since the guide track is written precisely in line with the data, a pair of directly opposite spots on opposite guide tracks will provide a precise alignment and spacing measurement when the media surface is imaged. By providing guide spots, the precise word boundaries may be established optically even in the event of catastrophic damage to a portion of the guide tracks.

The **DOTS** media would be read by a solid-state imager comprised of an oversampled linear array of monochrome photosites plus additional photosites to image the guide tracks. Figure 5 illustrates this arrangement.



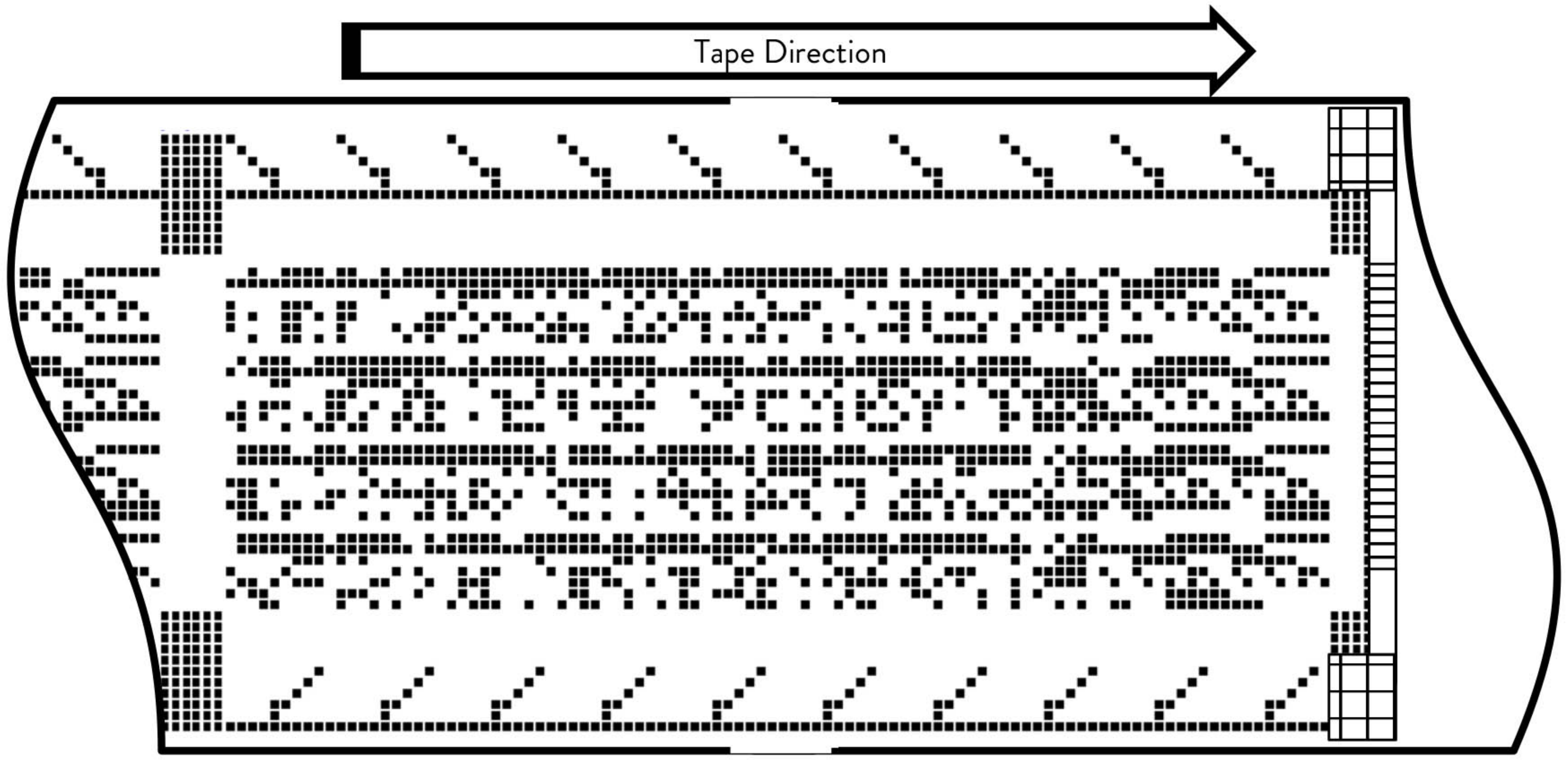



Fig. 5

As the media passes beneath the sensor array, an image of the media surface is built up line by line. A full image may be created upon detection of the guide spots.

This use of guide tracks which are created concomitantly with the data so as to provide precise fiducial reference for the imaged data, as well as the physical representation of the actual data itself, provides for a data storage scheme which is robust enough to accommodate variability in media velocity, tracking error and even physical deformation of the media itself. Figure 6 illustrates this concept.

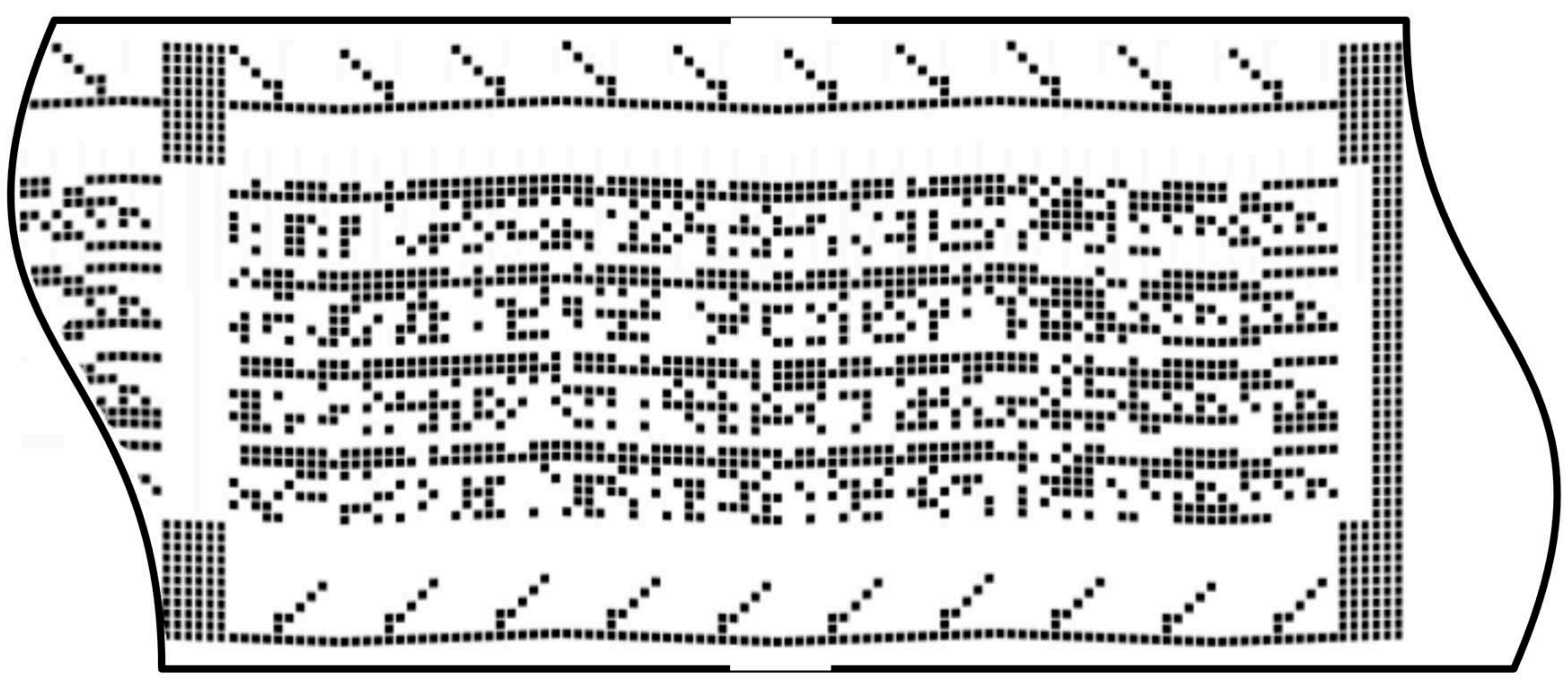


Fig. 6



# DOTS Technology

As can be seen, even though badly warped, it is readily apparent that the guide spots as well as the stagger spots can be easily used to determine how to align the data image when an imager captures the data and simple image processing is brought to bear.

So far, only the binary representation of data has been shown wherein binary bits are represented as spots of low reflectivity (“1s”) and areas where the metallic alloy reflectivity is unchanged (“0s”). Because of the optical nature of the DOTS media as well as the decision to utilize an optical representation of the data itself for storage of that data, the DOTS media lends itself to many types of data representation. The following figures illustrate three of the data storage concepts, which are possible. It is important to note that all four methods of data representation are supported by the DOTS media and reader/recorders because of the single, uniform method by which any data may be written and read.

It should be noted that further research is being done as to how the placement of guide spots and stagger spots may be done so as to provide meaningful timing or event data to the image processing system.

For purposes of this discussion, the guide tracks along with the guide spots and stagger spots are shown identically for each type of data representation.

Figure 7 represents the archiving of digital imagery by representing each bitplane of each of the 3 color channels (RGB) in an image as a binary image array. Figure 7 shows a frame from a 10bit DPX encoded RGB frame.

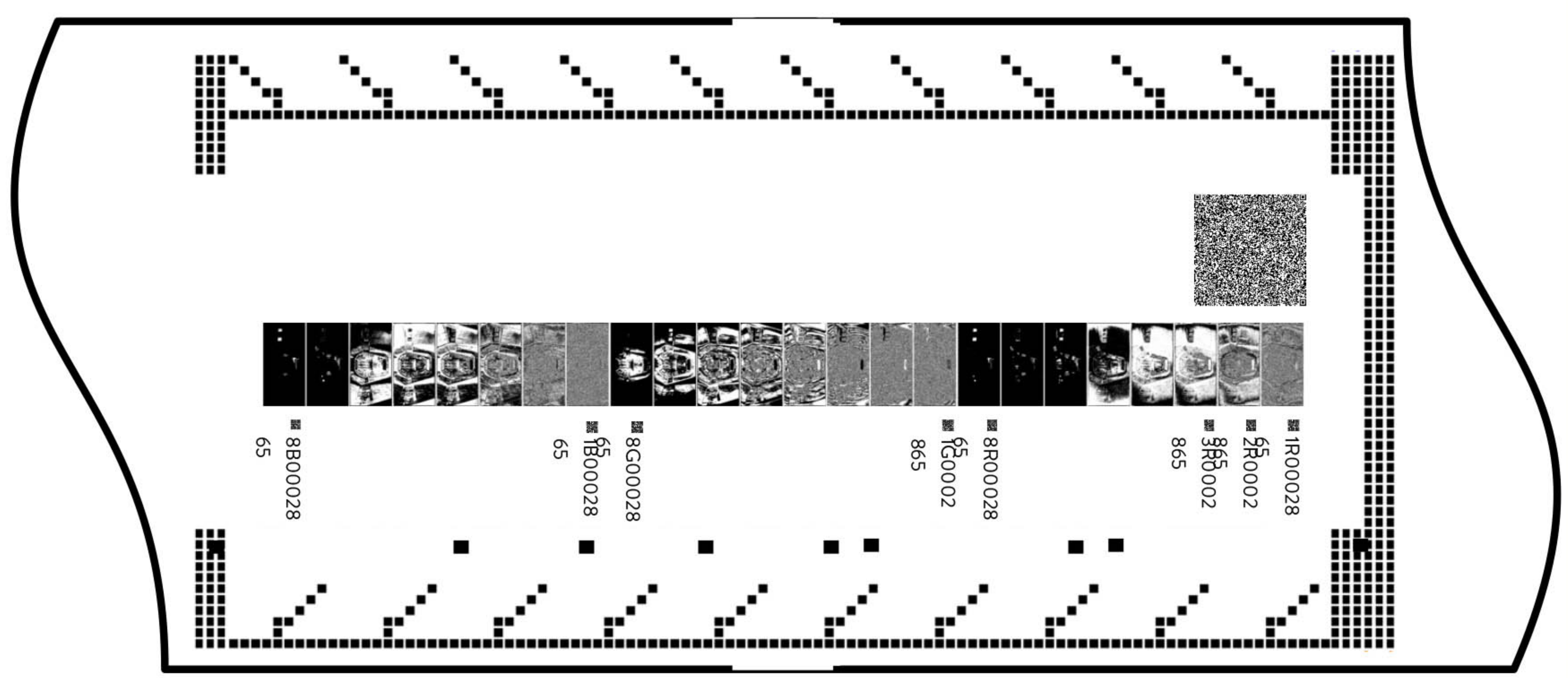


Fig. 7

# DOTS Technology

Next to the image are textual and QRcode elements which contain metadata describing the data. For example, the first binary image on the right is labeled “1R0002865”. This indicates that this image represents bitplane 1 of the Red channel of image frame number 2865. The next image to the left is labeled “2R0002865”. This indicates that this image represents bitplane 2 of the Red channel of image frame number 2865, etc. In image “1R0002865”, each “1” in the binary image indicates that the Red pixel at this location in the image frame, has the zero order (ones place) bit set, while in image “2R0002865”, each “1” in the binary image indicates that the Red pixel at this location in the image frame, has the first order (twos place) bit set. Thus by combining all ten binary images together by applying a logical OR operation to all of them, the full 10bit depth of the Red channel will be restored. By performing this operation on all the bitplane images of the Green and Blue channels, all three color channels will be fully restored and can be combined into a full color image.

This textual/2D barcode data representation is illustrated in Figure 8.

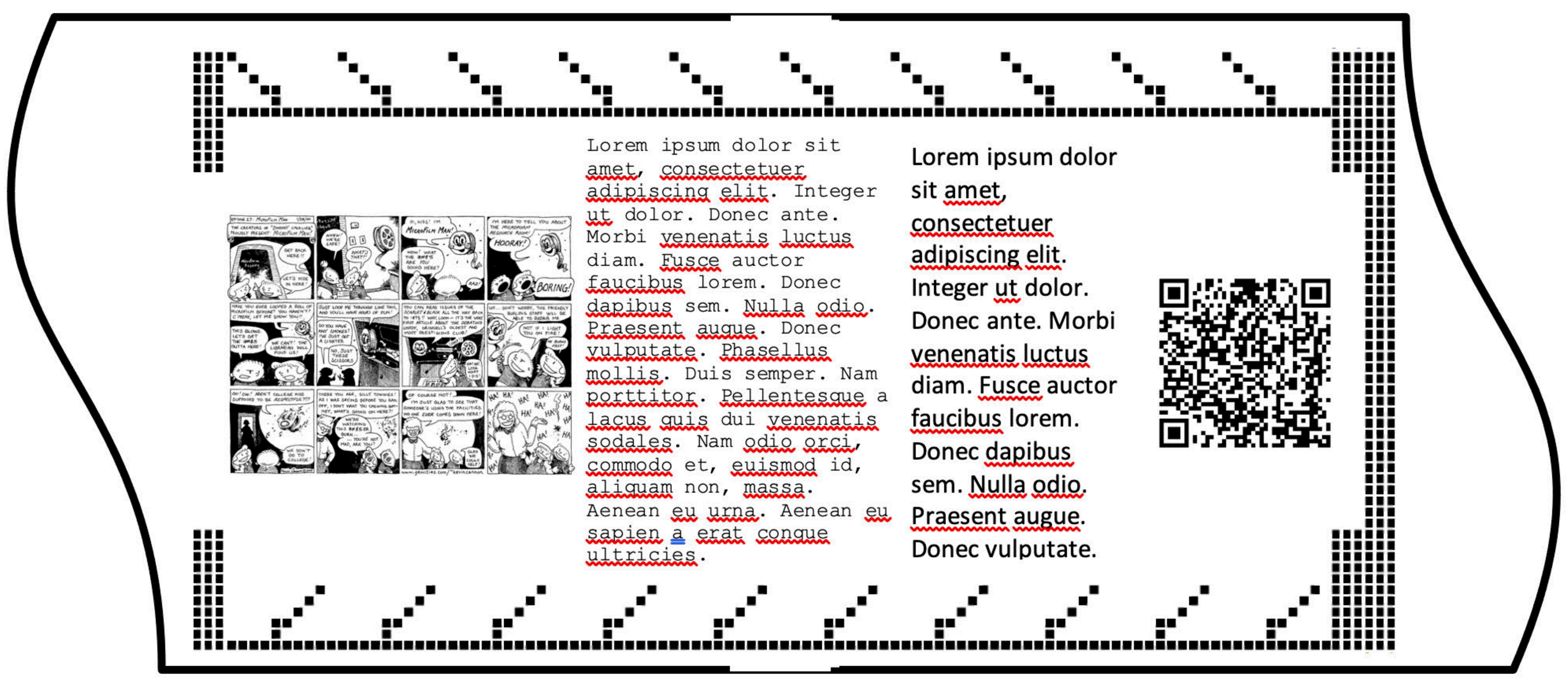



Fig. 8

This figure shows that the **DOTS** media may be used as a direct replacement for most applications which current utilize microfilm. The restriction that only monochrome imagery can be written requires that, as shown in Figure 8, any multi-bit imagery would need to be written as a series of binary bitplanes.



Figure 9 is a notional diagram of the way in which the **DOTS** media may be read.

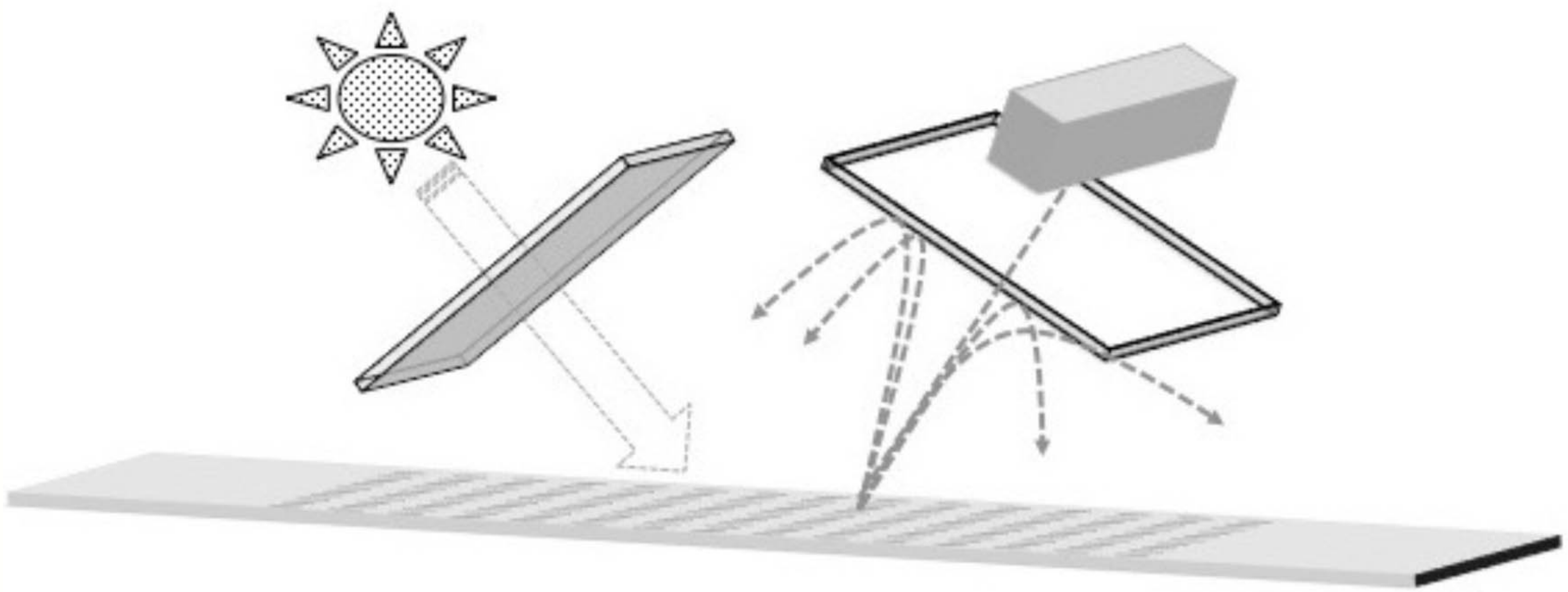


Fig. 9

A light source is directed through a polarizer onto the metallic surface of the **DOTS** media and is then reflected onto the imaging array. Between the media and the imaging array is a polarizer. The phase changed data spots exhibit very little specular reflection and a good deal of diffuse reflection while the non-phase changed alloy exhibits extremely high specular reflection and almost no diffuse reflection. Because of this difference, the highly specular reflections from the metallic alloy are filtered out by the polarizer showing up as black while the diffuse reflections from the data spots pass through to the detector showing up as white.

Because of the thickness of the phase change alloy and the mechanism of phase change, the phase-changed spots are readable from both the front and back sides of the support material. This would allow the use of multiple imagers to read/write the data.

# DOTS Technology

Here is a reduced resolution image of an actual scan of data saved on DOTS media by Group 47. It was produced by fastening a sample of the media to a DVD and then using a LightScribe© DVD burner to create the image. The image was scanned using a Canon CanoScan 8800F flatbed scanner at 2400 dpi. While crude, it illustrates the very high SNR which can be achieved by the media even in the face of incomplete phase-change, physical mishandling of the media and a very simplified reading setup. No enhancement other than cropping, rotation and resizing has been done on the image.

## Write Head

The **DOTS** Write Head shall consist of a laser source of sufficient power, along with associated optics, to effect a phase change in up to 10240 spots simultaneously, while guaranteeing the size, spacing and contrast requirements cited previously. The laser wavelength is selected depending on the focusing optics NA. Since the spot size diameter, D, is related to the laser wavelength, λ, and optics NA by

$$D=\lambda/NA$$

the most practical λ may be chosen based on the NA required for a 1μm spot. Should the NA requirements be impractical, the spot diameter may increase from 1 μm to up to 1.3 μm

At this time, the multiple simultaneous spots will be created through the use of a diffractive spatial light modulator (SLM), such as the Edge-E device from Alces Technology.

Operationally, the laser source beam(s) will be homogenized into a linear top-hat beam whose width matches the active width of the diffractive ribbons in the SLM (≈3.7 μm) and whose length is that of the SLM (≈25 mm) in the case of a 4096 pixel SLM. The SLM generates a spot when a ribbon is electromechanically depressed by a quarter wavelength of illumination (λ/4). This displacement creates a phase-delay in the reflected light proportional to half the wavelength (λ/2), whereby the reflected electric field amplitude can be described as proportional to ejϕ and in this case simplified to +1 and −1. In the next row, the impulse function of the optical system can be described as a pair of delta functions, spatially offset, and 180 degrees out of phase. When the reflected light off the MEMS ribbons is passed through the polarization displacement device and Analyzing Polarizer (this can be mathematically represented as the convolution of the electric field of the reflected light and the impulse function of the optical core), the output light is spatially defined as a pulse with width “p”. This rectangular pulse is an ideal pixel in the display and centered on the edge created in the MEMS array; hence the name “Edge-E”. In the non-depressed state, the Edge-E device ribbons act like mirrors, reflecting the incident illumination along the incident angle. Further information may be found in US Patent 7,940,448.

The Edge-E device as designed creates gaps between adjacent pixels. The **DOTS** data encoding scheme necessitates that the spot pitch is 120% of the spot diameter. In addition, the Edge-E device creates circular spots although the geometric shape of the data spots is irrelevant to its consequent detection.



## DOTS Technology

Figure 10 is a schematic block diagram of the DOTS systems Recorder/Reader for a cartridge container, but is effectively the same for an open reel design.

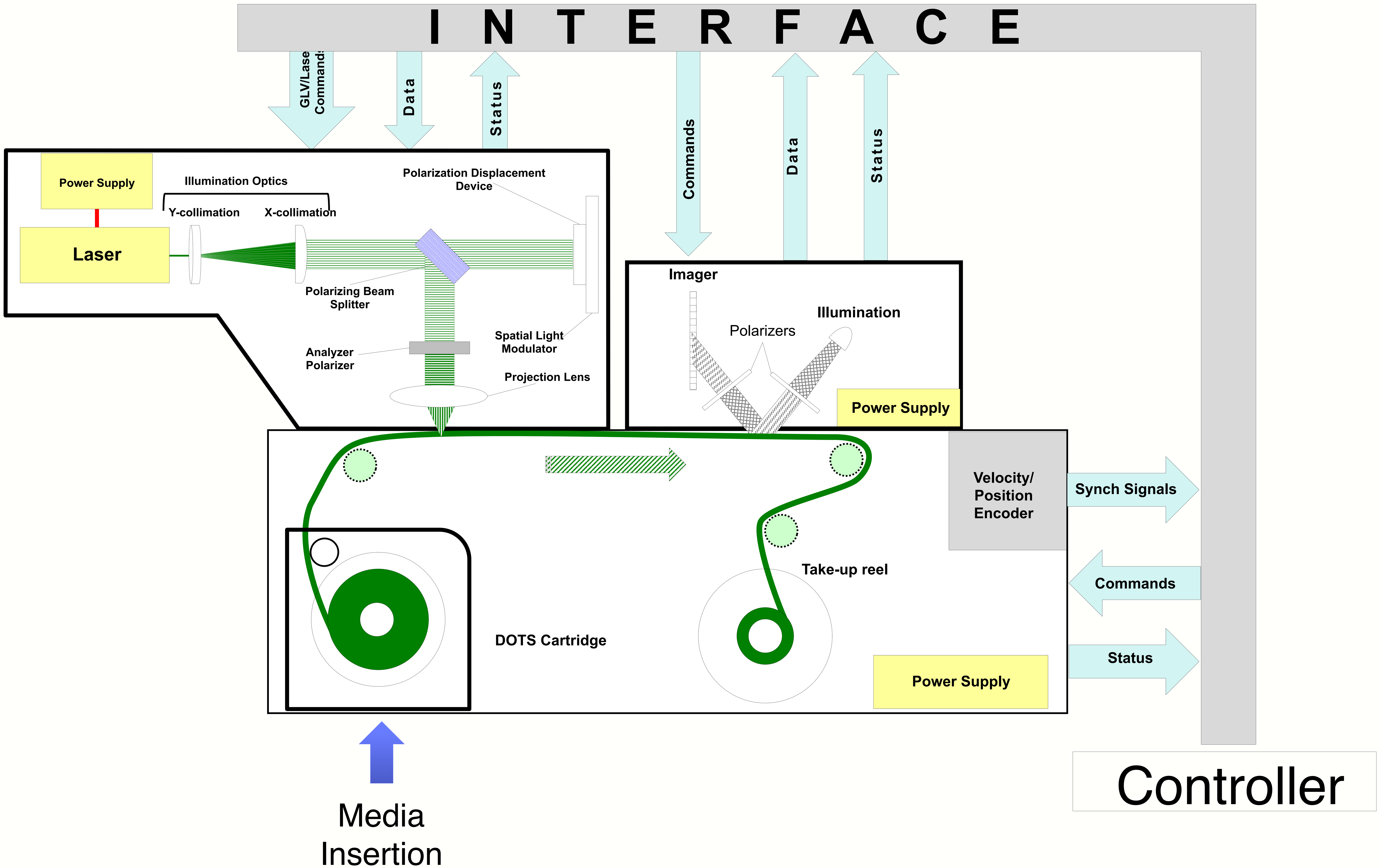



Fig. 10

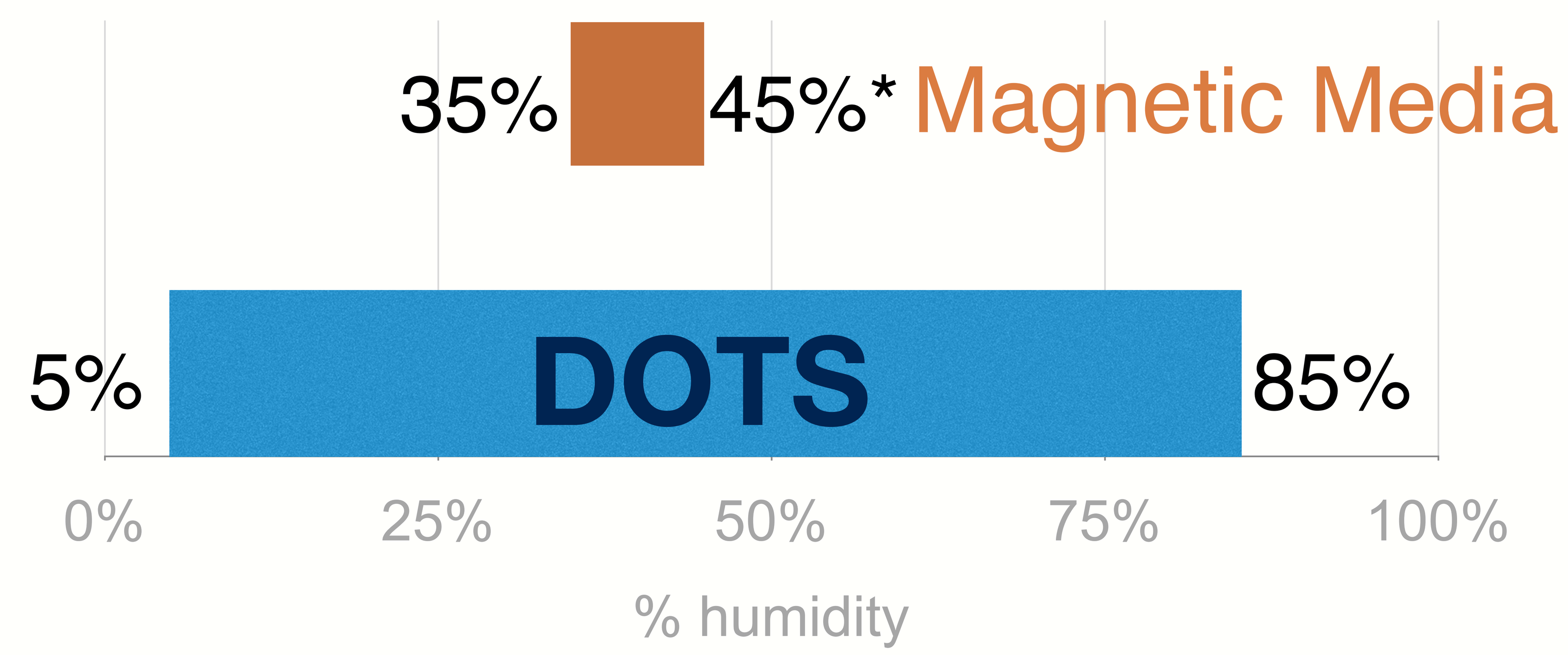
# G47 Visual Metal Alloy Tape - An Overview

- Media
  - Phase-change alloy sputtered onto archival (e.g. PET, PEN, Aramid) tape
- Recording mechanism
  - Laser heating to transition temperature (>150° C)
  - Multiple laser beams are used which eliminate the need for serpentine recording (1 µm spot)
  - Guide tracks are recorded at the same time as data is recorded (no preformatting)
- Read mechanism
  - Linear array image sensor
  - Image processing using guide tracks
- Data Encoding
  - **DOTS** is represented as optical bit-patterned media (BPM)
  - Options
    - Visible text and imagery
    - QR-Code or other symbology

# Technology Overview - Meeting the Metric for Data Provenance

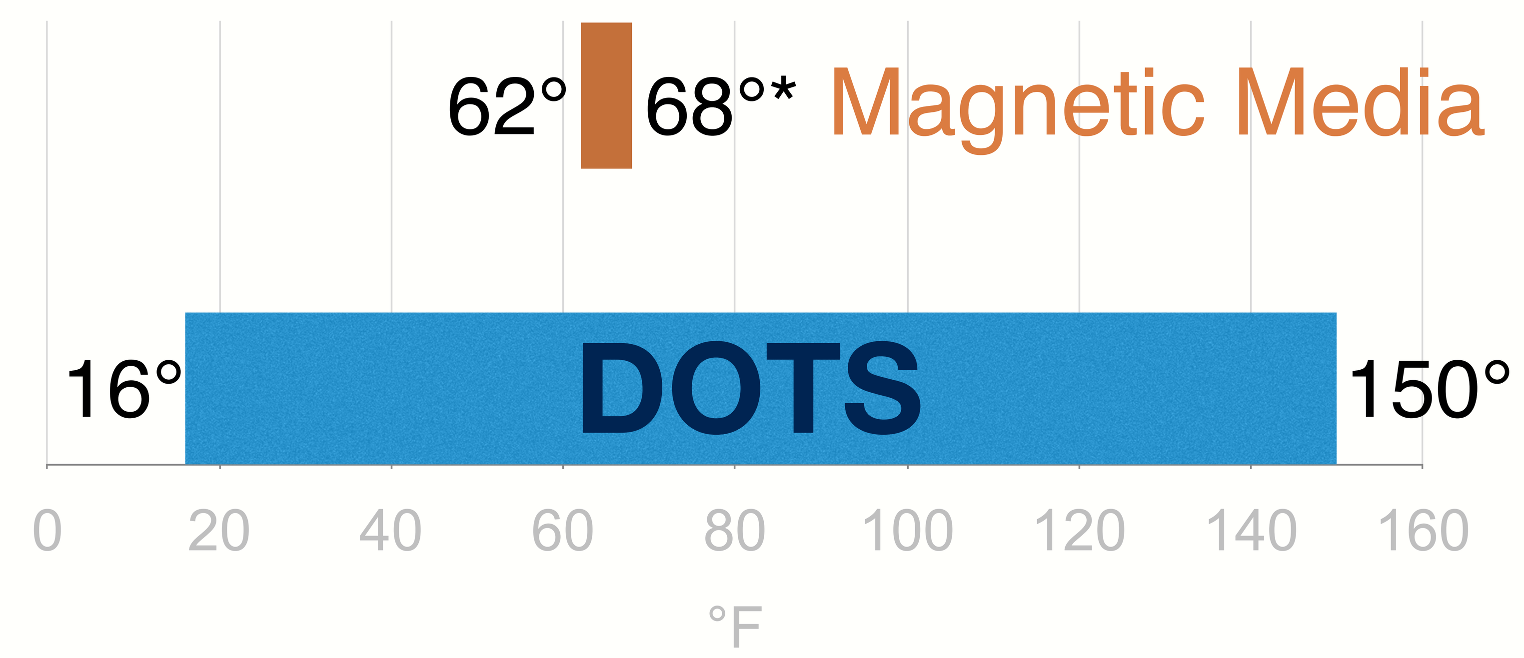
## Magnetic Media vs **DOTS** Visual Metal Alloy Tape

### Stable Humidity Range




* US National Archives and Records Administration Guidelines

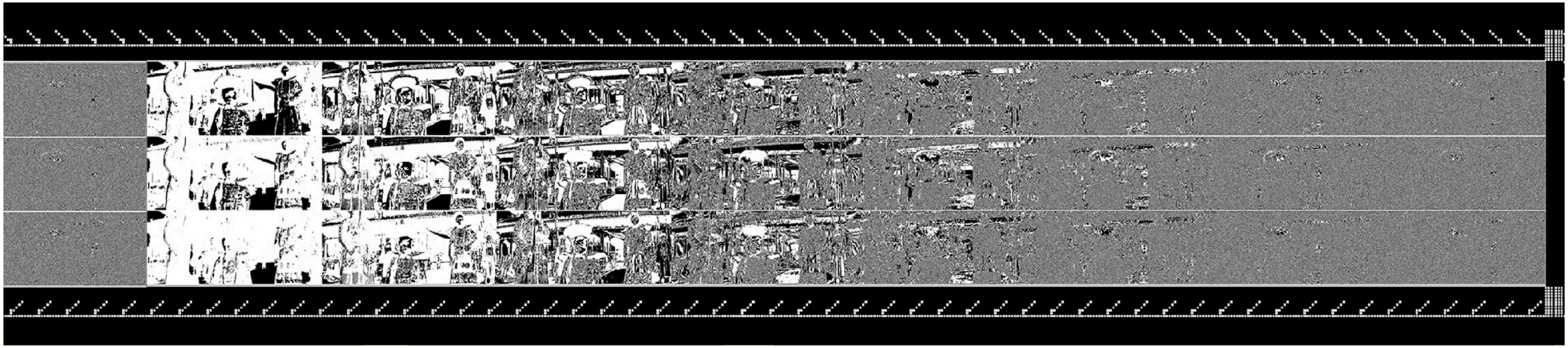
### Stable Temperature Range




## Additional Threats to Data Provenance

| | Magnetic Media | **DOTS** |
|---|---|---|
| Fingerprints / Human Touch | Corrupted Data | No Impact—Can be wiped off |
| Magnetic Interference | Corrupted Data | No Impact |
| Immersion in Water | Corrupted Data | No Impact |
| Hostile Electromagnetic Environment / EMP | Corrupted Data | No Impact |

# **DOTS** Data Scheme



- Bit Plane Image preservation is an image and sound preservation format ***in addition*** to the ability to store any standard files or applications with DOTS
  - Unique to **DOTS**, this format ensures forever readability of image and sound content
  - Immune from operating system incompatibility or file format dependencies

## To be clear….

**DOTS** can store *ANY* digital file – including images and sound

**DOTS** can store the same things you would save to your hard drive

**DOTS** can store any application like Photoshop, Microsoft Word, Excel, or any of the documents these programs create

# Encoding Data to Visual Metal Alloy Tape

Laser "writes" data
by heating metal alloy tape
to a specific temperature

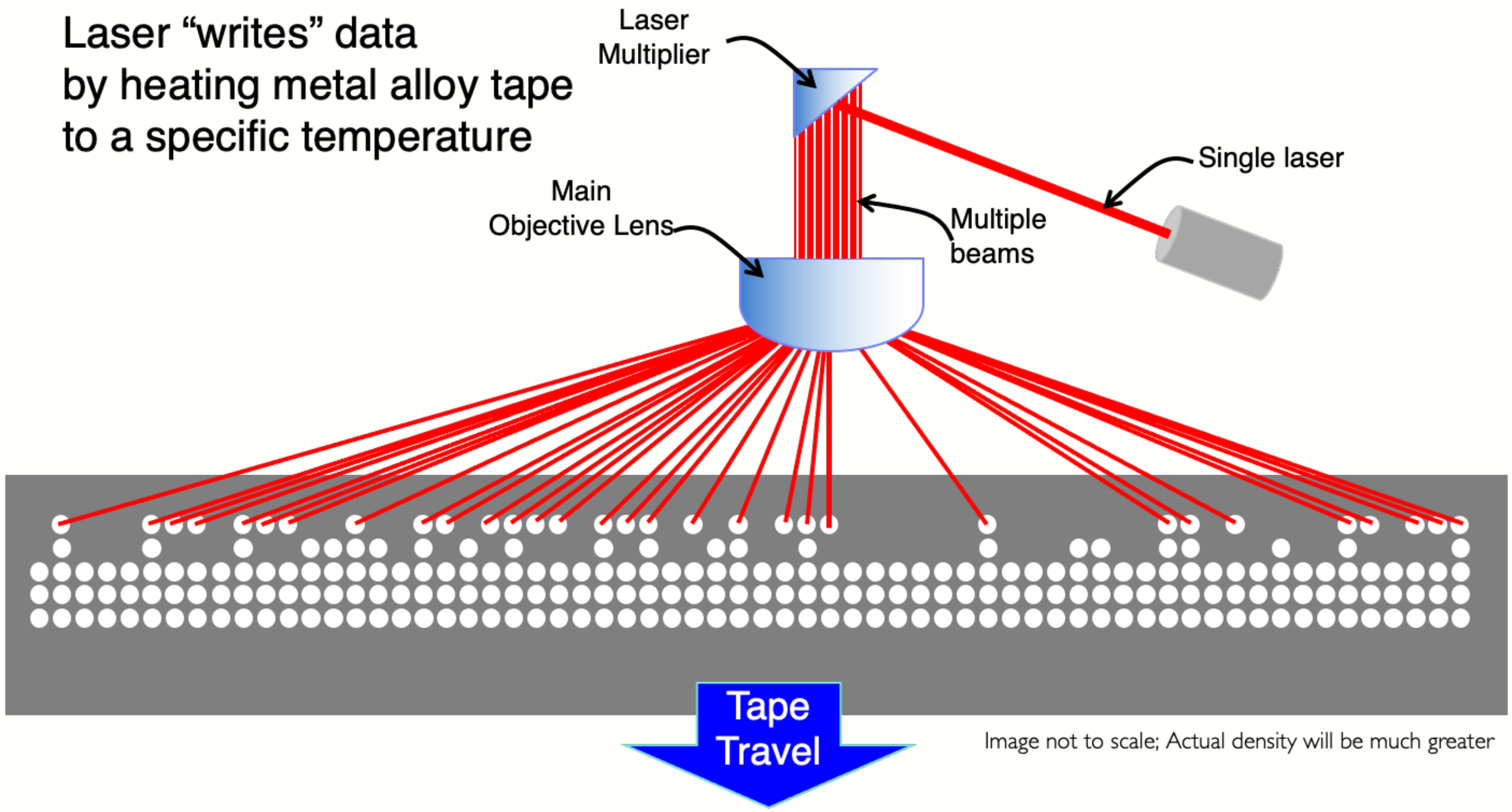



Image not to scale; Actual density will be much greater

Heat of laser causes instant phase change of the metal alloy tape at the molecular level
to create a microscopic visual representation of the data to be preserved.

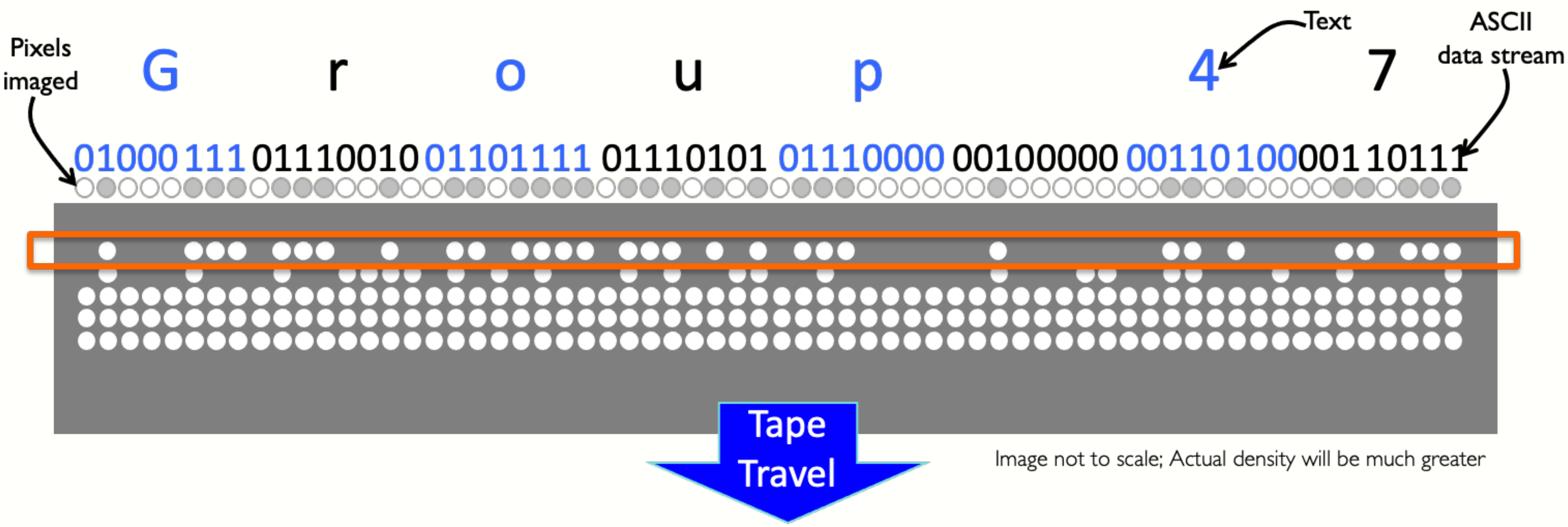


# Encoding Data to Visual Metal Alloy Tape

## Reading a Visual Representation of Digital Data

Digital encoding of the name “Group 47” using **DOTS**



# Encoding Data to Visual Metal Alloy Tape

## Data Density on ¾ inch G47 Visual Metal Alloy Tape

12,288 data spots per write cycle in First Customer Ship
One complete pass to fill up tape

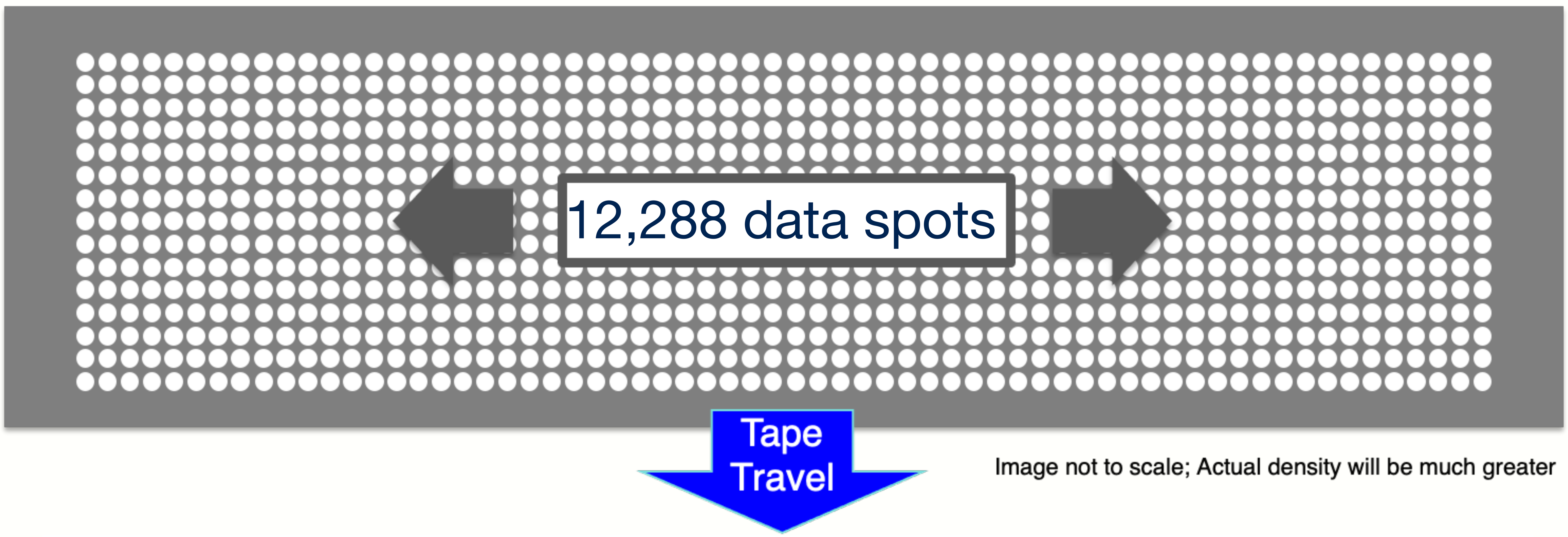



Image not to scale; Actual density will be much greater

G47 Tape read/write speed is 20cm/sec (7.87 in/sec)
For 1 GByte/second read/write speed

# Performance Specifications

## Data Transfer Speed

First Commercial Ship >1 GByte/sec

## Cartridge & Media Characteristics

| | |
|---|---|
| **Gen 1 Capacity:** 20TB Native | **Tracks per pass:** 15,000 |
| **Tape Speed:** 20cm/sec | **Passes to write entire tape:** 1 |
| **File System Support:** LTFS | **Head to Tape Contact:** None |
| **Reel Dimensions:** 12" Dia, ¾" Wide | **Media Life:** No less than 200 Years |
| **NVM Capacity:** Capacity TBD | **Corrected bit error rate (CBER) of $10^{-18}$** |
| **WORM capable:** Yes | **EMP/Magnetic Field Sensitivity:** None |
| **Tape Thickness:** TBD | **Long Term Storage Temperature:** 16-150°F (-9 – 66° C) |
| **Data, Text and Images:** Yes | **Long Term Storage Humidity:** 5%-85% |

All generations of **DOTS** media will be backwardly compatible

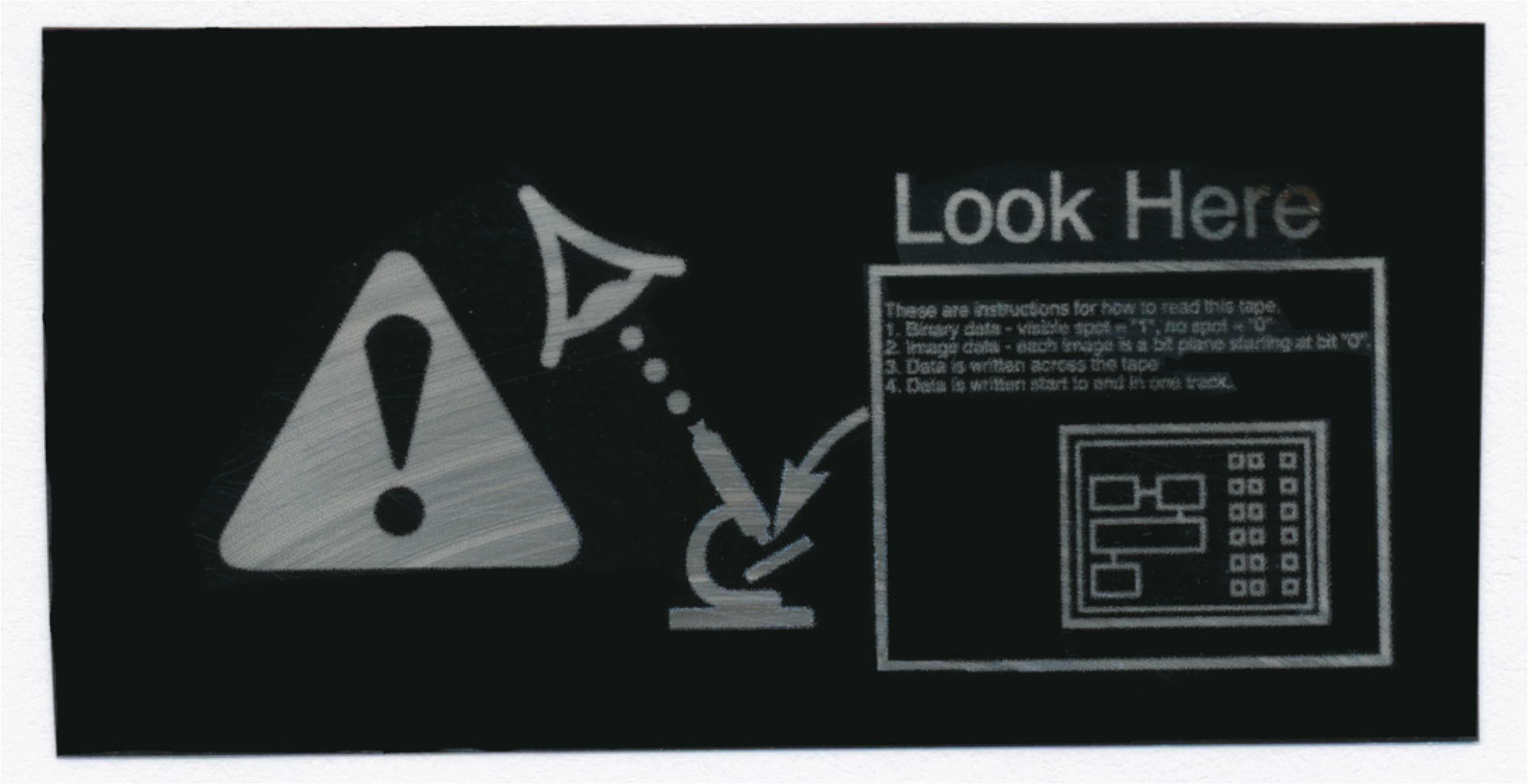


*all specifications are subject to change without notice or obligation and are presented as best estimate at the time of this document.

# “Rosetta Leader™”

- Because **DOTS** is visual, the leader of every tape will be written with human readable text that will have instructions and images on how to construct an optical reader and a full description (including computer source code) of the exact encoding scheme used to write the data.

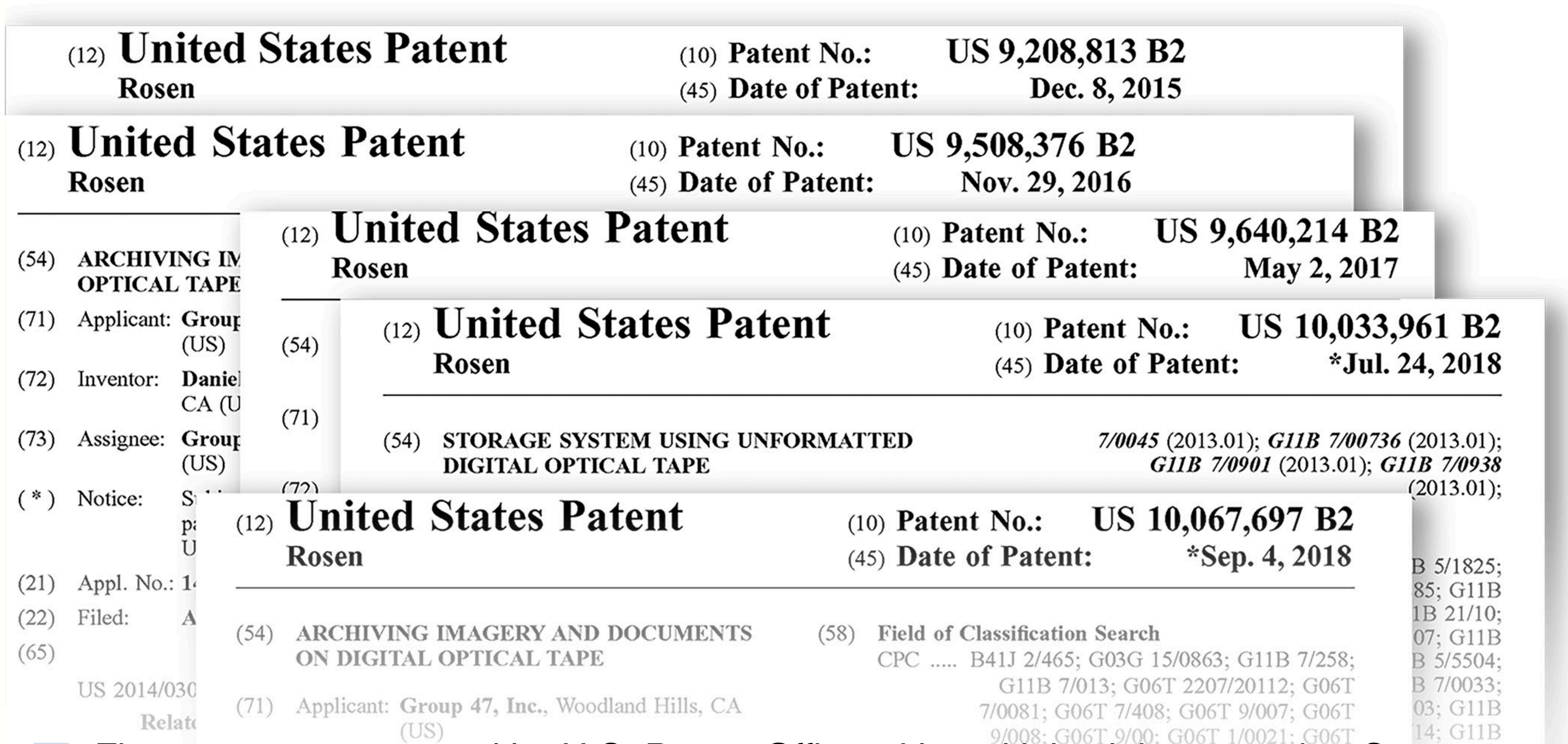
- Since **DOTS** supports LTFS, it is possible to set aside space for textual descriptions of what information is contained on that particular tape.





# Five NEW Patents Awarded to Group 47


(12) United States Patent
Rosen
(10) Patent No.: US 9,208,813 B2
(45) Date of Patent: Dec. 8, 2015

(12) United States Patent
Rosen
(10) Patent No.: US 9,508,376 B2
(45) Date of Patent: Nov. 29, 2016

(12) United States Patent
Rosen
(10) Patent No.: US 9,640,214 B2
(45) Date of Patent: May 2, 2017

(12) United States Patent
Rosen
(10) Patent No.: US 10,033,961 B2
(45) Date of Patent: *Jul. 24, 2018
(54) STORAGE SYSTEM USING UNFORMATTED DIGITAL OPTICAL TAPE

(12) United States Patent
Rosen
(10) Patent No.: US 10,067,697 B2
(45) Date of Patent: *Sep. 4, 2018
(54) ARCHIVING IMAGERY AND DOCUMENTS ON DIGITAL OPTICAL TAPE
(71) Applicant: Group 47, Inc., Woodland Hills, CA (US)
(58) Field of Classification Search
CPC ..... B41J 2/465; G03G 15/0863; G11B 7/258; G11B 7/013; G06T 2207/20112; G06T 7/0081; G06T 7/408; G06T 9/007; G06T 9/008; G06T 9/00; G06T 1/0021; G06T

- First new patent granted by U.S. Patent Office with multiple claims covering Group 47's unique visual approach for writing and reading digital data.
- The second new patent granted covers Group 47's Bit Plane Image method for archiving images (whether images of photos, videos, or documents) that removes all file format dependencies.
  - The new method for archiving images and sound has quickly become one of the most compelling aspects of **DOTS** for all potential customers, since, with it, **DOTS** can guarantee image and sound files can be read securely decades into the future, without concern for operating system.
- Three more patents have been granted, and twelve additional patents are ready to be filed



# Exponential Growth of Migration Compounds the Problem

Discussions tend to focus on one migration of data created in a given year, yet ignore the subsequent years of data created that will also require migration. Soon, you're faced with migrating multiple years of data in one year.

In a three-year migration cycle, you have the first year of data to migrate after three years. After 6 years it will be time for the second migration of year 1 and the first migration for year 4, and so on.

**LTO or Disk 3-year Migration Cycle**

| Migration / Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6th | | | | | | | | | | | | | | | | | | | 1 | 2 | 3 |
| 5th | | | | | | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| 4th | | | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 3rd | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 2nd | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1st | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| Current Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |

**LTO or Disk 5-year Migration Cycle**

| Migration / Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 4th | | | | | | | | | | | | | | | | | | | | | 1 |
| 3rd | | | | | | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 |
| 2nd | | | | | | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 1st | | | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| Current Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |

**DOTS™**

| Migration / Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Current Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |

10 Petabytes of LTO-7 takes over a year to migrate with one drive. Within a few years, it will take several years to migrate a single year's migration requirement.

This illustration assumes NO increase in new current-year data, which is unlikely.

Even with no increase in current data, there will quickly be insufficient time in the year to complete the migration of existing and new archive data.

The expected exponential increase in new archival data will further compound the challenge of migration.

# Contact


**Steve DeWindt**
Chairman & CEO
Phone: +1-714-202-7009
steve.dewindt@group47.com

**Rob Hummel**
President & Founder
Phone: +1-714-202-7009
rob.hummel@group47.com

**Daniel Rosen**
Chief Technology Officer
Phone: +1-714-202-7009
daniel.rosen@group47.com

**Jimmy Kemp**
EVP Federal Systems
Phone: +1-714-202-7009
jimmy.kemp@group47.com