Investigations on Color Microfilm as a Medium for Long-Term Storage of Digital Data

Christoph Voges, Volker Märgner, Tim Fingscheidt;
Technische Universität Braunschweig, Institute for Communications Technology (IfN); Braunschweig, Germany

Abstract

Digital data storage on microfilm is a promising alternative for long-term storage of digital data. Its estimated lifetime of up to 500 years and the availability of reading devices allow entirely new migration-free storage concepts. This paper presents investigations on the suitability of color microfilm as a medium for digital data storage being an alternative to conventional black-and-white film material. The main question we address is whether the advantage of three color channels justifies higher efforts and expenses related to this material. Therefore, an analysis based on several exposed test patterns has been performed. It turned out that the regarded film in combination with the employed exposure setup is very differently capable of storing data points depending on the color layer. Although black-and-white film material has several advantages in our opinion, special cases are pointed out where the use of color microfilm for digital data storage is attractive.

Introduction

In the past few years, microfilm has become an attractive medium for long-term storage of digital data. As an alternative to conventional storage media such as DVDs, CDs, hard drives, or magnetic tapes it offers estimated lifetimes of up to 500 years, depending on the specific film material and storage conditions (see, e.g., [1, 2]). Further advantages of this novel technology are the possibility to use common reading devices (film scanners) as well as the high data integrity level due to the genuine WORM (write once read many) character of the medium itself.

There have already been various investigations on microfilm-based digital data storage with respect to signal and information processing [3], hardware aspects [4], channel modelling [5], modulation coding [6], as well as storage capacities and error correction coding [7, 8]. A detailed review of this technology including an overview of relevant microfilm standards and related systems is provided in [8]. Laser recording technology (see, e.g., [4, 9, 10]) is widely used for digital data storage on microfilm to expose the film material. Systems involving three separately modulated laser beams can be employed to expose both black-and-white as well as color microfilms. However, recent scientific contributions mostly focus on black-and-white microfilm material for digital data storage applications. Furthermore, the laser recording system described in [4] has been developed for high-density exposure of digital data on black-and-white microfilm with high data transfer rates. Of course, this material has several advantages compared to color microfilm, such as the lower price or the less complex photochemical processing.

On the other hand, there are also various motivations for the use of color film material for data storage on microfilm. Firstly, all three color layers can be utilized leading to possible storage capacity improvements. Secondly, for hybrid storage of digital and analog data on the same medium (see [8]), the digital data can be stored along with color images on the same film. This is also an interesting solution for the storage of digital metadata. Some approaches to digital data storage on microfilm already make use of the advantages of color microfilm. Concerning the storage capacity for digital information we encounter the practical problem that the resulting characteristics of the color layers cannot be regarded as independent. The reason for this phenomenon is spectral overlap of the employed dyes leading to mutual influences, similar to so-called crosstalk in a communications system. Also, the optical properties can be different for each color layer.

This contribution provides an analysis dealing with the possibilities offered by color microfilm for digital data storage on film. Therefore, as a starting point, the relevant basic principles as well as the physical background of color microfilm are regarded. Important differences to monochrome film material are identified with respect to data storage applications. The second step is an analysis of dedicated test patterns that have been exposed to color microfilm material by means of a laser recorder. Bit error rates are measured and serve as an objective criterion to compare different alternatives for exposing the data points. A research microscope equipped with a high-resolution camera allows accurate imaging of these samples at various resolutions and therefore serves as a reference reading device. Finally, aim of this contribution is a set of recommendations regarding the use of color microfilm for digital data storage on microfilm based on the described analysis and experiments.

Data Storage on Microfilm

Current approaches to data storage on microfilm are based on laser recording technology (see, e.g., [5, 8]). Therefore, tiny data points are exposed to the film by means of a modulated laser beam that is moved over the film material. These data points represent logical ones or zeros, respectively. The grid space $d$, i.e., the distance between the exposure points is a crucial factor for the storage capacity. Furthermore, amplitude modulation can be used to store more than one bit within a single data point. A detailed analysis concerning the influence of these factors on the storage capacity and a discussion of the resulting trade-offs can be found in [7, 8]. These investigations show that a small grid space $d$ and binary modulation is a reasonable choice. Accordingly, we focus on this type of modulation in this paper.

Several laser recording devices exist for microfilm featuring different technical specifications (see, e.g., [4, 9, 10]). All film samples used for the analysis in this publication are exposed with the Arche laser recorder [10] as a state-of-the-art laser recording device that is capable of handling both black-and-white as well as color microfilm. Besides laser film recording, there are also alternative technologies to computer output microfilm devices (COM) that are
not in further focus of this paper (see, e.g., [11]).

The photosensitive layer of black-and-white film material (see, e.g., [1, 2]) basically consists of silver halide crystals (so-called grains) in gelatine. Simplified, during the exposure process, the laser beam causes photochemical reactions in which parts of the silver halide ions react to metallic silver and halide atoms. A chemical development process serves to transform the silver halide crystals containing such metallic silver atoms completely to metallic silver whereas the other grains remain unaffected – at least in the ideal case. As the remaining silver halide grains are still sensitive to light, a fixing process is required to remove them from the film material. However, the silver grains are not – also at least in the ideal case – affected by the fixing process and form the stable photographic image. For a detailed description of the photographic process please see [11, 12].

When regarding these photochemical reactions, three fundamental facts should be emphasized with respect to digital data storage on black-and-white microfilm: First of all, the photographic image consists of metallic silver grains ensuring a high degree of stability. As an example, the materials described in [1, 2] are expected to achieve a life expectancy of about 500 years. Secondly, the underlying photographic image formation process is well-understood and extensively investigated. Finally, it should be noted that both the material and the photographic process will presumably be available in the foreseeable future.

However, no color can be reproduced with this kind of film material. Also, the image formation during this process is negative, i.e., formerly exposed places appear black on the film and unexposed places appear white or transparent, respectively. Anyway, for storing data points in the context of data storage on film applications this is actually not relevant. The photographic reproduction of color requires both more complex film materials and chemical processes as described in the next section.

**Color Microfilm**

As opposed to black-and-white microfilm, color microfilm generally consists of several color layers. A widespread positive color microfilm material is Ilfochrome® Micrographic [13] that can be exposed, e.g., with the Arche laser recorder. It is available in two versions with different contrast, Type M and Type P, respectively, both being positive films. The Type M material exhibits a higher contrast compared to Type P and is the basis for the investigations within this paper. Both types of this positive film are processed in the P-5 process that mainly consists of three baths: developer, bleach, and fix [13] (as opposed to two baths, developer and fix, for the above-mentioned black-and-white process). Although the image formation process for Ilfochrome® Micrographic also involves silver halides, it is much more complex since the actual image finally consists of organic dyes and the silver is removed during the bleaching process (see, e.g, [12, 13] for more details on color film processing). Basically, there are three sensitive color layers for reproducing the colors blue, green, and red. Although the image for the Ilfochrome® Micrographic film is no longer composed of silver atoms – as it is for the above-mentioned black-and-white microfilm – but merely of organic dye molecules, it also provides an excellent long-term stability (see, e.g., [14]).

For traditional archiving of analog images, color microfilm has the clear advantage of preserving color information of colored documents, such as paintings, photographs, sketches, or drawings. When using laser film recorders or other COM devices, these images can be directly exposed to the film. On the other hand, compared to black-and-white microfilm, color microfilm is more expensive and – as already described – also the chemical processing is more complicated [13]. Accordingly, if color is not relevant, it is reasonable to use black-and-white microfilm instead.

When storing digital data in form of data points on color film it is straightforward to exploit the multiple color layers and use blue, green, and red data points simultaneously. However, it has to be taken into account that the color layers cannot necessarily be assumed as independent. When regarding the normalized spectral dye densities $D_b(\lambda)$, $D_g(\lambda)$, and $D_r(\lambda)$ (with $D_b(\lambda), D_g(\lambda), D_r(\lambda) \in [0, 1]$) of the Ilfochrome® Micrographic film depending on the light wavelength $\lambda$ in [13], it is obvious that there is a certain overlap of the spectral characteristics. The blue-sensitive dye has its maximum density at a light wavelength of about $\lambda_b = 425$ nm, the green-sensitive at approximately $\lambda_g = 570$ nm, and the red-sensitive dye has its maximum density at around $\lambda_r = 635$ nm as well as a second local maximum with a slightly lower density level at about $\lambda_{r2} = 685$ nm. All values $D_b(\lambda)$, $D_g(\lambda)$, and $D_r(\lambda)$ are provided in Table 1 for these wavelengths. Especially the blue-sensitive color layer is influenced by green and also red. Furthermore, the green characteristics are significantly influenced by the red-sensitive dye. This effect is similar to crosstalk in a communications system [15] and is referred to as spectral overlap in the following.

### Experimental Setup

To analyse the performance of color microfilm, several test patterns have been exposed to Ilfochrome® Type M microfilm [13] by means of the Arche laser recorder [10]. All pseudorandom patterns are exposed using binary modulation according to the recommendations given in [7]. These test patterns allow to determine gross bit error rates (i.e., bit error rates without any error correction – simply referred to as bit error rates or BERs in the following) by comparing the known written bit constellation to the bits after the read-out process. A research microscope with a high-resolution camera was employed as a reference setup to carry out the read-out process for our experiments. Synchronization, i.e., the precise detection of the position of each data point, was performed as described in [3].

Table 1: Normalized spectral dye densities $D_b(\lambda)$, $D_g(\lambda)$, and $D_r(\lambda)$ at different wavelengths $\lambda$ (values coarsely reconstructed from a figure in [13]).

<table>
<thead>
<tr>
<th>$\lambda$ (values coarsely reconstructed from a figure in [13]).</th>
<th>$D_b(\lambda)$</th>
<th>$D_g(\lambda)$</th>
<th>$D_r(\lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = \lambda_b = 425$ nm</td>
<td>1.00</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td>$\lambda = \lambda_g = 570$ nm</td>
<td>0.00</td>
<td>1.00</td>
<td>0.45</td>
</tr>
<tr>
<td>$\lambda = \lambda_r = 635$ nm</td>
<td>0.00</td>
<td>0.05</td>
<td>1.00</td>
</tr>
<tr>
<td>$\lambda = \lambda_{r2} = 685$ nm</td>
<td>0.00</td>
<td>0.00</td>
<td>0.95</td>
</tr>
<tr>
<td>$\lambda = \lambda_{r2} = 429$ nm</td>
<td>1.00</td>
<td>0.20</td>
<td>0.05</td>
</tr>
<tr>
<td>$\lambda = \lambda_{r2} = 529$ nm</td>
<td>0.05</td>
<td>0.80</td>
<td>0.20</td>
</tr>
<tr>
<td>$\lambda = \lambda_{r2} = 672$ nm</td>
<td>0.00</td>
<td>0.00</td>
<td>0.90</td>
</tr>
</tbody>
</table>
system (see also [7, 8, 16]) to cope with the gross BER. However, for our analysis the BER values are merely used as a criterion for the suitability of the different alternatives for exposing the data points. Accordingly, error correction coding is not within further scope of this contribution.

For a first analysis, the patterns have been exposed by using a single laser only, referred to as single-color exposure in the following. These experiments allow to study the performance and capabilities of each color layer separately. In a second step, the test patterns for blue, green, and red have been jointly exposed at the same place on the film (multi-color exposure). In turn, this allows to evaluate the interactions of the color layers especially due to spectral overlap. All analyses have been performed for grid spaces \( d = 6 \mu m \) and \( d = 9 \mu m \). Since the color layers are located at different depths within the film material, the focus was adjusted before imaging of each pattern during the read-out process. Note that the Arche laser recorder was originally developed for exposure of analog images and accordingly its settings are optimized to achieve good results in this context. For digital data storage on microfilm, the parameters have to be chosen differently. In our analysis, a certain set of exposure parameters is regarded that lead to good results. However, note that further optimization of the exposure parameters may lead to even better results in the future. To achieve comparable results for the bit error rates among the different test patterns, only those patterns have been employed for the analysis that are almost free of dust and scratches which can occasionally occur in a microfilm-based storage system (cf., e.g., [8]). This must also be considered when using the bit error rates of this analysis for redundancy estimations regarding error correction coding as presented in [7, 8].

For all BER measurements in this paper, the camera of the microscope was operated in grayscale mode. Three different color filters were used to distinguish between the blue, green, and red color channels. Therefore, the filters were inserted into the optical path of the microscope’s illumination. Accordingly, the film samples were either illuminated by blue, green, or red light. All filters exhibit a transmission bandwidth of approximately 50 nm at center wavelengths \( \lambda_{Fb} = 429 \text{ nm} \) (blue), \( \lambda_{Fg} = 529 \text{ nm} \) (green), and \( \lambda_{Fr} = 672 \text{ nm} \) (red), respectively. These wavelengths are also given in Table 1 along with the corresponding normalized spectral dye densities \( D_{b}(\lambda) \), \( D_{g}(\lambda) \), and \( D_{r}(\lambda) \). The blue wavelength \( \lambda_{Fb} \) is chosen to be located near the maximum value of \( D_{b}(\lambda) \). Table 1 shows that especially \( D_{g}(\lambda) \) still exhibits a considerable value at \( D_{g}(\lambda_{Fb}) = 0.20 \). For the blue dye this cannot be significantly improved since the position of the maximum of \( D_{b}(\lambda) \) corresponds approximately to a local minimum of \( D_{g}(\lambda) \) (see [13] for a graphical representation of the normalized spectral dye densities of the Ilfochrome Micrographic microfilm). The green filter was chosen to a center wavelength of \( \lambda_{Fg} = 529 \text{ nm} \) being about 41 nm apart from the maximum value of the green-sensitive dye. This approach has the clear advantage of reducing the influence of the red-sensitive dye. In a similar way, the filter wavelength \( \lambda_{Fr} = 672 \text{ nm} \) was chosen to be at a distance of about 37 nm from the maximum of \( D_{r}(\lambda) \).
at a wavelength where the influence of both the green-sensitive and the blue-sensitive dyes can be neglected since $D_g(\lambda_F) = 0.00$ and $D_b(\lambda_F) = 0.00$.

### Measurement Results

Figures 1 to 6 show microscopical images of the investigated test patterns. Those patterns based on single-color exposure can be regarded in Figures 1 and 2 for $d = 9 \, \mu m$ and $d = 6 \, \mu m$, respectively. It is directly obvious from these images that the data points appear very clear for both the blue and the green channels. Of course, the data points are better distinguishable for the larger grid space $d = 9 \, \mu m$. When regarding the patterns exposed with the red laser, it appears that the data points are barely visible, due to both resolution and contrast. Also, these data patterns appear noisy. Obviously, the green and blue data points are substantially better suited for digital data storage on microfilm. It is even questionable from regarding these images, whether the red data patterns can be reasonably employed to store digital data for the investigated grid spaces ($d = 9 \, \mu m$ and $d = 6 \, \mu m$) and exposure parameters.

In Figures 3 and 5, microscopical images are provided where blue, green, and red data patterns are jointly exposed at the same positions on the film (colors can only be seen in PDF version of this paper that is available online and not printed in the conference proceedings). Images involving the color filters with center wavelengths $\lambda_{Fb}$, $\lambda_{Fg}$, and $\lambda_F$, are provided in Figures 4 and 6. Again, the green and the blue exposed channels provide very clear data points whereas the red exposed channel leads to merely noisy data patterns. When regarding the blue exposed channels in Figures 4(a) and 6(a), the spectral overlap can be observed since the green data points are still slightly visible as dark data points when applying the blue filter with $\lambda_{Fb} = 429 \, nm$. This is a prospective source of bit errors.

As an objective criterion for the analysis, BERs have been measured on the exposed pseudorandom test patterns as described in the last section. The results of this analysis are presented in Table 2. Due to their poor quality, no reliable read-out of the red test patterns has been possible. Accordingly, only bit error rates for the blue and green exposed channels are given in Table 2. For single-color exposure, very low BER values are achieved for blue and green at both grid spaces $d = 9 \, \mu m$ as well as $d = 6 \, \mu m$. When comparing these low BER values it should be considered that the total number of data points (or bits, since binary modulation was employed) for the investigation was 10,000, meaning that a BER of 0.0001 corresponds to a single bit error only. When regarding multi-color exposure, it can be observed that the green channel leads to similarly low bit error rates for $d = 6 \, \mu m$ and $d = 9 \, \mu m$. However, concerning the blue channel, the bit error rate for multi-color exposure leads to significantly higher values compared to green or single-color exposure (blue and green), respectively. Furthermore, for multi-color exposure the BER value for blue at $d = 6 \, \mu m$ has turned out to be highly dependent on the lightening conditions. As an example, a stronger illumination has caused the BER to rise from 0.0650 to a value of even 0.2554. On the other hand, the green channel seems to be almost unaffected by spectral overlap effects.

Since the bit error rates are approximately identical for $d = 6 \, \mu m$ and $d = 9 \, \mu m$ (except for blue when using multi-color exposure), it can be concluded that intersymbol interference (ISI) (see also [6, 17]) can be neglected for both grid spaces and single-color exposure (blue and green) as well as multi-color exposure (at least for green). Accordingly, bit errors must be assumed to occur due to residual errors (e.g., small dust particles that cannot be avoided even when carefully handling the film samples) and of course spectral overlap, especially for blue within the multi-color exposure experiments.

### Discussion and Recommendations

The experiments described in the last two sections show that the color layers are differently suited for storing digital data in form of data points. The green channel performs very well and exhibits low BERs for both single color as well as multi-color exposure at both grid spaces. For single-color exposure, blue offers similarly low BER results but is susceptible to spectral overlap in a multi-color exposure setup. Also, for $d = 6 \, \mu m$ a strong dependence on the lightening conditions has been observed for the blue channel when using multi-color exposure. It should be emphasized that all BERs presented in Table 2 for green and blue can be reasonably handled by error correction codes. Please see also [7, 8] for discussions about these codes and redundancy considerations. On the other hand, for the investigated setup, red leads to no sufficient results for grid spaces $d = 6 \, \mu m$ and $d = 9 \, \mu m$. Of course, wider grid spaces and larger exposure points may finally solve this problem but this would then result in unattractive storage capacities due to the inversely quadratic dependency of the storage capacity on the grid space $d$ (see [6–8]). At least for the regarded setup and exposure parameters, the effort to store data using the red-sensitive layer is questionable compared to its benefits.

On the other hand, it is straightforward to store data in the blue-sensitive and green-sensitive color layers simultaneously. However, through this approach the overall storage capacity obviously cannot be doubled and a factor to green or blue single-color exposure accordingly takes on values less than two. The reason for this factor not being exactly two is spectral overlap leading to higher redundancies required for error correction coding of the blue channel compared to single-color exposure only. For a practical application it should be very carefully weighted if the use of two color channels is really appropriate. Regarding the read-out process, the required scanning or imaging equipment for color is certainly more complex compared to a setup for single-color exposure or black-and-white film, respectively.

Also, it should be evaluated if the use of color microfilm is really useful since the write process also requires color laser recorders (or generally color COM devices) and the chemical processing of the color film material is more complicated compared to black-and-white film as discussed earlier. Moreover, color film material is gen-

<table>
<thead>
<tr>
<th>grid space</th>
<th>exposed channel(s)</th>
<th>analysed channel</th>
<th>bit error rate (BER)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$6 , \mu m$</td>
<td>blue</td>
<td>blue</td>
<td>0.0004</td>
</tr>
<tr>
<td>green</td>
<td>green</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>blue, red, green</td>
<td>blue</td>
<td>0.0650$^*$</td>
<td></td>
</tr>
<tr>
<td>blue, red, green</td>
<td>green</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td>$9 , \mu m$</td>
<td>blue</td>
<td>blue</td>
<td>0.0007</td>
</tr>
<tr>
<td>green</td>
<td>green</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>blue, red, green</td>
<td>blue</td>
<td>0.0029</td>
<td></td>
</tr>
<tr>
<td>blue, red, green</td>
<td>green</td>
<td>0.0002</td>
<td></td>
</tr>
</tbody>
</table>

$^*$Value highly dependent on lightening conditions!
Conclusions

In this contribution we analyzed color microfilm as a medium for long-term storage of digital data. The main question we address is whether higher efforts and expenses related to this material compared to conventional black-and-white microfilm justifies its use. As an objective criterion, bit error rates have been measured for single-color exposure (i.e., exposure by means of only a single laser) as well as multi-color exposure (i.e., data patterns of different color jointly exposed at the same place on the film) at different grid spaces. To ensure comparable results, only clean data patterns have been selected. For the investigated setup it turned out that reasonable data storage is possible at a grid space of $d = 6 \mu m$ for both green and blue exposed patterns. On the other hand, no practicable storage of digital data has been possible for the patterns exposed with the red laser – at least not for the selected exposure settings and setup parameters.
For blue and green multi-color exposure it turned out that the storage capacity is less than twice the capacity compared to single-color exposure employing green or blue data points only. The reason for this is spectral overlap due to the spectral characteristics of the different dye materials that are used for the color layers.

Regarding the applications of color microfilm, hybrid recording of digital information along with analog photographic images is highly attractive. The digital information can, e.g., be digital metadata or even a digital version of the human readable analog photographic image. Due to the less complex read-out setup it is meaningful to use only a single color for data storage in cases where storage capacity is a minor issue. On the other hand, when storing only digital information on microfilm (or for hybrid storage involving black-and-white image information only) the use of black-and-white microfilm instead is recommended because both the exposure as well as the read-out equipment is more complex for color microfilm. Also, color microfilm is generally more expensive and requires more complicated chemical processing.

As an outlook on further developments it has to be remarked that digital data storage on microfilm is a relatively new development. By optimizing the exposure parameters, the read-out setup, the exposure hardware, or even the film material, even better results may be achieved in the future. Also, optimized advanced signal processing and error correction coding for digital data storage is currently under development. After all, large scale practical tests are required for this new technology that is a highly attractive solution to store digital information for several hundred years.

Acknowledgments

The Arche laser recorder was developed at Fraunhofer IPM. All analyzed film samples have been exposed with the Arche laser recorder through service provider media de lux GbR, Offenburg, Germany and Fraunhofer IPM in Freiburg, Germany.

References


Author Biography

Christoph Voges received his Dipl.-Ing. degree in Electrical Engineering from Technische Universität Braunschweig, Germany, with a major in information technology in 2005. During his studies he visited the University of Southampton, UK, as an exchange student. After graduating, he joined the Institute for Communications Technology in Braunschweig, Germany. His particular research interests are modulation, channel coding, and channel models for microfilm as a medium for long-term data storage. He is also a member of the ITG Technical Committee 3.4 ‘Film Technology’ of the Association for Electrical, Electronic, and Information Technologies (VDE), Germany.

Volker Märgner received his Dipl.-Ing. and Dr.-Ing. degrees in Electrical Engineering from Technische Universität Braunschweig, Germany, in 1974 and 1983, respectively. Since 1983 he has been working at Technische Universität Braunschweig. Currently he is a member of the research and teaching staff at the Institute for Communications Technology. His main areas of research are image processing and pattern recognition.

Tim Fingscheidt received his Dipl.-Ing. and Dr.-Ing. degrees in 1993 and 1998, respectively, from RWTH Aachen, Germany. After further research on joint source and channel coding at AT&T Labs, Florham Park, NJ, U.S.A., he joined Siemens AG and later Siemens Corporate Research, where he led teams and standardization activities on speech signal processing and transmission. Since 2006 he is Professor for Signal Processing at the Institute for Communications Technology at Technische Universität Braunschweig, Germany. His research interests include source and channel coding as well as speech signal processing.