

Un groupe de travail commun à deux Académies

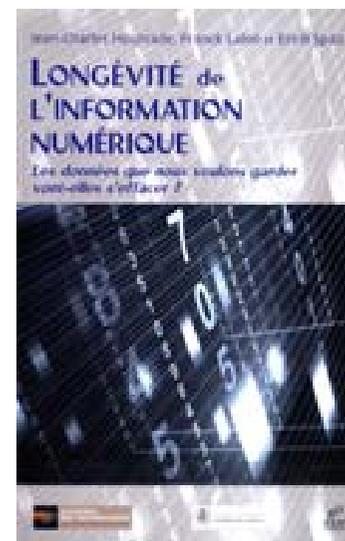
Rapport du
groupe PSN (pérennité des supports numériques)
commun aux deux Académies, des Sciences et des Technologies

Membres du groupe :

Erich Spitz, Académie des Sciences et Académie des Technologies, Président
Jean-Charles Hourcade, Académie des Technologies
Franck Laloë, LKB/ ENS, rapporteur

16 auditions des experts, des laboratoires, des industriels et des utilisateurs
entre septembre 2008 et juin 2009

8 mars 2011 ANR
Académie des Sciences



Préservation à long terme de l'information numérique

Un sujet qui concerne la plupart des activités de la société:

- Données scientifiques (CERN, NASA, CNES, bases de données biologiques, stellaires, chimiques, etc.)
- Données industrielles et commerciales (plans, comptabilité, etc.)
- Données administratives (Etat, collectivités locales, départements, cadastre, suivi des carrières et retraites)
- Données médicales (dossiers des patients, résultats examens, photos, etc.)
- Données personnelles (souvenirs familiaux, comptes, etc.)
- Patrimoine artistique (numérisation d'œuvres, BNF, INA, ..)
- etc.

L'UNESCO estime en Exaoctets (10^{18}) la production annuelle d'information numérique

Une partie de cette information doit être conservée longtemps (des décennies ou plus); ceci pose un problème qui, à l'étude, se révèle très préoccupant.

Plan de l'exposé

- I. Les DONE (Disques Optiques Numériques enregistrables), un support universellement utilisé
- II. L'archivage et ses méthodes
- III. Vieillessement naturel et artificiel des DONE
- IV. La situation actuelle

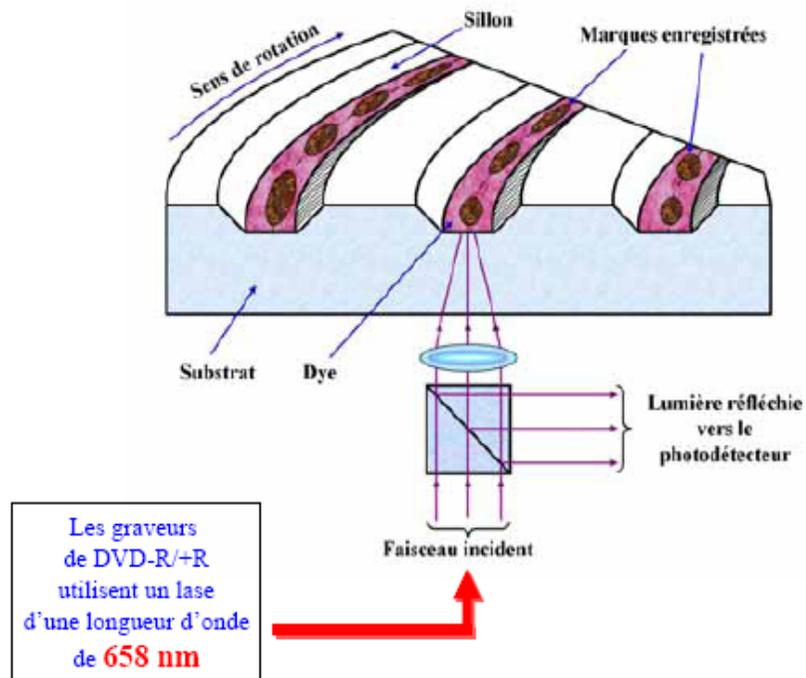
I. Les DONE, un support universellement utilisé

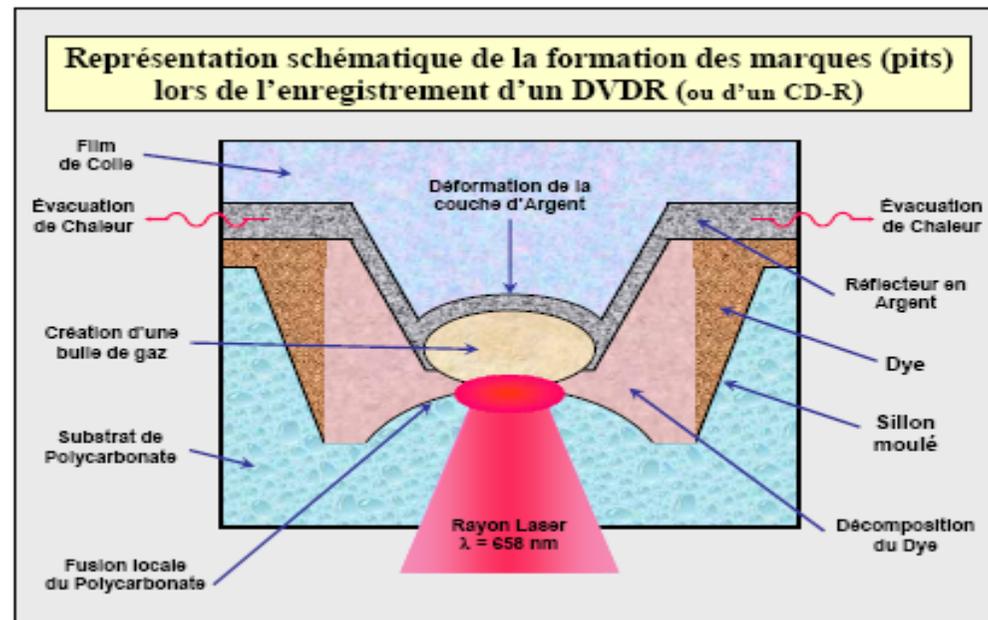
Les disques optiques numériques **enregistrables** (DONE*) sont très commodes d'emploi, rapides à « graver », compacts (envois par la poste,...)

On fabrique chaque année plus de 10 milliards (10^{10}) de DONE (disques optiques numériques enregistrables), CD-R, DVD-R, BD-R.

*attention à ne pas confondre avec disque pressé

Structure générale d'un DVD Enregistrable : DVD-R ou DVD+R





L'effet du laser est une petite explosion locale, qui brûle le colorant, crée une cavité dans le polycarbonate et déforme plus ou moins la couche métallique (argent ou aluminium).

Tous ces effets se combinent de façon complexe pour modifier la réflectivité du disque lors de la lecture, et permettre la détection optique des signaux numériques.

Les DONE (*) sont-ils fiables?

On a cru pendant longtemps qu'ils permettraient une conservation extrêmement longue des données numériques, pour deux raisons:

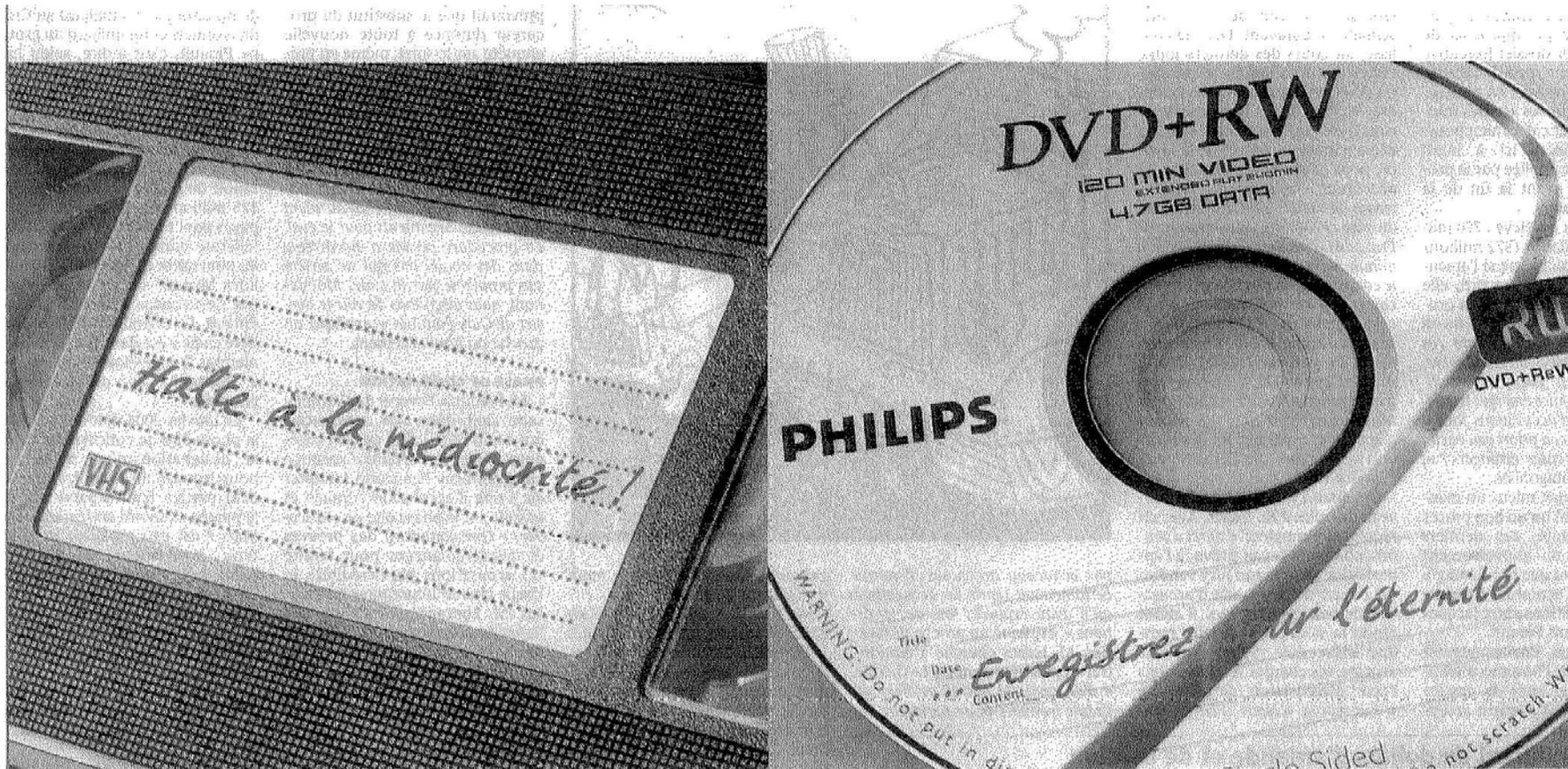
-Ils ne s'usent pas à lecture (pas de contact mécanique)

- Ils utilisent des codes de correction d'erreur qui permettent des recopies instantanées et pratiquement à l'infini (la copie peut être meilleure que l'original, si ce dernier est dégradé).

Mais sont-ils vraiment fiables à long terme? Conservent-ils l'information enregistrée lorsqu'on les stocke soigneusement dans un coffre fort pour des décennies? Peuvent-ils jouer le rôle que jouent les archives sur papier, soit familiales, soit institutionnelles, qui durent des siècles?

Quel est le message de la publicité au grand public?

(*) DONE = Disque Optique Numérique Enregistrable



Pour ne plus voir la qualité de vos enregistrements sur VHS disparaître avec le temps, entrez, **pour moins de 500 euros**, dans le monde du DVD Recorder. Avec la nouvelle gamme de lecteurs-enregistreurs DVD Philips simples d'utilisation et accessibles à tous, vous êtes assurés de **conserver pour l'éternité des enregistrements parfaits et inaltérables.**

www.philips.com/dvdr

Le Monde, 13 décembre 2003

Vos FILMS et VIDÉOS de famille sauvegardés à vie sur DVD !

*DVD DE FAMILLE: le spécialiste de la sauvegarde
de vos archives audiovisuelles familiales.*

TRANSFERT SUR DVD DE TOUS LES FORMATS
Super 8 - 8 mm - 9,5 mm - 16 mm
VHS - Hi8 - mini DV - Umatic - Betacam - Diapositives

TEST GRATUIT sur 3 minutes de film.
Prêt sur place d'une visionneuse pour repérage
et sélection de vos films.

SERVICE DE MONTAGE A LA CARTE
des meilleurs moments de vos films et vidéos.
TRANSFERT SUR CD AUDIO
de vos bandes, cassettes audio et disques vinyl.

Enlèvement gratuit sur Paris de vos films par le coursier DVD de Famille.

DVD DE FAMILLE

Depuis plus de 7 années d'existence, des milliers de clients ont confié à DVD DE FAMILLE leur patrimoine familial audiovisuel. DVD DE FAMILLE est régulièrement cité par les médias comme le prestataire vous garantissant l'expérience, la qualité, la rapidité, au juste prix.

57, RUE RENNEQUIN
75017 PARIS
Métro : Péreire

Ouvert du lundi au samedi
de 10 h à 19 h, sans interruption.

2, RUE DE LONGCHAMP
92200 NEUILLY / SEINE
Métro : Pont de Neuilly

Ouvert du lundi au vendredi
de 10 h à 13 h, et de 14 h à 19 h.

Devis immédiat par téléphone : 01 44 05 02 22

www.dvddefamille.com

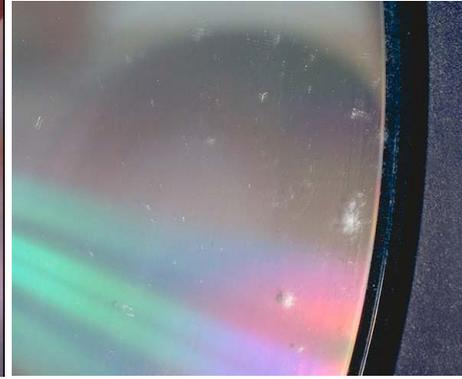
Mais, dans la réalité...



cloques



moisissures



exposition à la
lumière

+ délitement des couches du sandwich par diffusion des molécules d'eau; etc.
En chimie: Phase d'induction et puis phase rapide de destruction

Les DVD sont des dispositifs complexes, comprenant beaucoup de composants organiques (polycarbonate, colorant), et mettant en jeu en parallèle des processus d'écriture et de lecture divers qui sont assez mal connus.

II. L'archivage à long terme(*) et ses méthodes

(*) dizaines ou centaines d'années.

L'archivage à long terme est un processus complexe

Collecte des informations
Numérisation
Création de contenus numériques natifs

- Structuration des données et documents
- métadonnées
- Formats, etc.

Conservation dans le temps
(stratégie active ou passive)

- Stratégie active ou passive
- Infrastructures et personnels (vérification, migrations, etc.)
- Supports: accidents (crash, inondations, usure, manipulations, ..)

VIEILLISSEMENT NATUREL

Communication (serveurs rapides, etc.)
Outils (moteurs) de recherche intelligents
Agrégation
Enrichissement
Certification (intégrité, conformité)
Confidentialité si requis

- Autoroutes de l'information
- Propriété intellectuelle (DRM, etc.)
- Evolution en fonction de la demande (enrichissements, liens etc.).

axe des temps

Le rapport n'aborde pas tous les aspects de l'archivage, en particulier ne parle pas de :

- Conserver toute l'information; il est évident **qu'une forte sélection est nécessaire** parmi toute l'information que nous voyons passer chaque jour
 - Les formats, les logiciels*, les accidents et pannes soudaines, les mauvaises utilisations
 - Les questions de matériel de gravure ou de lecture
 - Les normes
 - Les solutions futuristes, les prototypes de laboratoire
- .. mais uniquement du vieillissement des supports eux-mêmes. C'est le cœur du problème, la composante incontournable qui commande la solution des autres problèmes.

Quelles sont les stratégies d'archivage numérique possibles?

- La stratégie **passive** « Archive et Oublie », traditionnelle, la plus naturelle.
- La stratégie **active** de migration perpétuelle (BNF, CNES, banques, etc.).
- La sous-traitance à un prestataire de service (rôle de Internet); **active**.
- Le retour à l'analogique (rapport du Science and Technology Council); **passive**.

Seule la seconde est appliquée à grande échelle par des institutions dans le monde entier. Elle met généralement en jeu de gros robots (SUN, IBM) manoeuvrant des bandes magnétiques dans des cassettes; elle demande toutefois de gros moyens, en matériel, locaux, et personnel. Est-elle transposable au niveau des individus et des familles?

Quels sont supports d'information numérique?

(hormis DONE)

- Les **bandes magnétiques**; économiques, elles permettent de stocker de grandes quantité d'information, mais en accès relativement lent.

Elles vieillissent: le support devient cassant, l'oxyde se décolle. Durée de vie estimée: **quelques années (moins de 10 ans)**.

- Les **disques durs magnétique**: grande capacité, accès très rapide, leur capacité a énormément augmenté, les prix ont fantastiquement baissé. Prix grand public de 2 To: 150 Euros environ!

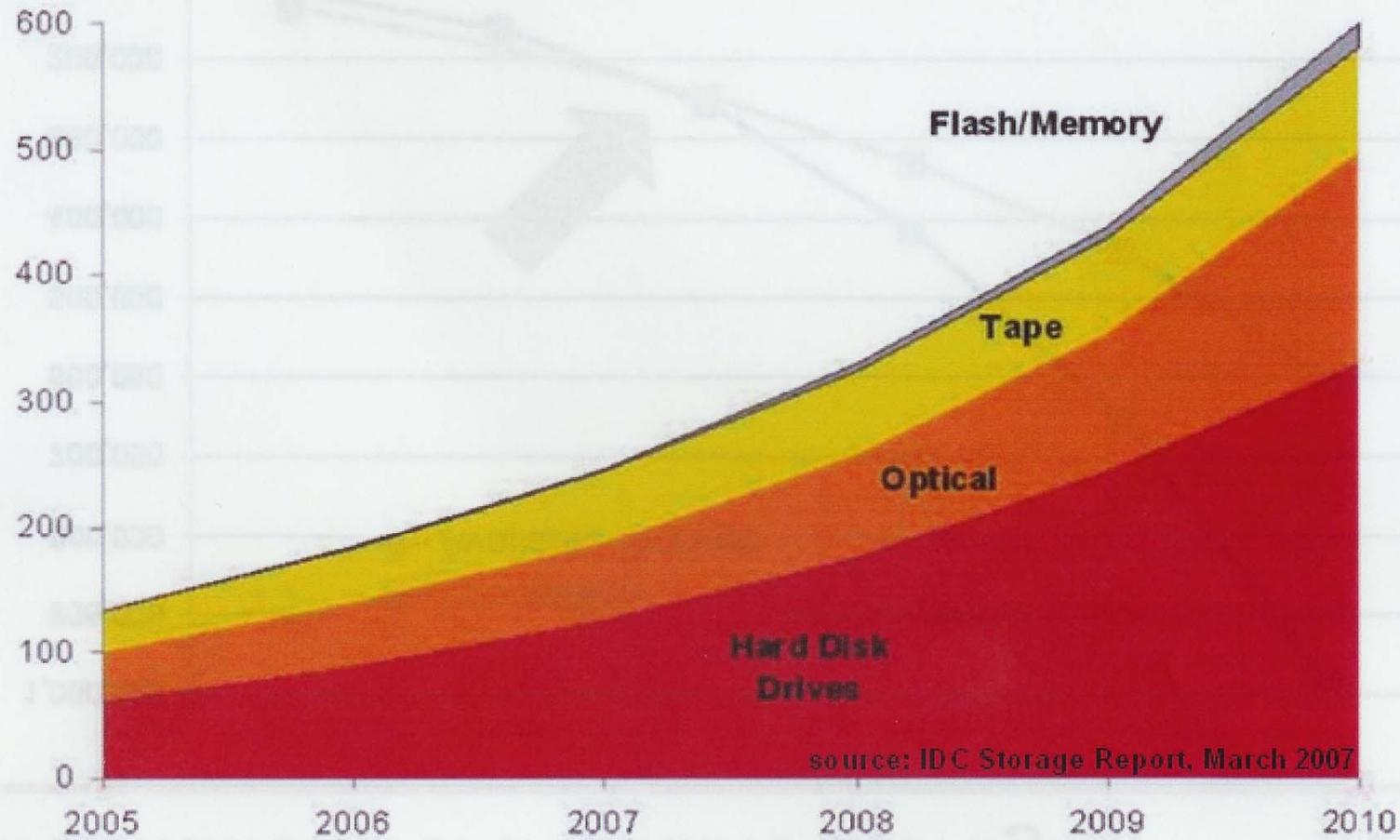
Ils sont sujets aux pannes soudaines (crashes), leur vieillissement est difficilement prévisible. Durée de vie moyenne estimée: **quelques années**.

- Les **mémoires flash** sont de plus en plus répandues, rapides et très commodes. Elles conservent l'information sous forme de charges électriques, et ne sont pas conçues pour conserver l'information pendant des années. On parle de **5 ans environ, parfois moins**.

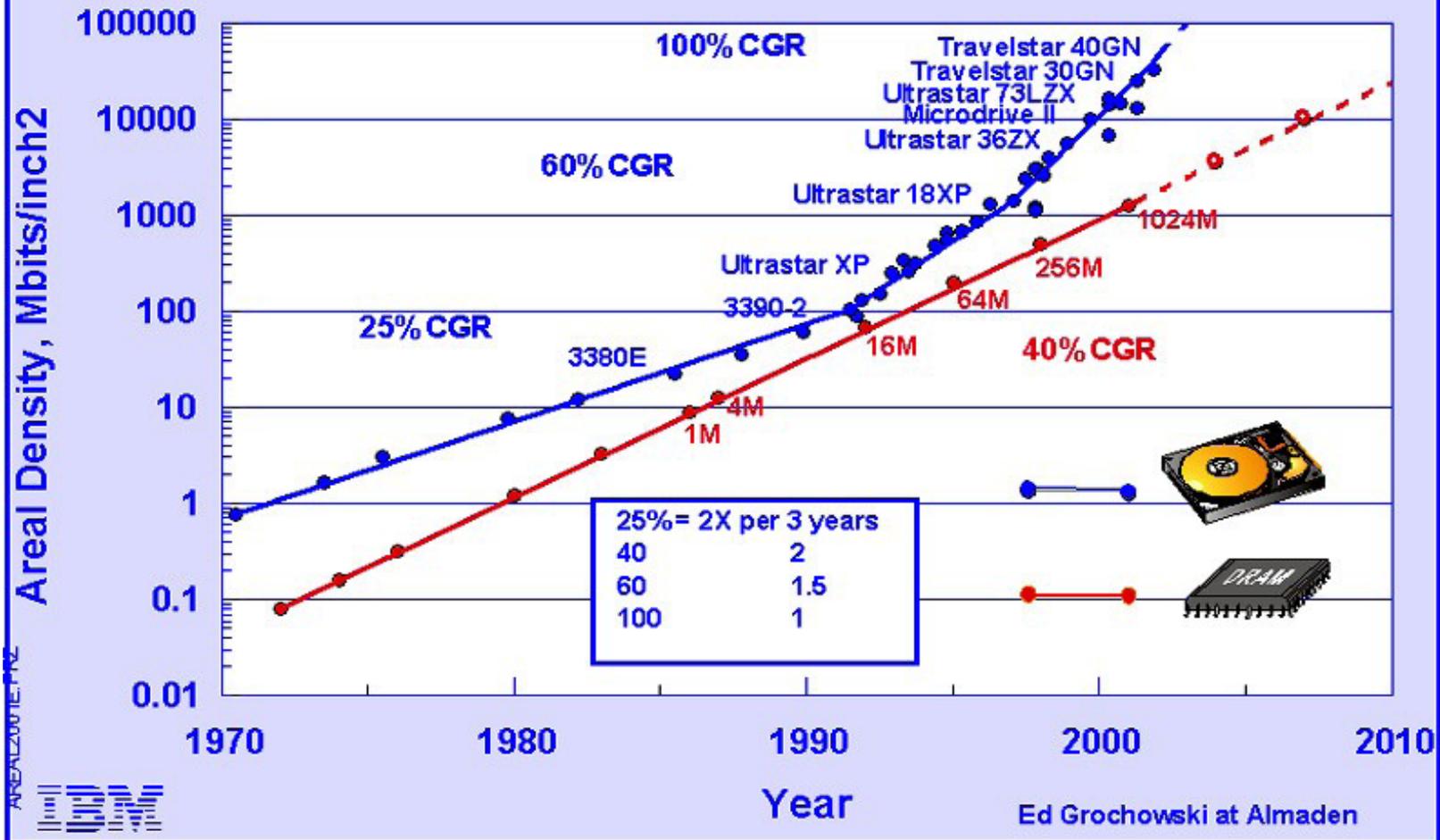
- Il existe bien de projets destinés à mettre sur le marché de nouveaux supports.

Available storage worldwide

Exabytes



Areal Density of Magnetic HDD and DRAM



Paramètres pour choisir/développer une solution

Capacité

Densité de stockage (surface et
mémoires multivaluées)

Temps d'accès

Accès séquentiel ou aléatoire

Vitesses d'écriture et de lecture

Cyclage (réécriture)

Temps de rétention (stockage ?)

Technologie existante / concurrente

Codage

Coût

III. Vieillesse naturel ou artificiel des DONE

Mesures de vieillissement NATUREL des DVD (études du LNE)

Etudes du LNE (Laboratoire National d'Essais) pour la DAF (Direction des archives de France) et pour l'INA (Institut national audiovisuel).

- Etude du vieillissement de 70 DVD du commerce gravés en 2000. Beaucoup sont endommagés en 2004, et **15% sont illisibles** (donc impossible à recopier sur un support neuf)
- Etude de 2006 de DVD des archives départementales: beaucoup de supports sont devenus illisibles, et donc irrécupérables
- Etude d'une collection privée de 100 DVD achetés en gravés au cours des années entre 1994 et 2004 (celle de F. Laloë)

J. Perdereau, Y. Saunders et J.M. Lambert, LNE (laboratoire national d'essais)

Conclusions de ces études

- La qualité n'est pas homogène, même dans un modèle donné d'une marque donnée; elle dépend des lots de fabrication
- Certains disques sont déjà détériorés en quelques années, présentant des taux d'erreur élevés. D'autres, plus rares, n'ont que peu varié en dix ans.
- Dans le pire des cas (rare), un CD-R peut être devenu illisible en un an, ce qui rend sa recopie impossible (perte totale des données)
- Les disques « archivage » ou « gold » ne sont pas nécessairement les meilleurs; parfois même, ils sont moins bons.
- Des études supplémentaires sont nécessaires; dans l'état actuel du marché et des connaissances, **il n'est pas possible d'émettre une recommandation qui garantisse une conservation de quelques décennies**, autre que la recopie régulière et perpétuelle des données.

J. Perdereau, Y. Saunders et J.M. Lambert, LNE (laboratoire national d'essais)

Mesures de vieillissement ARTIFICIEL; la « théorie » officielle, les normes

On suppose que le processus de vieillissement résulte d'un seul processus physico-chimique (inconnu), et qui possède une énergie d'activation bien définie.

lois d'Arrhenius et d'Eyring

$$t = AT^a e^{\Delta H/kT} e^{(B+C/T) \times RH}$$

avec

t	Durée de vie
A	Constante de temps
T^a	Facteur pré-exponentiel de température
ΔH	Energie d'activation par molécule
k	Constante de Boltzmann ($1,3807 \times 10^{-23}$ J/molécule degré K)
T	Température (Kelvin)
B, C	Constantes RH
RH	Humidité relative

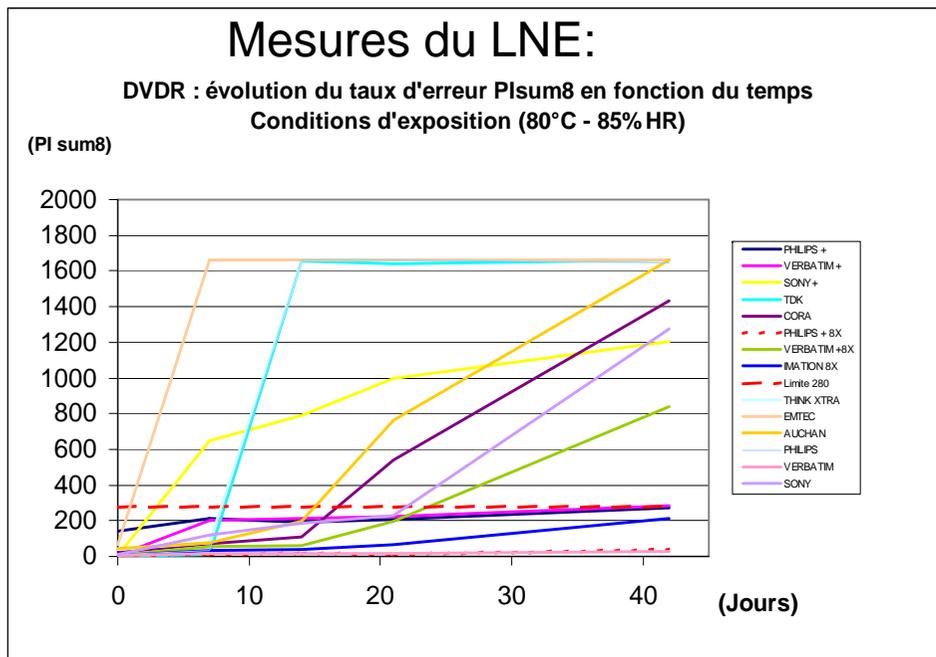
Vieillissement en étuve, 80 C, 85% d'humidité; norme ISO 10955

Mais...

- Cette méthode suppose que le processus de vieillissement est unique, avec une énergie d'activation unique et bien déterminée. Or, on sait bien qu'en pratique
 - les processus sont multiples
 - on a généralement des bandes d'énergie, et pas des énergies caractérisée par une seule valeur
- Elle conduit à des extrapolations avec de grands facteurs (exponentiels); une petite erreur sur l'énergie d'activation conduit à des erreurs d'un facteur 1000 sur le résultat final. Bien sûr, chacun peut utiliser cette flexibilité à son gré pour influencer les résultats dans la direction qu'il désire.
- Elle demanderait de faire des mesures à plusieurs températures pour déterminer cette énergie, ce qui en pratique est rarement fait.. ce qui permet de manipuler à son gré les exponentielles.
- L'extrapolation repose sur un modèle qui prévoit que l'évolution **temporelle** du taux d'erreurs est exponentiel. Est-ce le cas en pratique?

Qu'en est-il vraiment ?

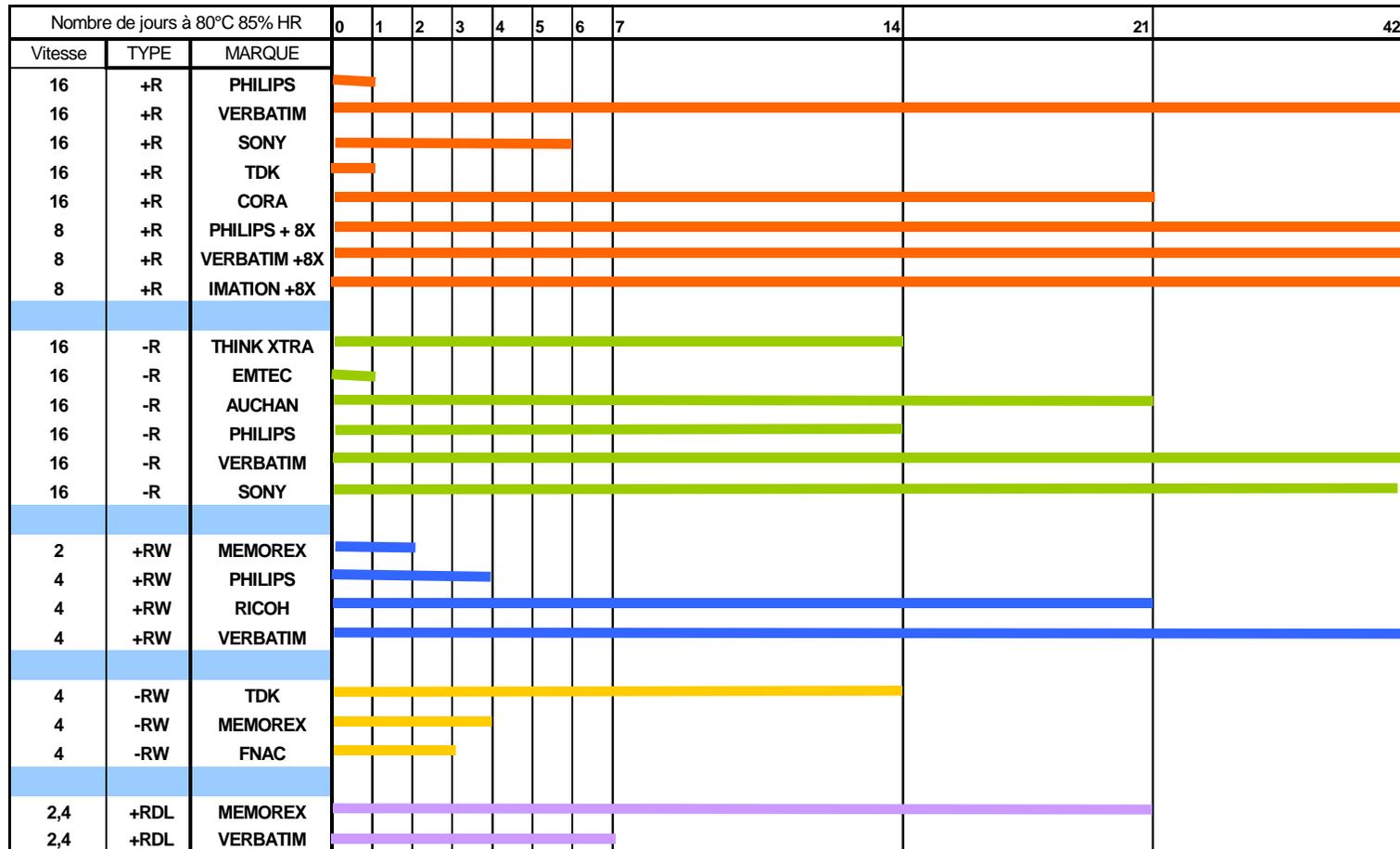
Les courbes temporelles observées n'ont rien d'exponentielles!
L'extrapolation vers les conditions d'utilisation normales, qui mettrait en jeu des exponentielles de l'ordre de 100 ou 1000, n'est donc pas raisonnable.



Conclusion: on ne sait pas relier quantitativement vieillissement artificiel à vieillissement naturel

On peut au mieux espérer une relation qualitative, à savoir que les disques qui résistent le mieux sous conditions artificielles sont également ceux qui résisteront le mieux en vieillissement naturel.

Exemples de mesures de vieillissement artificiel



Documents fournis par J. Perdereau (LNE)

Qu'en ressort-il ?

- De façon générale, il est très difficile de tirer des règles des observations.
- Avec le même modèle de la même marque acheté à un an d'intervalle, on peut obtenir d'abord de très bons résultats, puis très mauvais. Il est donc impossible de recommander telle ou telle marque, ou de se fier au prix des disques.
- Les disques labellisés « gold » ou « archivage » (plus chers) sont parfois meilleurs, parfois pires que les disques ordinaires
- Visiblement les constructeurs ne maîtrisent pas les chaînes de production, qui ne sont pas toujours les mêmes. Ils ne maîtrisent pas non plus les colorants qui sont utilisés dans les disques, qu'ils achètent sans pouvoir en connaître la composition exacte, tenue secrète.
- On a l'impression que la **complexité des disques enregistrables actuels est trop grande, et qu'il est pratiquement impossible de prédire la longévité de tels édifices.**

Etudes à l'étranger

- On trouve un certain nombre d'études, mais sporadiques. Le NIST avait une activité dans ce domaine, mais l'a abandonnée il y a quelques années
- Les constructeurs font des études de contrôle de qualité, mais ne les rendent pas publiques. **Le marché pousse uniquement vers la course à la capacité et à la rapidité d'écriture.**
- Un consortium de constructeurs japonais, « Bifrostec », a été créé. Il dispose d'un laboratoire bien équipé et semble prêt à faire des mesures systématiques.

Le but est de créer un « logo » de garantie de longévité pour des DONE, pour des périodes de 30 ans ou 50 ans.

Mais est-ce vraiment possible avec la conception actuelle des DONE, bien trop compliquée et difficiles à contrôler?

IV. La situation actuelle

Situation générale des DONE

- La consommation mondiale des disques optiques numériques (DON), tous formats confondus, est énorme: plus de 10 milliards. Mais la part de marché des DON diminue, devant les progrès spectaculaires des autres supports: disques durs, mémoires flash, et d'autres qui devraient apparaître bientôt.
- La course à la capacité est perdue; les projets sur les DON parlent de 100 Go, alors que la capacité des disques durs du commerce atteint 2 TO (pour moins de 200 Euros!). Les prix par TO des DONE ne sont pas plus faibles, les débits pas extraordinaires non plus .
- En revanche, en ce qui concerne la préservation à long terme, c'est le principe du DONE qui semble le plus prometteur: séparation des données et du lecteur, simplicité conceptuelle, absence de contact mécanique, etc..
- Il est possible qu'à long terme (dix ans ou plus), la seule chance de survie des DON(E) soit la mise sur le marché de produits qu'on puisse absolument garantir sur le très long terme.
- Le **disque en verre**, le **Millenniata** (garanti un millénaire!),
- Et le **Data Tresor Disc**

Des enjeux économiques significatifs (1)

1. **Stratégie passive**: seuls les DONE semblent envisageables à relativement court terme. Deux marchés en parallèle:
 - Le stockage à court terme
 - L'archivage

Le premier occulte le second. Il est cependant menacé, à cause des progrès fantastiques des autres techniques (disques durs, flash). Le second marché est passé sous silence.

Le déni actuel du problème de vieillissement risque d'entraîner une perte de confiance.

Des enjeux économiques significatifs (2)

1. **Stratégie active**: migrations et tests constants. C'est la solution professionnelle majoritaire.

Peut-on la généraliser aux besoins de l'ensemble de la population nationale?
C'est un problème de coût.

On peut estimer les besoins d'une famille à 1 To environ (à cause de la video) .
En extrapolant à partir des données du CNES ou de l'INA (qui archivent entre 0.5 et 2 Po), on arrive à des prix compris entre **300 et 3000 Euros par To et par an**.

A l'échelle du pays, la dépense se situerait entre 3 et 30 milliards d'Euros avec les technologies actuelles.

Comment faire mieux?

Techniquement, le problème ne semble pas insurmontable, pourvu qu'on s'y attelle. Il existe de multiples procédés permettant de créer de petites inscriptions dans des conditions bien mieux contrôlées que dans les DONE actuels. Diverses initiatives ont été prises, parmi lesquelles une qui semble particulièrement prometteurs:

Le disque de verre trempé ou les informations sont inscrites par lithogravure. Très stable (mesures LNE). C'est un disque de ce type qui a été envoyé dans une navette spatiale.

Il faudrait investir dans le développement de nouvelles technologies pour mettre sur le marché de nouveaux produits très stables à un prix raisonnable.

Que disent les mesures des disques gravés dans le verre?

Disque en verre trempé



Mesures au LNE sur les disques gravés dans le verre



Descriptif de l'étude

- **Conditions climatiques particulièrement sévères:**
 - **Température : 90°C**
 - **Humidité relative : 85%**
- **Durée d'étuvage:**
 - **Par période de 125 heures**
 - **Total actuel : 500 heures**



Conclusion

- Nous sommes à la période de basculement où chacun va se rendre compte de l'acuité du problème à cause des pertes de données.
- Comme souvent, il faudra une catastrophe pour que le problème soit pris à bras le corps. Le problème n'attire pas l'attention du fait de son côté négatif, et des effets d'ombre des réseaux, du « cloud computing », etc.
- Le pays prévoit de dépenser des centaines de millions d'Euros pour numériser le patrimoine artistique et culturel, mais personne ne sait comment ce patrimoine sera transmis dans un siècle...
- En tant que supports courants de sauvegarde ou d'archivage, les DVD ont peut-être perdu la partie commerciale devant les disques durs, du fait des progrès énormes de ces derniers.
- Il semble qu'ils restent les meilleurs candidats pour la mise au point d'un réel support d'archivage numérique à très long terme. La gravure dans un matériau stable, verre par exemple, semble la solution la plus prometteuse.

Recommandations du rapport PSN

Un message d'alerte au grand public, aux établissements et à l'administration, doit être lancé. Le problème n'est pas spécifique à la France, mais mondial. Une action au niveau européen, ou dans le cadre de l'UNESCO, semblerait souhaitable, afin de faire prendre conscience de l'urgence d'une politique concertée dans ce domaine. A l'échelon national et européen, nous proposons quatre recommandations:

- 1. Débloquer les études sur le sujet.**
- 2. Éviter la perte des compétences dans le privé et le public**
- 3. Favoriser l'innovation et l'apparition d'une offre industrielle de qualité**
- 4. Élaborer une véritable politique d'archivage numérique**

Université de Montréal :

« L'information numérique est devenue essentielle à presque tout ce que les Canadiens font dans leur vie quotidienne..... Une partie de notre patrimoine numérique est déjà disparue et une grande partie est en péril »

Bibliothèque et Archives du Canada:

« Au cours des prochaines années, tous les Canadiens devraient vraisemblablement être touchés par les problèmes de conservation de l'information. La fragilité des données numériques et la rapide obsolescence de leurs supports menacent les précieuses collections personnelles et familiales d'œuvres musicales, de DVD et de photographies.

Pourtant, peu de Canadiens mesurent l'ampleur du problème. Aussi est-il important de sensibiliser les créateurs, les représentants des gouvernements, les organes de financement et le grand public. Il faut notamment attirer leur attention sur les risques de perte d'information numérique et ses conséquences, de même que préciser les stratégies et les mesures à prendre pour éviter une telle perte. Nous aurons besoin de leaders dans ce domaine. Il faudra également fournir des exemples percutants de perte ou de risques de perte d'information ».

Il suffit remplacer « les Canadiens » par « les Français » et puis par « les citoyens du monde » pour conclure....

Nanoscale Reversible Mass Transport for Archival Memory

G. E. Begtrup, W. Gannett, T. D. Yuzvinsky, V. H. Crespi, and A. Zettl

Nano Lett., 2009, 9 (5), 1835-1838 • Publication Date (Web): 28 April 2009

Downloaded from <http://pubs.acs.org> on May 15, 2009

