DATA FORMAT

FILES & DATA DESCRIPTION FORMAT
FOR SPECTRA - MATRICE - EVENT
MEMORY - DISK - TAPE

CRN Strasbourg **G. ZEHNACKER**

1 Introduction

A partir des différents articles cités en référence , le présent papier propose un format de données en mémoire , disque et bande valables pour des informations de type variable . Nous nous attacherons à préciser les idées retenues et à les concrétiser par un format de données aussi proche que possible en mémoire et disque et répondant à la norme ANSI pour la bande . Il est important de regrouper dans un même document tous les choix retenus à la fois pour les spectres et les évènements bruts . Le présent document ne discutera pas de la structure des évènements bruts mais de leur format de stockage sur bande et sur disque .

2 Quelques idées directrices

2.1 Types d'Informations

Les informations manipulées en Physique Nucleaire sont de plusieurs types :

— spectre mono , bi ou multiparamétrique (matrice)
— événements bruts
— data quelconque (positions de pic , emplacement de zônes marquées , valeur).

Il est important d' accompagner les valeurs numériques (par ex : canaux ou counts pour les spectres) par des informations descriptives (appelés suivant les cas : label , descripteur de spectres , attributs , bloc header , titre , commentaires etc) . Dans la suite de l'article nous retiendrons le terme label . La taille du label devrait être identique pour tous les types d'informations . Bien entendu la taille des données numériques pourra être variable .

2.2 Medias

Les informations seront mémorisées sur bande magnétique , cartouche Exabyte , disque et en mémoire centrale .

2.2.1 Les spectres

Les spectres sont soit le résultat de l'acquisition , soit le résultat de traitement . Dans les 2 cas au moment de la création du spectre (les canaux) , le label de spectre doit être aussi créé. Celui ci suivra le spectre toute sa vie . Nous créerons donc en mémoire une zône regroupant tous les labels et sur disque un fichier de labels . Dans les deux cas il est possible d'accéder rapidement aux canaux ; dans un cas par un pointeur , dans l'autre par un nom de fichier . Les labels et les données ne seront donc pas nécessairement contigus . Cela simplifiera aussi le "hardware" du DMI prévu pour former directement des histogrammes .

En mémoire , la place sera demandée au fur et à mesure des besoins . Le système fournissant un pointeur sur la zône fournie . Par la même les zônes obtenues ne seront plus jointives . Une zône ainsi obtenue pourra contenir un ou plusieurs spectres ; les labels étant aussi mis à jour en cascade .

La question se pose sur la possibibilité de stocker les canaux d'un spectre dans un fichier ou de stocker les canaux de tous les spectres dans un fichier .

La première solution permet si on y stocke aussi le label , une recherche rapide si on est certain que le spectre existe tout en permettant une connaissance rapide sur tous les spectres grâce au fichier des labels . Il est évident que tous ces spectres devront se trouver sous un répertoire réservé aux spectres . La deuxième solution éventuellement ne nécessite pas un répertoire spécifique pour les spectres , mais de toute façon l'accès est moins rapide et les fichiers riquent de devenir trop gros .

Le choix à faire est donc entre les 2 solutions ou une solution mixte (1 fichier pour les petits spectres 4 à 8 k et un fichier pour les spectres multiparamétriques (de quelques kilo ou mega octet) . De toute façon , il est préférable d'avoir un fichier de données pour un spectre multiparamétrique. Notons aussi qu'il n'y a pas de différence de structure des données pour les spectres mono , bi ou multiparamétriques .

2.2.2 Les événements bruts

Les événements créés par l'acquisition seront stockés soit sur Exabyte , soit sur bande par l'intermédiaire d'un buffer en mémoire centrale . Pour y accéder nous pourrions utiliser le pointeur associé au label d'événements créés au moment du démarrage de l'acquisition . Ce label s'intégrera dans la table des labels au même titre que les spectres . Le stockage sur bande ou exabyte s'intégrera dans la norme ANSI qui sera définie plus loin . Le dépouillement en géneral se fait directement à partir de la cassette exabyte . Mais il peut arriver que pour d'autres expériences qu' EUROGAM , nous soyons amenés à stocker des évènements bruts sur disque . Dans ce cas nous ferons une recopie de la bande magnétique sur disque , en conservant la structure de la bande . Un fichier comportera un ou des blocs d'évenement bruts . Une cassette Exabyte pourra contenir plusieurs 'run'.

2.2.3 Les autres types de données

D'autres informations peuvent être nécessaires aux cours des expériences . Celles ci seront caractérisées par un label et des valeurs numériques . Leur structure sera connue dans le label . Les valeurs numériques seront stockées dans un bloc identifié différement des spectres et des événemnts bruts .

2.3 Structure

Les informations comporteront toujours un label et des valeurs numériques quelque soit le type d'informations .

2.3.1 label

Pour définir le label nous nous sommes inspirés de la norme bande IN2P3 (5) . Nous résumons schématiquement quelques unes des zônes du label . Les informations sont stockés en général en ASCII dans des blocs de 80 octets .

— nom du run , du propriétaire
— nom et numéro du spectre
— date de création
— type (spectre,matrice,events,data quelconque)
— taille et type de la donnée élementaire (en bits , entier ou reel)
— taille des dimensions
— zone d'usage libre (commentaires,valeurs d'etalonnage)

2.3.2 données

Sur bande ces données sont regroupéffs en bloc > à 512 octets et < à 16384 octets. Sur disque comme pour tout fichier UNIX il n'y a pas de structure de blocs. Le fichier est une suite d'octets accessibles directement .

3 Compléments pour la réalisation

3.1 mémoire , ram disque , disque

La table de labels est automatiquement implantée en mémoire après chaque login . Lors de chaque création de spectres , de buffer d'événements bruts ou de données numériques quelconques la table de label est mise à jour et une demande de place mémoire stockera son emplacement (valeur d'un pointeur) dans la table de labels. Chaque destruction de spectre provoquera la mise à jour de la table des labels et la restitution de place . Bien entendu les zônes de données numériques ne sont pas contigues .

Sur disque un stockage d'informations créera la table des labels si elle n'existe pas ou la mettra à jour si elle existe déjà . Puis on créera le fichier des donnés numériques. Il est souhaité de créer si nécessaire un répertoire spécifique informations (spectres , événements ou données quelconques). Une commande d'écriture spécifique sur disque est à créer . La commande cp n'est donc pas directement utilisable .

La table des labels sera accessible indépendamment des valeurs des canaux. Une liste de spectres sera faite à partir de cette table (ou fichier) labels sans avoir besoin de scruter tous les spectres .Nous proposons le format specificié dans l'article de J.ZEN

3.2 bande

Le format retenu est conforme à la norme ANSI qui a été adapté pour les besoins de la Physique Nucléaire .

— chaque bande contient des blocs label de volume (VOL1 , UVL1 ,UVL2)
— un fichier de données est précédé d'un groupe de bloc (HDR1 , HDR2 , UHL1 , UHL2) séparé par un tape marque et suivi d'une tape marque puis d'un groupe de blocs (EOF1 , EOF2 , UTL1 , UTL2)
— les blocs autres que de données ont 80 octets
— chacune de ces entités se termine par une tape marque
— chaque bande se termine par une double tape marque

Tout en étant conforme à la norme , nous proposons quelques choix pour simplifier l'accés aux informations. Nous allons les résumer:

— les fichiers devraient être de type monovolume
— les blocs de données seront de taille fixe multiple de 4 octets > = à 512 octets
— les champs commencent tous par un espace
— les valeurs numériques sont cadrées à droite et complétées à gauche par des zéros ou des espaces
— les zônes non utilisées devront être mises à blanc

5 Rappel de la norme bande

— Label de volume : VOL1 , UVL1 ,UVL2
— Label début de fichier: HDR1 , HDR2 , UHL1 , UHL2
— Fichier de données :COMMENT , SCALER , PARAM , EVENTH ,EVENTD , SPECTH , SPECTD
— Label fin de fichier: EOF1 , EOF2, UTL1,UTL2

6 Bibliographie

— Introducing the SDB spectrum access system sept 1985 , NSF , Daresbury
— Accessing spectra , Feb 1988 , NSF , Daresbury
— Norme spectres sur disquettes , mai 1988 , IPN / CNSM
— Système de gestion des fichiers de spectres , Nov 1988 , CRN Strasbourg

— Sur la définition d'un standard bande jan 87 CNSM Orsay B.MERLANT
— Spectrum database structure sept 1990 CRN Strasbourg J.ZEN
— Eurogam on line spectrum access jul 1990 NSF Daresbury D.BRIGHTLY
— Event by Event S.KOSIONIDES