

В. И. ЖУК

(Москва)

## О НЕКОТОРЫХ МЕТОДАХ НАКОПЛЕНИЯ И ВНЕШНЕЙ СОРТИРОВКИ СПЕКТРОМЕТРИЧЕСКИХ ДАННЫХ НА МАЛОЙ ЭВМ С МАГНИТНЫМИ ДИСКАМИ

**Введение и постановка задачи.** Задачу регистрации спектрометрических данных в многомерных ядерно-физических экспериментах в общем виде можно формально описать следующим образом. На вход системы регистрации поступает поток исходных признаков  $X$ , каждый из которых характеризует некоторое событие и может быть рассмотрен как многомерный вектор:

$$X = (x_1, \dots, x_n),$$

где  $x_1, \dots, x_n$  — элементарные исходные признаки. Например, признак  $X$  может состоять из двух элементарных признаков ( $n=2$ ): кода времени пролета частицы и кода состояния детекторов, в котором каждый разряд фиксирует состояние определенного детектора.

Для общности рассмотрения можно представить, что система сначала преобразует первичные признаки  $X$  в некоторые вторичные признаки  $Y_\psi$ , а затем сортирует последние, как будет показано ниже. Из вектора  $X$  может быть образован вектор, состоящий из  $j$  признаков  $Y_\psi$ :

$$(Y_1, \dots, Y_j) = \varphi(X), \quad (1)$$

где  $j$  — размерность вектора  $Y$  — зависит от  $X$ .

Функция (1) определяется задачами физического эксперимента. Среднее значение  $j$  для потока векторов  $X$  обозначим через  $r$  и назовем коэффициентом расщепления. В простейшем случае  $Y \equiv X$  и  $j \equiv r \equiv 1$ .

Сортировка признаков  $Y_\psi$  состоит в формировании адресно-упорядоченного массива, в котором значение любого элемента равно количеству признаков  $Y_\psi$ , имеющих одно определенное значение, которое взаимно-однозначно соответствует адресу этого элемента в массиве. Такой массив называют «спектром»; значения признаков  $Y$  сопоставляют «каналам спектра»; значение элемента этого массива называют «отсчетом», а сортировку — «поканальной» [1—4]. Заметим, что индекс  $j$  в (1), как правило, означает номер спектра. В ряде случаев оказывается удобным рассматривать регистрацию спектров как процесс сортировки признаков  $X$  по сложным критериям, включающим в себя преобразование (1) и критерии поканальной сортировки признаков  $Y_\psi$ . Поэтому в общем случае будем говорить о сортировке некоторых признаков  $U$ , которые при необходимости могут быть замещены признаками  $Y_\psi$  или  $X$ . По аналогии с другими видами сортировки [5] поканальную сортировку будем называть внутренней, если сортируемый и отсортированные массивы размещены в оперативной памяти (ОП), и внешней, если по крайней мере отсортированный массив размещен во внешней памяти.

Развитие многомерных спектрометрических измерений в ядерной физике характеризуется увеличением суммарного количества каналов одновременно регистрируемых спектров от десятков тысяч до миллионов. Это количество часто превышает объем поля данных (ОП) регистратора. В этих случаях используют накопление неупорядоченного массива спектрометрических данных в порядке их поступления (в дальней-

шем неинкрементное накопление)\* и последующую поканальную их сортировку [1—4, 6]. В работах [1, 2, 6, 7] описаны специализированные цифровые средства для накопления и внешней поканальной сортировки спектрометрических данных. Имеются сообщения о системах накопления и внешней поканальной сортировки спектрометрических данных непосредственно на универсальных ЭВМ с использованием магнитных лент [8], барабана [9] и дисков [10]. С точки зрения повышения быстродействия наибольший интерес из них представляют системы с магнитными барабанами и дисками. Однако в [9, 10] рассмотрены только случаи, при котором во внешней памяти размещен лишь отсортированный массив (спектр), а сортируемые массивы размещаются только в ОП. При этом в [9] рассмотрена система, в которой полный объем спектра не превышает объема поля данных внешней памяти, а в [10] описана ассоциативная система, предназначенная для использования в ситуации, когда количество каналов спектра превышает объем поля данных внешней памяти. Кроме того, системы, рассмотренные в [9, 10], соответствуют случаю  $Y \equiv X$ ,  $j \equiv r \equiv 1$ . Между тем большой практический интерес представляет также более сложный общий случай, описываемый функцией (1) при  $r \geq 1$ .

Важнейшей задачей оптимальной организации неинкрементного накопления и внешней поканальной сортировки спектрометрических данных является минимизация суммарного времени всех вынужденных перерывов в измерениях, т. е. времени, затрачиваемого на сортировку и приостановки в процессе накопления [1]. В данной работе рассматриваются некоторые методы решения этой задачи (преимущественно для случая, когда измерения проводятся в соответствии с (1) при  $r \geq 1$ ) на основе малой ЭВМ, имеющей два канала прямого доступа (КПД) и накопитель на магнитных дисках (НМД) в качестве внешней памяти, которую будем называть дисковой (ДП). Ограничим рассматриваемый случай условием, что полный объем спектра не превышает объема поля данных по ДП, но при этом в отличие от [9] в ДП, помимо отсортированного массива, может быть также размещен (если это окажется целесообразным) и сортируемый массив.

Работа построена следующим образом. В разделе 1 рассматриваются основные принципы регистрации спектров с предварительным внешним накоплением сортируемого массива и обосновывается выбор алгоритма внешней сортировки. В разделе 2 описывается метод ускорения выбранного алгоритма с помощью предварительной группировки признаков по зонам в процессе их внешнего накопления. В разделе 3 рассматриваются вопросы организации внешнего накопления с группировкой признаков по зонам и совмещением во времени процессов приема исходных признаков, их обработки и передачи в ДП. В разделе 4 дается сравнительная оценка суммарных потерь времени на перерывы в измерениях при регистрации спектров с предварительным внешним накоплением, группировкой признаков по зонам и последующей внешней сортировкой. Все численные примеры даны для случаев использования ЭВМ фирмы «Hewlett Packard» на базе процессора HP2100A и НМД типа HP7900A с подвижными головками.

**1. Прямой и инверсный алгоритмы внешней сортировки.** Рассмотрим сначала внешнюю поканальную сортировку накопленного в ДП неупорядоченного массива  $D$  признаков  $U$ , в качестве которых в зависимости от принятого способа регистрации могут быть использованы признаки  $X$  или  $Y$ . Сформированный в результате этой сортировки адресно-упорядоченный массив  $S$  отсчетов  $q$  размещается в ДП. Разобьем массивы  $D$  и  $S$  на элементарные массивы  $d$  и  $s$  соответственно.

\* Этот вид накопления часто называют «накоплением в неинтегрирующем режиме» [4].

Пусть при этом соблюдается соотношение.

$$v = \omega_v m + \omega_q n, \quad (2)$$

где  $v$  — объем поля данных ОП в словах\*;  $m$  — объем массива  $d$ , выраженный в количестве элементов  $U$ ;  $n$  — объем массива  $s$ , выраженный в количестве элементов  $q$ ;  $\omega_v$ ,  $\omega_q$  — количества слов, затрачиваемых на кодирование элементов  $U$  и  $q$  соответственно ( $\omega_v$  и  $\omega_q$  не обязательно целые).

Представляются очевидными два основных алгоритма непосредственной внешней поканальной сортировки, которые будем называть прямым и инверсным алгоритмами. В обоих алгоритмах используется внутренняя сортировка массива  $d$  по признакам, соответствующим номеру формируемого элементарного массива  $s$ . При этом в ОП одновременно размещается один массив  $d$  и один массив  $s$ . Сформированные в ОП массивы  $s$  передаются в ДП.

Прямой алгоритм состоит в поочередном формировании массивов  $s$ . Для полного формирования каждого массива  $s$  производится поочередная передача в ОП и внутренняя сортировка всех массивов  $d$ . Аналогичный алгоритм используется в анализаторах с промежуточным запоминанием данных на магнитной ленте [1, 2].

Инверсный алгоритм состоит в том, что каждый элементарный массив  $d$  передается в ОП только один раз; для каждого массива  $d$  в ОП поочередно передаются элементарные массивы  $s$ . После передачи очередного массива  $s$  в ОП производится внутренняя сортировка массива  $d$  по признакам, соответствующим номеру массива  $s$ , после чего массив  $s$  возвращается в прежнее место ДП.

В общем случае для внутренней сортировки признак  $U$  проверяется по некоторому условию отбора, соответствующему номеру  $i$  формируемого элементарного массива  $s_i$ . Если это условие выполняется, то по значению  $U$  определяется адрес элемента в массиве  $s_i$  и увеличивается на единицу отсчет  $q$ , содержащийся по этому адресу. При  $U=X$  сложность условия отбора зависит от функции (1). При  $U=Y$  это условие сводится к тому, что признак  $Y$  должен попадать на интервал, соответствующий номеру  $i$ . Время внутренней сортировки должно быть минимизировано, так как существенно влияет на длительность внешней сортировки.

Существуют оптимальные значения параметров  $m$  и  $n$ , которые можно найти, минимизируя время внешней сортировки при постоянном объеме  $v$ , определенном в (2).

Можно показать, что при  $U=Y$  и оптимальных параметрах  $m$  и  $n$  прямой алгоритм примерно в  $1,5 \div 2$  раза быстрее инверсного вследствие того, что в последнем случае при сортировке каждого массива  $d$  каждый массив  $s$  передается между ОП и ДП дважды.

Суммарные потери времени на внешнее накопление и сортировку массива  $D$  объемом  $M$  можно представить в виде

$$T_{\text{нр}} = T_p + T_{\text{нр}}, \quad (3)$$

где  $T_p$  — время внешней сортировки массива  $D$ ;  $T_{\text{нр}}$  — время накопления массива  $D$  в ДП до начала сортировки;  $\rho$  — номер алгоритма, в частности,  $\rho=1, 2$  — номер, присвоенный прямому и инверсному алгоритмам соответственно.

При больших значениях  $N$  величиной  $T_{\text{нр}}$  в (3) можно пренебречь по сравнению с величиной  $T_p$ . Например, погрешность в определении  $T_{\text{нр}}$  при  $N=64 K$  ( $K=1024$ ),  $M=10^6$ ,  $v=10 K$ ,  $U=X \equiv Y$ ,  $r \equiv 1$  не превышает 3,5% для прямого алгоритма и 2,5% для инверсного.

\* Здесь под словом понимается элемент данных, адресуемый в ОП. Например, в ЭВМ HP2100A слово содержит 16 двоичных разрядов.

Обычно количество  $Q$  сортируемых признаков  $U$  больше, чем максимально возможный объем  $M$ , отводимый в ДП для размещения массива  $D$ . Поэтому поток признаков  $U$ , имеющий объем  $Q$ , необходимо накапливать и сортировать частями, представляющими массивы  $D$ . В этом случае суммарные потери времени с учетом (3) составят

$$T_{\text{пр}}^{(Q)} = \frac{r \cdot Q}{M} \cdot T_{\text{пр}}. \quad (4)$$

При известном способе внешней сортировки [9] в отличие от рассмотренного выше не накапливают массив  $D$  в ДП, а производят внешнюю сортировку каждого элементарного массива  $d$ , полученного в ОП в процессе накопления исходных признаков  $X$ , поступающих от датчика. Алгоритм внешней сортировки этого массива  $d$  является частью рассмотренного выше инверсного алгоритма, и его характеристики помечаются ниже индексом  $\rho=0$ . Для ускорения внешней поканальной сортировки этот массив  $d$  может быть предварительно преобразован, например, упорядочен [9]. Можно показать, что если массивы  $d$  не упорядочены монотонно перед их сортировкой, то при достаточно больших  $N$  и  $M=Q_x$  получим

$$T_{\text{п}0} \approx T_2 \approx T_{\text{п}2} > T_{\text{п}1}.$$

Как будет показано на рис. 1 в разделе 4, величина  $T_{\text{п}0}$ , определенная с учетом предварительного монотонного упорядочения массивов  $d$  для случая, когда  $U=X=Y$ ,  $r=1$ , при неизменных остальных условиях, может быть больше величины  $T_1$ , даже если время упорядочения равно нулю. Заметим, что в случае (1) при  $r>1$  предварительное упорядочение массивов  $d$  не имеет смысла, так как приводит к увеличению  $T_{\text{п}0}$ .

Из условия  $T_{\text{п}1}^{(Q)} < T_{\text{п}0}$  несложно найти минимально допустимое значение  $M$  для прямого алгоритма.

Таким образом, при достаточно больших  $M$  и  $N$  наиболее быстрым из рассмотренных алгоритмов внешней сортировки является прямой алгоритм. Однако время сортировки этим алгоритмом с оптимизированными параметрами  $m$  и  $n$  при больших значениях  $M$  и  $N$  остается большим, особенно при  $r>1$ . Для дальнейшего уменьшения потерь времени при регистрации спектров предлагаются рассматриваемые ниже способы регистрации, использующие прямой алгоритм внешней сортировки с предварительной внешней группировкой признаков по зонам в процессе их накопления.

**2. Ускорение внешней сортировки с помощью группировки признаков по зонам памяти в процессе накопления.** Группировка событий по зонам памяти использовалась для ускорения внешней сортировки [1] и для регистрации наиболее вероятных событий при их ассоциативной сортировке [3]. Однако в этих случаях формировались сгруппированные массивы отсчетов, а не массивы сортируемых признаков. В работе [9] предложена внутренняя группировка признаков после накопления сортируемого массива в ОП перед его внешней поканальной сортировкой с целью ускорения последней. Эта группировка осуществлялась путем внутреннего монотонного упорядочения по шкале ключевых признаков, размещенных в старших разрядах элементов массива [5]; при этом группы размещались в ОП в соответствии с их длиной, а не в фиксированных зонах. Здесь рассматривается применение группировки признаков  $Y$  в процессе их накопления с целью последующего ускорения внешней сортировки. При этом признаки  $Y$  группируются в неупорядоченные массивы по фиксированным зонам памяти; элементы массива, принадлежащие разным группам, могут не содержать в себе ключевой признак группы и могут быть равными.

В этом разделе рассматривается внешняя сортировка, проводимая по окончании группировки признаков в ДП, а в следующем — организация накопления с группировкой признаков.

Поле  $P_D$  данных в ДП, предназначенное для размещения массива  $D$  объемом  $M$ , разбивается на  $z$  зон  $P_D^{(\lambda)}$ , в которых располагаются сгруппированные массивы  $D_\lambda$ , имеющие объемы  $M_\lambda$ . Здесь  $\lambda=1, 2, \dots$ ;  $z$  — номер зоны. Поле  $P_S$  отсчетов в ДП, предназначенное для размещения массива  $S$  объемом  $N$ , содержащего все полученные спектры, разбивается на такое же количество зон  $P_S^{(\lambda)}$ , в которых размещаются массивы-спектры  $S_\lambda$ , имеющие объемы  $N_\lambda$ . Каждой зоне  $P_S^{(\lambda)}$  соответствует зона  $P_D^{(\lambda)}$ , т. е. в зоне  $P_D^{(\lambda)}$  должны содержаться только те сортируемые признаки  $Y$ , которые соответствуют каналам спектра  $S_\lambda$ . Массивы  $D_\lambda$  и  $S_\lambda$  состоят, как и в случае  $z=1$ , из элементарных массивов  $d$  и  $s$  соответственно. Объемы  $m$  и  $n$  массивов  $d$  и  $s$  не зависят от  $\lambda$ .

Для сокращения потерь времени, связанных с повторением циклов накопления и сортировки, желательно, чтобы все зоны  $P_D^{(\lambda)}$  были в максимально возможной степени заполнены признаками. Мы рассмотрим способ разбиения ДП на зоны, который характеризуется тем, что объемы зон  $P_S^{(\lambda)}$  одинаковы, а объемы зон  $P_D^{(\lambda)}$  могут быть разными. Можно показать, что такой способ целесообразнее других, так как в нем легче обеспечивается кратность величин  $N_\lambda$  и  $n$ , необходимая для уменьшения времени сортировки. Кроме того, этот способ удобнее во многих случаях, когда массивы  $S_\lambda$  соответствуют различным физическим спектрам, имеющим одинаковое количество каналов. Однако требования к степени заполнения зон  $P_D^{(\lambda)}$  могут быть снижены и во многих случаях сняты, если сортировать только заполненную часть каждой зоны  $P_D^{(\lambda)}$ , для чего необходимо использовать информацию о границе заполнения этой зоны.

При внешней сортировке каждый массив  $D_\lambda$  сортируется только по признакам, относящимся к соответствующей номеру  $\lambda$  зоне  $P_S^{(\lambda)}$ .

Время внешней сортировки массива  $D$  прямым алгоритмом после группировки этого массива по зонам

$$T_3 \approx M \left[ \frac{\omega_Y}{m_0} \left( t_{g\min} + t_0 \frac{b_1 - 1}{b_1} \right) + \frac{t_0}{m} + (\tau_n + \bar{\tau}_{c\min}) \omega_Y \right] \left[ \frac{N}{nz} \right] + M (\bar{\tau}_{c\max} - \bar{\tau}_{c\min}) \omega_Y + z \left\{ 3(\theta - 1) t_{g\min} + 2 \left[ \left[ \frac{n\omega_q}{m_0} \right] (t_{g\min} + t_0) + \tau_n \omega_q n \right] \right\} \left[ \frac{N}{nz} \right], \quad (5)$$

где  $\frac{b_1}{a_1} = \frac{m\omega_Y}{m_0}$ ;  $\theta = \frac{t_{g\max}}{t_{g\min}}$ ;  $[x] = x$ , если  $x$  — целое;  $[x] = [x] + 1$ , если  $x$  — нецелое;  $[x]$  — целая часть числа  $x$ . Здесь  $t_{g\max}$ ,  $t_{g\min}$  — максимальное и минимальное время перемещения головок ДП;  $t_0$  — среднее время ожидания сектора ДП;  $\tau_n$  — среднее время передачи одного слова между ОП и ДП;  $m_0$  — объем цилиндра ДП (в словах);  $a_1$ ,  $b_1$  — взаимно простые числа;  $\bar{\tau}_{c\max}$ ,  $\bar{\tau}_{c\min}$  — среднее время внутренней сортировки, приходящееся на одно слово массива  $d$  в случаях выполнения ( $\bar{\tau}_{c\max}$ ) и невыполнения ( $\bar{\tau}_{c\min}$ ) условия отбора признака  $Y$  в соответствии с номером формируемого массива  $s$ ; остальные величины определены выше.

Оптимальные параметры  $z$ ,  $n$ ,  $m$  находятся из условий минимизации величин  $T_3$  и  $[N/nz]$  при заданном  $v$  с учетом (2). Как будет показано в разделе 3, величина  $z$  ограничена сверху при заданном  $v$  соотношениями (11), (13), которые также должны быть учтены при оптимизации параметров. При достаточно большом  $v$  оптимальное количество зон удовлетворяет соотношению

$$z = N/n. \quad (6)$$

Время внешней сортировки при группировке признаков по зонам

Количество спектров $j$	Количество каналов в одном спектре	Количество каналов во всех спектрах $N$	Количество зон $z$	Объем $l$ в словах	Объем $v$ поля данных ОП в словах		
					10K	18K	26K
					Время сортировки $T_3$ , с		
1	64K	64K	1	8K	242	216	198
				16K	—	129	116
			4	16K	—	35,34	32,1
			8	8K	36,1	32,84	32,04
37	8K	296K	37	8K	46,39	43,19	42,84
16	64K	1024K	64	16K	—	67	56,69

При выполнении условия (6) отпадает необходимость проверки условий отбора в процессе внутренней сортировки, вследствие чего время внутренней сортировки в (5) может быть сделано минимально возможным. В этом случае  $\tau_{\max} \equiv \tau_{\min} = \text{const}$ .

Если при длине  $l$  машинного слова имеем

$$\log_2 N > l, \quad (7)$$

то для уменьшения величин  $\bar{\tau}_c$  и  $T_3$  необходимо выбрать  $z$  так, чтобы  $\omega_Y = 1$ :

$$z \geq N/2^l. \quad (8)$$

В этом случае каждому признаку  $Y$  будет соответствовать некоторый номер зоны, не входящий в состав этого признака и формируемый в процессе накопления. При таком задании величины  $z$  каждый спектр, например, может состоять из одной зоны или может быть поделен на зоны.

В таблице приведены значения времени  $T_3$ , рассчитанные по формуле (5), в зависимости от объема  $v$  поля данных ОП при  $\omega_Y = \omega_q = 1$ ,  $M = 10^6$ .

3. Группировка признаков по зонам в процессе их накопления. Рассмотрим организацию предварительной группировки признаков по зонам  $P_D^{(k)}$  на диске. В процессе накопления поле данных ОП, имеющее объем  $v$ , разделено на поле  $H$ , в которое по каналу прямого доступа (КПД) поступают исходные признаки  $X$  от их датчика, и поле  $G$ , разделенное, в свою очередь, на  $W$  зон  $G_\gamma$  ( $\gamma = 1, \dots, W$ ), имеющих объемы  $v_\gamma$ :

$$v = v_H + v_G, \quad (9)$$

$$v_G = W \bar{v}_\gamma, \quad (10)$$

$$W = z + \alpha, \quad (11)$$

где  $v_H$ ,  $v_G$  — объемы полей  $H$  и  $G$  в словах;  $\bar{v}_\gamma$  — средний объем зоны  $G_\gamma$  в словах;  $z$  — количество зон при сортировке;  $\alpha$  — дополнительное целое количество зон,  $\alpha \geq 0$ .

Примем ограничения:

$$v_H \geq \omega_x, \quad v_\gamma \geq v_c, \quad (12)$$

где  $\omega_x$  — количество слов в признаке  $X$ ,  $v_c$  — объем сектора диска в словах, причем величина  $v_7$  кратна величине  $v_c$ .

Из (9), (10), (12) получим

$$W \leq \frac{v - \omega_x}{v_c}. \quad (13)$$

Одновременно с приемом признаков  $X$  через КПД в поле  $H$  программа просматривает поле  $H$ , преобразует принятые признаки  $X$  в признаки  $Y$ , согласно (1), и распределяет (группирует) признаки  $Y$  по зонам  $G_7$ . Признаки  $Y$  заполняют соответствующие им зоны  $G_7$ , в результате чего в каждой зоне может накапливаться неупорядоченный массив признаков  $Y$ .

Отметим два способа организации заполнения и просмотра поля  $H$ . При первом из них поле  $H$  разделено на две части, каждая из которых поочередно заполняется через КПД. Пока заполняется одна часть поля  $H$ , другая его часть просматривается программой, осуществляющей преобразование (1) и группировку признаков  $Y$ . При втором способе поле  $H$  не разделено на части. Перед каждым включением КПД поле  $H$  заполняется нулями. Одновременно с заполнением поля  $H$  через КПД программа просматривает поле  $H$  в направлении его заполнения, следит за положением текущей границы заполнения поля  $H$ , сравнивая содержимое ячеек поля  $H$  с нулем, и поочередно обрабатывает признаки  $X$  в порядке их поступления. Новое заполнение поля  $H$  начинается после обработки массива, накопленного в поле  $H$  в результате предыдущего заполнения этого поля.

Недостаток 1-го способа состоит в увеличении потерь времени на включение КПД, а недостаток 2-го способа — в затратах времени на очистку поля  $H$  перед каждым его заполнением. Некоторое преимущество 2-го способа состоит в том, что он позволяет организовать контроль интервалов времени между событиями.

Рассмотрим теперь способы организации заполнения зон  $G_7$  в ОП. В зависимости от значения  $\alpha$  удобно выделить 5 таких способов:

1)  $\alpha=0$ ; 2)  $\alpha=1$ ; 3)  $z > \alpha > 1$ ; 4)  $\alpha=z$ ; 5)  $\alpha > z$ .

Выбор  $\alpha$  зависит от величин  $v$  и  $z$  [см. (9) — (13)]. При  $\alpha=0$  имеем  $\gamma=\lambda=1, \dots, z$ . Зоны  $G_7$  неделимы на части. Как только заполнится любая зона  $G_\lambda$ , ее содержимое передается в соответствующую зону  $P_D^{(\lambda)}$  диска независимо от степени заполнения других зон  $G_\lambda$ . Передача массивов из поля  $G$  на диск и прием признаков  $X$  в поле  $H$  осуществляется по двум КПД соответственно и поэтому могут быть совмещены во времени. Работа программы формирования массивов в зонах  $G_\lambda$  приостанавливается на время передачи массива из зоны  $G_\lambda$  на диск. Таким образом, прием признаков  $X$  в поле  $H$  может быть совмещен либо с работой программы формирования массивов в зонах  $G_\lambda$ , либо с передачей массивов из зон  $G_\lambda$  на диск.

При остальных способах ( $\alpha \geq 1$ ) наличие запасных зон дает возможность совместить во времени три процесса: прием признаков  $X$  в поле  $H$ , передачу массивов из зон  $G_7$  на диск и формирование массивов в зонах  $G_7$ . Во всех способах при  $\alpha \geq 1$  возможно возникновение очередей для передачи массивов из зон  $G_7$  на диск. При этом время остановок формирования массивов в зонах  $G_7$  уменьшается с увеличением  $\alpha$ . Весьма удобным представляется третий способ ( $\alpha=z$ ), при котором  $\gamma=1, \dots, 2z$  и каждой зоне  $P_D^{(\lambda)}$  соответствует зона  $G_\lambda$ , разделенная на две зоны  $G_7$ , которые будем называть подзонами. Массив из каждой заполненной подзоны передается в соответствующую зону  $P_D^{(\lambda)}$  диска независимо от степени заполнения других подзон. Подзоны, относящиеся к одной зоне  $G_\lambda$ , заполняются поочередно. Формирование

массива в одной из этих подзон может быть совмещено с передачей на диск массива из другой подзоны.

Необходимые условия совмещения во времени процессов приема признаков  $X$  в поле  $H$ , формирования массивов в зонах  $G_T$  и передачи их на диск будут:

для 1-го способа организации заполнения поля  $G$

$$T_D + T_X - T_H \leq 0, \quad (14)$$

для остальных способов

$$\max \{T_D, T_X\} - T_H \leq 0, \quad (15)$$

где  $T_D$  — время передачи на диск массива  $D$ ;  $T_X$  — время формирования и записи в поле  $G$  всех признаков  $Y \in D$  из признаков  $X$ ;  $T_H$  — время накопления признаков  $X$ , из которых сформированы все признаки  $Y \in D$ , т. е. «чистое» время измерения;  $\max \{T_D, T_X\}$  — максимальная из величин  $T_D$  и  $T_X$ . Здесь

$$T_D = \omega_Y (t_{g.c.p.} + t_0) \sum \frac{Mr}{v_Y} + M\omega_Y \tau_H, \quad (16)$$

$$T_X = M\tau_X/r, \quad T_H = M/Ir, \quad (17)$$

где  $t_{g.c.p.}$  — среднее время перемещения головок диска;  $r$  — коэффициент расщепления признака  $X$ ;  $\tau_X$  — среднее время преобразования признака  $X$  в  $r$  признаков  $Y$  и записи этих признаков в поле  $G$ ;  $I$  — интенсивность потока признаков  $X$  (признак/с); остальные величины определены выше.

Для удобства реализации и улучшения совмещения процессов при накоплении примем, что при любых значениях  $\alpha$  объемы  $v_T$  зон  $G_T$  одинаковы. Это ограничение представляется очевидным при  $\alpha \geq 1$  в тех случаях, когда запасная зона может замещать любую другую зону. Для случаев, когда все зоны  $G_T$  жестко соответствуют определенным зонам  $P_T$  на диске (при  $\alpha=0$ ,  $\alpha=z$ ), альтернативный вариант состоит в выборе объемов  $v_T$  в таком соответствии с распределением признаков по зонам, чтобы выровнять частоты обращения к диску из разных зон  $G_T$ . По сравнению с вариантом, когда объемы  $v_T$  равны, этот выбор приводит к увеличению вероятности возникновения очередей запросов обращения к диску.

Заметим, что при выполнении условий (14), (15) также возможны некоторые приостановки измерений, возникающие с некоторой вероятностью в ситуациях, при которых заполнение поля  $H$  заканчивается в моменты, когда передача массива из зоны  $G_T$  на диск или обработка массива в поле  $H$  не закончены. Можно показать, что вероятность возникновения этих ситуаций также зависит от статистического распределения признаков  $v$  по зонам и обратно пропорциональна объему  $v_H$  поля  $H$ .

В качестве необходимого условия минимальности времени остановок при накоплении, не зависящего от статистического распределения признаков  $Y$  по зонам и общего для всех рассмотренных способов заполнения поля  $G$ , можно выбрать минимум функции  $T_{он}$ :

$$T_{он} = T_D + T_{вкл} + T_X, \quad (18)$$

где  $T_{вкл}$  — суммарное время включения КПД, через который поступают признаки  $X$  при накоплении массива  $D$ ; остальные величины определены выше;

$$T_{вкл} = \frac{M\omega_X}{r v_H} t_{вкл} \quad (19)$$



( $t_{\text{вкл}}$  — длительность одного включения КПД, через который поступают признаки  $X$ ).

Величина  $T_{\text{нз}}$  суммарных потерь времени при накоплении массива  $D$  с внешней группировкой признаков по зонам в общем случае принимает значение на интервале

$$0 \leq T_{\text{нз}} \leq T_{\text{он}}, \quad (20)$$

где функция  $T_{\text{он}}$  определена в (18).

Если процессы приема признаков  $X$ , их обработки и передачи массивов на диск полностью не совмещены, то значение  $T_{\text{нз}}$  максимально. Если эти процессы совмещены полностью, то можно принять, что  $T_{\text{нз}} \approx 0$ .

При равенстве объемов зон  $G_T$  величину  $\bar{v}_T$  можно найти из условия минимизации функции (18) при постоянном  $W$  с учетом (16), (17), (19), (9), (10), (12), (13).

Для уменьшения среднего времени  $t_{\text{ср}}$  перемещения головок можно расположить зоны  $P_D^{(\lambda)}$  на диске в соответствии с интенсивностями поступления в них признаков. В центре поля  $P_D$  разместим зону с максимальной интенсивностью, а по обе стороны от нее — зоны с меньшей интенсивностью в порядке ее убывания. Распределение интенсивностей по зонам может быть получено в процессе эксперимента на первой стадии измерений и в дальнейшем может уточняться. Полученное распределение может быть использовано для автоматического определения оптимальных размеров зон  $P_D^{(\lambda)}$  и их оптимального размещения на диске.

Заметим, что совмещение во времени приема признаков  $X$  с их преобразованием в признаки  $Y$  и с передачей признаков  $Y$  на диск приводит к увеличению мертвого времени ЭВМ как измерителя потока событий, и при этом мертвое время становится случайной величиной. Поэтому для обеспечения постоянства минимально возможного мертвого времени измерения на входе ЭВМ можно установить разравнитель небольшого объема (один-два регистра) либо увеличить мертвое время датчика признаков  $X$ .

**4. Суммарные потери времени при регистрации спектров с предварительной внешней группировкой признаков по зонам.** Величина  $T_{\text{нз}}$  суммарных потерь времени на остановки измерения при внешнем накоплении с группировкой признаков и при последующей внешней сортировке определяется формулами (3), (4) при  $\rho=3$ . Оптимальные значения рассмотренных выше параметров  $W$ ,  $z$ ,  $m$ ,  $n$ ,  $v_T$ ,  $v_H$  алгоритмов накопления и сортировки могут быть найдены из условий минимизации значения функции потерь при  $T_{\text{нз}}=T_{\text{он}}$ :

$$T_{\text{нз max}} = T_z + T_{\text{он}}. \quad (21)$$

Среднее значение  $T_{\text{нз}}$ , так же как и среднее значение  $T_{\text{нз}}$ , зависит от степени совмещения во времени процессов приема признаков  $X$ , формирования массивов в зонах ОП и передачи их на диск. Поэтому ниже для сравнения способов регистрации спектров будем использовать крайние значения  $T_{\text{нз}}$ , соответствующие интервалу (20).

На рис. 1—3 приведены зависимости величин потерь  $T_{\text{нз}}$  при  $\rho=0, 1, 3$  от объема  $v$  поля данных ОП. Эти зависимости получены для случая, когда при внешней сортировке каждому элементарному массиву  $s$  и каждой зоне соответствует хотя бы один признак из сортируемого их множества. На рис. 1 приведены графики функций  $T_{\text{нз min}}$ ,  $T_{\text{нз max}}$  (кривые 1—4 соответственно) для следующих условий:  $X \equiv Y$ ,  $r \equiv 1$ ,  $\omega_X = \omega_Y = \omega_q = 1$ ,  $N = 64K - 1$ ,  $M = 10^6$ ,  $z = 8$  при  $\rho = 3$ . На рис. 2 показаны функции  $T_{\text{нз min}}$  (кривая 1),  $T_{\text{нз min}}^{(Q)}$  для  $Q = 10^8$  и  $M = 10^6, 10^5, 10^4$  (кривые 2—4 соответственно) при остальных услови-

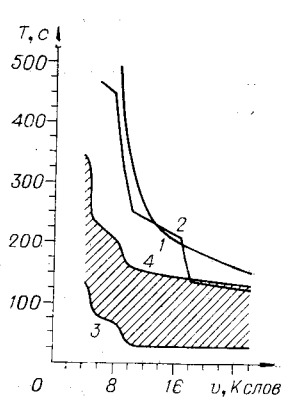


Рис. 1.

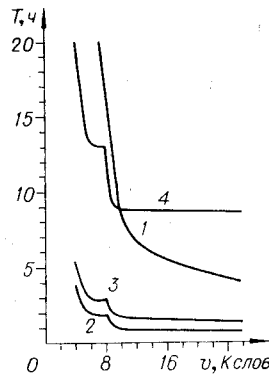


Рис. 2.

ях, принятых для кривых на рис. 1. На рис. 3 приведены функции  $T_{п0 min}$ ,  $T_{п3 min}$ ,  $T_{п3 max}$  (кривые 1—3 соответственно) для следующих условий при определенной функции (1):  $j_{max}=37$ ,  $r=2$ ,  $\omega_x=2$ ,  $\omega_x=\omega_d=1$ ,  $N=296K$ ,  $z=37$ ,  $M=10^6$ . Некоторые значения  $T_{п3 min}$ , полученные для этих условий, были приведены в таблице. На этих графиках величины  $T_{п3 min}$  и  $T_{п3 min}^{(0)}$  найдены, согласно (3), при  $T_{п3}=0$ .

Функция  $T_{п0 min}$  на рис. 1, 2 определена при условии, что массивы  $d$  упорядочены за время, равное нулю, а на рис. 3 — при условии, что массивы  $d$  не упорядочены.

Из рис. 1 следует, что для  $X \equiv Y$ ,  $r \equiv 1$  и  $N=64K$  максимальные потери  $T_{п3 max}$  при регистрации спектров с группировкой и сортировкой по зонам меньше минимальных потерь  $T_{п0 min}$  при внешней сортировке множества массивов  $d$  без предварительного накопления их на диске. Например, при  $v=10K$  уменьшение потерь времени характеризуется отношениями:

$$T_{п0 min}/T_{п3 max} \approx 2; T_{п0 min}/T_{п3 min} \approx 9.$$

Из рис. 2 видно, что по сравнению с сортировкой без предварительного накопления массивов  $d$  на диске регистрация спектров с группировкой и сортировкой по зонам эффективнее в заданных условиях при  $M \geq 10^5$ . Заметим, что уменьшение потерь  $T_{п3}$  замедляется в данном случае также при  $M \geq 10^5$ . Например, при  $v=10K$  потери  $T_{п3 min}$  уменьшаются примерно в 5 раз с увеличением  $M$  от  $10^4$  до  $10^5$  и примерно в 1,6 раза с увеличением  $M$  от  $10^5$  до  $10^6$ .

Регистрация с внешней группировкой и сортировкой по зонам в наибольшей степени эффективна при  $r > 1$ ,  $\omega_x > 1$ , больших значениях  $N$  и  $M$ . Например, для условий, соответствующих рис. 3, уменьшение потерь времени при  $v=10K$  характеризуется отношениями:

$$T_{п0 min}/T_{п3 max} \approx 7,5; T_{п0 min}/T_{п3 min} \approx 89.$$

**Заключение.** Приведенный анализ методов накопления и внешней сортировки спектрометрических данных на малых ЭВМ с магнитными дисками дает возможность провести рациональный выбор этих методов и оптимизацию их параметров, что позволяет существенно сократить потери времени на вынужденные перерывы измерения. По сравнению с методом регистрации без предварительного накопления признаков на диске предложенный метод регистрации спектров с предварительным внешним накоплением и группировкой признаков по зонам позволяет сократить потери времени на перерывы в измерении при некоторых реальных условиях на порядок и более. Этот метод положен в основу разработки программного обеспечения систем автоматизации спектрометрического эксперимента на ЭВМ HP2100A.

Из полученных результатов следует, что для сокращения потерь времени на вынужденные перерывы в измерении

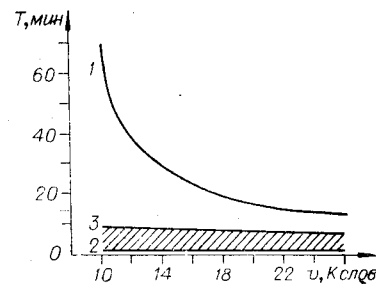


Рис. 3.

при регистрации спектров на ЭВМ с магнитными дисками достаточно большого объема целесообразно одновременно осуществлять следующие меры:

- 1) применение так называемого прямого алгоритма внешней поканальной сортировки с предварительным накоплением на диске массива, подлежащего сортировке;
- 2) предварительное преобразование исходных признаков событий в сортируемые признаки одинарной длины на стадии внешнего накопления;
- 3) предварительная группировка признаков по зонам поля данных диска на стадии внешнего накопления и внешняя сортировка заполненных частей зон;
- 4) совмещение во времени процессов приема исходных признаков в оперативную память, их преобразования с группировкой и передачи массивов на диск на стадии внешнего накопления;
- 5) оптимизация параметров алгоритмов внешнего накопления и внешней поканальной сортировки;
- 6) выбор объемов зон в поле данных на диске и их размещения в зависимости от статистического распределения признаков событий по зонам.

#### ЛИТЕРАТУРА

1. Л. А. Маталин, В. И. Смирнов. Накопление и обработка информации при многомерном анализе.— В кн.: Труды 6-й конференции по ядерной электронике. Т. 3. Ч. 1. М., Атомиздат, 1965.
2. С. С. Курочкин. Многомерные анализаторы. М., Атомиздат, 1968.
3. Ю. Д. Федоров. Методы многомерной регистрации.— В кн.: Труды 7-й конференции по ядерной электронике. Т. 1. Ч. 3. М., Атомиздат, 1970.
4. В. М. Цупко-Ситников. ЭВМ в задачах физики низких энергий.— В кн.: ЭВМ в экспериментальной физике. Лекции второй школы ОИЯИ. г. Алушта, 3-7 мая 1970 г. Дубна, изд. ОИЯИ, 1970.
5. В. М. Глушков, В. П. Гладун, Л. С. Лазинский, С. Б. Погребинский. Обработка информационных массивов в автоматизированных системах управления. Киев, «Наукова думка», 1970.
6. Р. Г. Оффенгенден, Ф. Н. Березин, П. В. Лапшук, Г. Б. Любанский, П. Н. Светличный. Дискретная система для измерения многомерных распределений вероятностей. Киев, «Наукова думка», 1969.
7. Р. Г. Оффенгенден, Ф. Н. Березин, П. В. Лапшук, А. М. Шур. Миллионно-канальная измерительная система для ядерно-физических измерений.— В кн.: Труды 7-й конференции по ядерной электронике. Т. 1. Ч. 3. М., Атомиздат, 1970.
8. B. Souček, R. J. Sprigad. Megachannel analyzers.— "IEEE Trans. Nucl. Sci.", 1966, vol. NS-13, № 1.
9. Р. Г. Оффенгенден, Л. А. Головач, С. И. Пилипчак. Алгоритм для измерения распределения вероятностей произвольного входящего потока с помощью ЭВМ.— В кн.: Автоматизация научных исследований (Материалы III Всесоюзной школы). Рига, «Зинатне», 1972.
10. D. C. Hensley. List or Sort?— Some experience with the ORIC multiparameters data acquisition system.— "IEEE Trans. Nucl. Sci.", 1973, vol. NS-20, № 1.

*Поступила в редакцию 10 июня 1975 г.*

УДК 631.327.11

**В. Л. КИПОТЬ, Ю. К. СИТНИКОВ**

*(Казань)*

### **СИНТЕЗ УСТРОЙСТВ ВЫВОДА ДУГ ОКРУЖНОСТЕЙ И ЭЛЛИПСОВ НА ЭКРАН ДИСПЛЕЯ**

При выводе графических элементов из ЦВМ на экран устройства графического воспроизведения информации (УГВ) необходимо осуществить преобразование цифрового описания элемента в отклоняющие