

СООБЩЕНИЯ  
ОБЪЕДИНЕННОГО  
ИНСТИТУТА  
ЯДЕРНЫХ  
ИССЛЕДОВАНИЙ  
ДУБНА



7366

P10 - 7366

Экз. чит. зала

Г. Элер, П.М. Гопыч , Г.В. Винель,  
В. Хабенихт, Л.А. Вылова

ЭКСПРЕСС-ПРОГРАММА  
ОБРАБОТКИ СПЕКТРОВ ЭПОС

Подгонка изолированных  
и слабoperекрывающихся пиков

**1973**

ЛАБОРАТОРИЯ ЯДЕРНЫХ ПРОБЛЕМ

P10 - 7366

Г. Элер, П.М. Гопыч\*, Г.В. Винель,  
В. Хабенихт, Л.А. Вылова

**ЭКСПРЕСС-ПРОГРАММА**

**ОБРАБОТКИ СПЕКТРОВ ЭПОС**

Подгонка изолированных  
и слабоперекрывающихся пиков

---

\* Харьковский государственный университет

**Научно-техническая  
библиотека  
ОИЯИ**

## 1. Введение.

Большинство пиков, имеющих в спектре, расположены обычно достаточно далеко друг от друга и являются либо изолированными, т.е. такими, влиянием на которые соседних пиков можно пренебречь, либо слабоперекрывающимися. Огибающая кривая для группы таких пиков имеет явно выраженную структуру, число максимумов этой кривой равно числу пиков, и все они могут быть найдены простой процедурой поиска, описанной в /1/. Если число образующих группу пиков невелико, то ее подгонка, как правило, не встречает трудностей. В противном случае, при решении задачи о разложении такого участка каким-либо из итерационных методов возможны затруднения, связанные с необходимостью многократного обращения матриц высокого порядка. Поэтому приходится предъявлять жесткие требования к объему памяти ЭВМ, ее быстродействию и к длине машинного слова. В настоящей работе описана созданная на основе метода статистической подгонки /2/ процедура для подгонки изолированных и слабоперекрывающихся пиков, которая является составной частью экспресс-программы обработки спектров ЭПОС /3/. Эта процедура, благодаря математической простоте используемого в ней метода, характеризуется высокой скоростью работы и свободна от перечисленных выше недостатков применения итерационных процедур. Подробно обсуждается задание начальных областей изменения параметров, приведена конкретная реализация алгоритма подгонки на языке АЛГОЛ-60.

## 2. Задание начальных параметров и областей их вариации.

В качестве исходных данных для подгонки используются результаты первичной обработки спектра первой частью программы ЭПОС (ЭПОС-I) /3/, которая автоматизирует поиск явно выраженных пиков, разбиение спектра на интервалы и приближённое первоначальное определение положений и амплитуд найденных пиков и параметров фона. Если для некоторого интервала полученные таким образом результаты явно неудовлетворительны, то производится, как это описано в /4/, дополнительный анализ, в результате которого экспериментальный спектр в этом интервале преобразуется в спектр без фона, и параметры пиков определяются заново. Новые приближенные значения положений и амплитуд не искажены влиянием сложного фона. В разных местах программы процедура подгонки работает с разным числом свободных параметров: либо три свободных параметра для каждого пика (положение, амплитуда, полуширина), либо два (положение и амплитуда). В соответствии с этим предусмотрена возможность вычисления в необходимых случаях приближённого значения полуширины пика по нескольким точкам на его склонах.

Для эффективной работы процедуры подгонки важное значение имеет правильное определение начальных областей вариации исходных параметров: с одной стороны, они не могут быть слишком широкими, т.к. в этом случае вероятность улучшения  $\langle \chi^2 \rangle$  резко уменьшается и, следовательно, возрастает время, необходимое для точного нахождения параметров; с другой стороны, они не могут быть слишком узкими, т.к. в этом случае значение каждого параметра приближается к истинному его значению очень мелкими шагами, и достижение минимума  $\langle \chi^2 \rangle$  затруднено или может оказаться невозможным, потому что в процессе подгонки при каждом улучшении приближения происходит

дальнейшее сужение областей вариации и дальнейшее уменьшение скорости приближения к истинным значениям параметров. Обстоятельством, позволяющим оптимально выбрать области вариации параметров, является возможность использования функций, количественно определяющих зависимость точности результатов эксперимента от его статистики. Эти функции и соответствующие графики приведены в работе /5/. Выбирая исходные области вариации начальных параметров пропорциональными этим ошибкам, необходимо учитывать, что предварительно найденные приближенные значения параметров иногда случайно могут быть определены гораздо точнее, чем это бывает обычно. В таких случаях  $\langle \chi^2 \rangle_1$  перед подгонкой и значение  $\langle \chi^2 \rangle_0$ , которое надо достичь в результате подгонки, одного порядка. В этом случае начальный диапазон изменения параметров необходимо сузить. Аналогично, бывают случаи, когда случайно найденные первоначальные параметры хуже обычных, тогда  $\langle \chi^2 \rangle_1$  может отличаться от  $\langle \chi^2 \rangle_0$  на несколько порядков, значения исходных параметров далеки от их истинных значений, и поэтому начальные диапазоны вариации необходимо соответственно расширить. Критерием качества исходных параметров, с помощью которых автоматически задаются оптимальные области вариации, служит отношение  $\langle \chi^2 \rangle_1 / \langle \chi^2 \rangle_0$ . В этом отношении величина  $\langle \chi^2 \rangle_0$  пока точно не известна, поэтому используется некоторая приближённая оценка. Таким образом, при задании начальных областей вариации надо знать и учитывать качество первоначального приближения и ожидаемый после подгонки результат.

Рабочие массивы, содержащие начальные параметры (массив pp) и начальные области их вариации (массив var) формируются вспомогательной процедурой parlim (см. блок-схему на рис.1 а). Здесь же для всего интервала в целом вычисляется значение  $\langle \chi^2 \rangle$  перед подгонкой  $\langle \chi^2 \rangle_1 = \text{clim}$  и приближенно оценивается то значение  $\langle \chi^2 \rangle$ ,

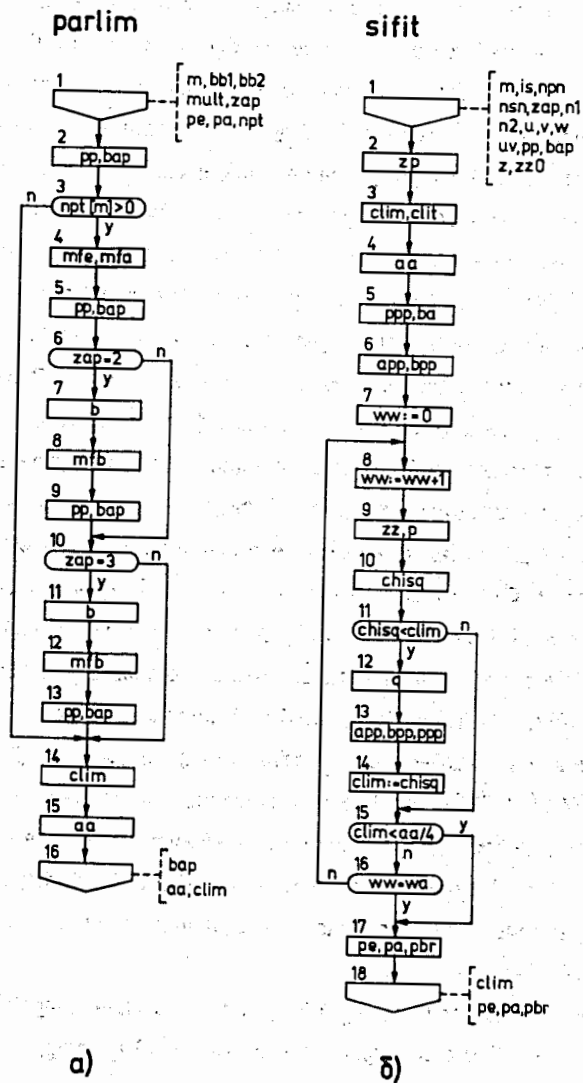


Рис. I

которое надо достичь в результате подгонки  $\langle \chi^2 \rangle_0 = \text{crit} \approx aa$ .

Вначале, в блоке (2) вычисляются исходные значения и области их вариации для четырех параметров фона. Если число пиков в  $m$ -ом интервале  $npt(m) > 0$  (3), то в (4...13) формируются массивы  $pp, bap$ , содержащие начальные значения параметров всех пиков, фона и области их вариации. Если число свободных параметров при подгонке  $zap=2$ , то полуширина (7) вычисляется по формуле  $b = \sqrt{bb1+bb2 \cdot k}$  ( $bb1, bb2$  - коэффициенты, описывающие зависимость полуширины от номера канала,  $k$  - номер канала), если  $zap=3$  (10), то полуширина пика вычисляется в (11) по нескольким точкам на его склонах так, как это делается в процедуре *intval* /1/. Начальные области вариации для каждого  $i$ -ого параметра вычисляются по формуле  $var(i)=mf$ , причём, для положений  $mf=mfe$  (4), для амплитуд  $mf=mfa$  (4), для полуширин  $mf=mfb$  (8,12). Аналитические выражения для ошибок в значении каждого параметра пика, которые связаны только со статистикой эксперимента ( $mfe, mfa, mfb$ ), приведены ниже /5/ :

$$mfe = (0.001 + (\sqrt{fon/a_1} + 0.5/\sqrt{a_1})/3) \cdot b_1,$$

$$mfa = (0.002 + 1/\sqrt{a_1} + 0.5\sqrt{fon/a_1}) \cdot a_1,$$

$$mfb = (0.002 + (1 + (fon/a_1)^{0.6})/a_1^{0.6}) \cdot b_1,$$

Здесь  $a_1$  - амплитуда,  $b_1$  - полуширина,  $fon$  - фон в месте расположения данного пика. В блоках (14,15) для интервала в целом вычисляются значения  $clim$  и  $aa$ .

### 3. Описание процедуры подгонки.

В программе ЭПОС делается попытка наиболее рационально организовать процесс обработки. Поэтому, если результаты предварительного анализа некоторого интервала спектра в первой части программы

(ЭПОС-1) являются уже удовлетворительными, то в таком интервале может сразу работать процедура подгонки *fit*. Если результаты предварительной обработки заведомо неудовлетворительны, то такой интервал подвергается дополнительному анализу, в результате которого происходит вычитание фона из экспериментального спектра и более мелкое разбиение интервала на подинтервалы. Подинтервалы, содержащие изолированные пики или группы слабоперекрывающихся пиков, имеющие сейчас экспериментальные значения без фона, подгоняются с помощью процедуры подгонки *sifit*. *Fit* и *sifit* вполне аналогичны, отличие их состоит только в том, что при работе *fit* для подгонки используются параметры пиков и параметры фона, а при работе *sifit* — только параметры пиков (экспериментальный спектр в соответствующем интервале фона уже не содержит). Т.к. в обеих процедурах заложен один и тот же принцип и отличие имеется только в деталях, далее объясняется подробно только работа процедуры *sifit*.

Блок-схема процедуры *sifit* показана на рис. 16. В блоке (2) вычисляется число свободных параметров при подгонке  $z_p = z_{ap} \cdot uv$  ( $z_{ap}$  — число свободных параметров для каждого пика,  $uv$  — число пиков в подинтервале). В (3) для рассматриваемого подинтервала находится первоначальное значение  $\langle \chi^2 \rangle$

$$\langle \chi^2 \rangle_1 = \text{clim} = \text{clit} = \left[ \sum_{i=n1}^{n2} \left( \frac{z(i) - \text{teo}(i)}{z(i) + ac} \right)^2 \right] / (n2 - n1 + 1) \cdot 10^6, \quad (I)$$

где  $n1, n2$  — начальный и конечный номера канала в подинтервале,  $z(i)$  — экспериментальное значение (без фона),  $\text{teo}(i)$  — теоретическое значение в  $i$ -ом канале, учитывающее, если это необходимо, вклад нескольких пиков

$$\text{teo}(i) = \sum_{j=1}^{uv} a_j \cdot \exp(-w((x_i - x_j)/b_j)^2). \quad (II)$$

$b_j$  — полуширина,  $x_j$  — положение,  $a_j$  — амплитуда  $j$ -ого пика, для симметричного гауссиана  $w = 2.7725887$ . Константа  $ac$  в формуле (I) добавлена, чтобы исключить возможные нулевые значения в знаменателе. Выбор константы в значительной мере произволен, можно взять её равной, например, 5% от средней амплитуды пиков в подинтервале. Формула для  $\langle \chi^2 \rangle$  особенно чувствительна к слабым отклонениям теоретических и экспериментальных значений вблизи основания пика, где значения  $z(i)$  малы. В (4) оценивается граничное значение  $\langle \chi^2 \rangle$  для подинтервала, которое надо достичь в процессе подгонки:

$$aa = \frac{10^6}{(n2 - n1 + 1)} \sum_{i=n1}^{n2} \frac{1}{z(i) + ac}.$$

Заметим, что найденные процедурой *parlim* аналогично тому, как здесь описано, величины  $\text{clim}$  и  $aa$  в *sifit* использованы быть не могут, т.к. они относятся ко всему интервалу в целом, а не к его отдельному подинтервалу. Значения  $\text{clim}$  и  $aa$ , найденные в *parlim*, нужны при работе процедуры *fit* и во внешнем блоке программы для управления. После того, как вычислены значения  $\text{clim}$  и  $aa$  для подинтервала, можно учесть влияние точности нахождения первоначальных параметров на ширину начальных диапазонов изменения параметров. Эти диапазоны вычисляются в (5):

$$ba(i) = f(\text{clim}/\text{crit}) \cdot \text{var}(i), \quad (III)$$

$$f(\text{clim}/\text{crit}) = 1 + 0.5 \cdot \ln(\text{clim}/\text{crit}), \quad (IV)$$

где  $\text{crit} = aa/4$  — оценка граничного значения  $\langle \chi^2 \rangle_0$ , которого надо достичь в результате подгонки,  $\text{var}(i)$  — диапазон вариации, вычисленный в *parlim* для  $i$ -ого параметра и учитывающий статистическую точность эксперимента. Формулы III, IV найдены эмпирически.

В блоке (6) для каждого параметра находятся нижняя и верхняя границы его области вариации (массивы  $arr$  и  $bpr$ , соответственно). Границы располагаются симметрично относительно начальных значений параметров из массива  $ppr$ . Теперь, когда готовы исходные данные для подгонки, она проводится  $wa$  раз в (8...16). Если граничное значение  $\langle \chi^2 \rangle$  достигнуто (15), то дальнейшая подгонка производиться не будет, и происходит выход к бл. (17), где суммируются полученные результаты. Эта часть процедуры  $sifit$ , использующая методику статистической подгонки, оформлена для наглядности в виде самостоятельной процедуры  $stafit$ , алгольный текст которой дан в приложении.

В (9) для каждого параметра генерируется свое, равномерно распределенное в области изменения этого параметра квазислучайное число  $z_j$ , с помощью которого находится новое случайное значение этого параметра в области его вариации. Полученные таким образом случайные значения для всех  $z_j$  параметров запоминаются в массиве  $p$ . С их помощью в (10) вычисляется новое значение  $\langle \chi^2 \rangle$  для подинтервала (оно называется  $chisq$ ) по формулам I, II, в которых, однако, вместо  $x_j, a_j, b_j$  используются соответствующие значения из массива  $p$ .

Если  $chisq < clim$  (11), значит, в результате этого шага подгонки получено улучшение аппроксимации экспериментальных данных, в (12) вычисляется параметр  $q = \sqrt{chisq/clim}$ , управляющий сужением диапазонов вариации, в (13) набор параметров из массива  $p$  запоминается в массиве  $ppr$  и находятся новые нижние и верхние границы

$$\begin{aligned} arr(i) &= p(i) - q \cdot (1-xb) \cdot ba(i)/2, \\ bpr(i) &= p(i) + q \cdot (1+xb) \cdot ba(i)/2, \end{aligned} \quad (V)$$

где  $xb = (p(i) - ppr(i)) / (bpr(i) - arr(i))$ . В (14) переменная  $clim$  получает значение, равное  $chisq$ . Если условие (11) не выполняется, то значения параметров и области их вариации остаются без изменения.

При  $ww < wa$  (число шагов меньше максимально допустимого) через (8) переходим к (9), где вычисление новых случайных значений параметров повторяется и т.д., пока либо не будет достигнуто предельное значение  $crit = aa/4$ , либо не исчерпаются все возможные  $wa$  шагов.

Качественно ход подгонки показан на примере (см. рис. 2). Для примера взят сконструированный пик без статистики, параметры которого были заданы при конструкции и поэтому точно известны. Сплошные линии обозначают нижнюю и верхнюю границы диапазона изменений каждого параметра, штрихованная линия обозначает ход изменения самого параметра в процессе подгонки. Верхние точки этих линий обозначают соответствующие начальные значения, крестики показывают истинные значения положения, амплитуды и полуширины, т.е. те, которые были заложены при конструкции. Отклонение начальных значений параметров от их истинных значений обычно для результатов ЭПОС-I. Слева, на рисунке показан номер шага подгонки  $ww$ , в котором было получено улучшение приближения, и соответствующее ему значение  $\langle \chi^2 \rangle$ , вверху даны шкалы для всех трех параметров пика. При каждом улучшении  $\langle \chi^2 \rangle$  границы областей вариации изменяются по формулам (V), в результате чего ход процесса подгонки для каждого параметра выглядит как изломанная, сужающаяся книзу полоса. На рисунке отчетливо видна асимметрия областей изменения параметров. Эти области открываются больше в том направлении, изменение параметра в котором только что дало улучшение в  $\langle \chi^2 \rangle$ . Внизу ширина каждой полосы соответствует точности определения данного параметра в конце подгонки (в нашем примере - 15 шагов). При увеличении числа шагов приближение продолжает улучшаться и при  $ww=30$   $\langle \chi^2 \rangle = 3.5$ , а при  $ww=100$   $\langle \chi^2 \rangle = 2.2 \cdot 10^{-4}$ . При  $ww=15$  для примера на рис. 2 отклонение в положении  $\Delta b/b = 7.2 \cdot 10^{-4}$ , в амплитуде  $\Delta a/a = 1.7 \cdot 10^{-4}$ , в полуширине  $\Delta b/b = 7.1 \cdot 10^{-4}$ .

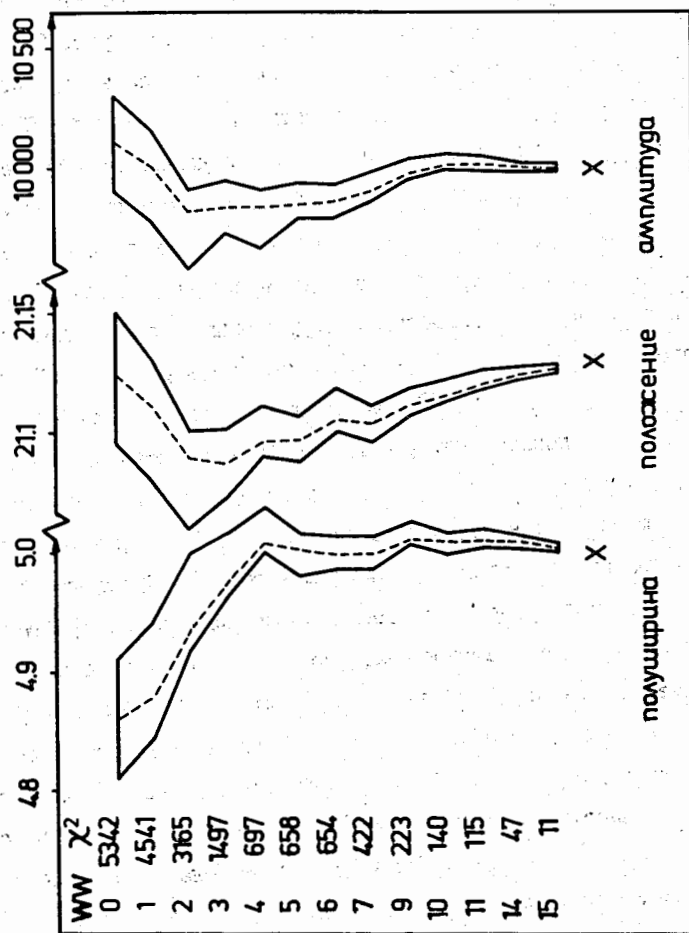


Рис. 2

#### 4. Обсуждение свойств процедуры подгонки.

Поскольку процедуры *fit* и *sifit* используют статистический метод подгонки, их результаты носят существенно статистический характер. При подгонке одного и того же спектра несколько раз с помощью, например, процедуры *sifit*, результаты всех подгонок будут, вообще говоря, разные. Действительно, в разных подгонках используются различные наборы случайных чисел *zz*, и поэтому закон, по которому происходит изменение параметров и границ их областей вариации, не является строго детерминированным. Общий характер такого изменения всегда одинаков (изломанная сужающаяся книзу полоса на рис. 2), однако каждый частный случай отличается от другого, и отличие носит случайный характер. Это означает, что подгонка может заканчиваться в различных точках вблизи минимума  $\langle \chi^2 \rangle$  в зависимости от того, какой конкретный набор квазислучайных чисел *zz* используется процедурой подгонки и по какому пути в результате этого шло приближение подгоняемых параметров от их первоначальных значений до значений этих параметров вблизи минимума  $\langle \chi^2 \rangle$ . Разброс конечных результатов, а, значит, и точность методики подгонки зависит от выбора квазислучайных чисел *zz*. Эту зависимость демонстрируют рисунки 3 и 4. Тот же сконструированный пик, с теми же начальными параметрами, что и на рис. 2, подгонялся с 350 разными наборами чисел *zz*. Результаты для каждого параметра пика и для  $\langle \chi^2 \rangle$  показаны в виде гистограмм. У каждой гистограммы приведено соответствующее значение относительного среднеквадратичного отклонения, масштаб по горизонтали всегда отложен в единицах этого отклонения. Из рисунков видно, что большинство значений, получаемых в результате подгонки, имеют точность, которая достаточна для большинства применений. Количество хороших подгонок преобладает, однако имеются



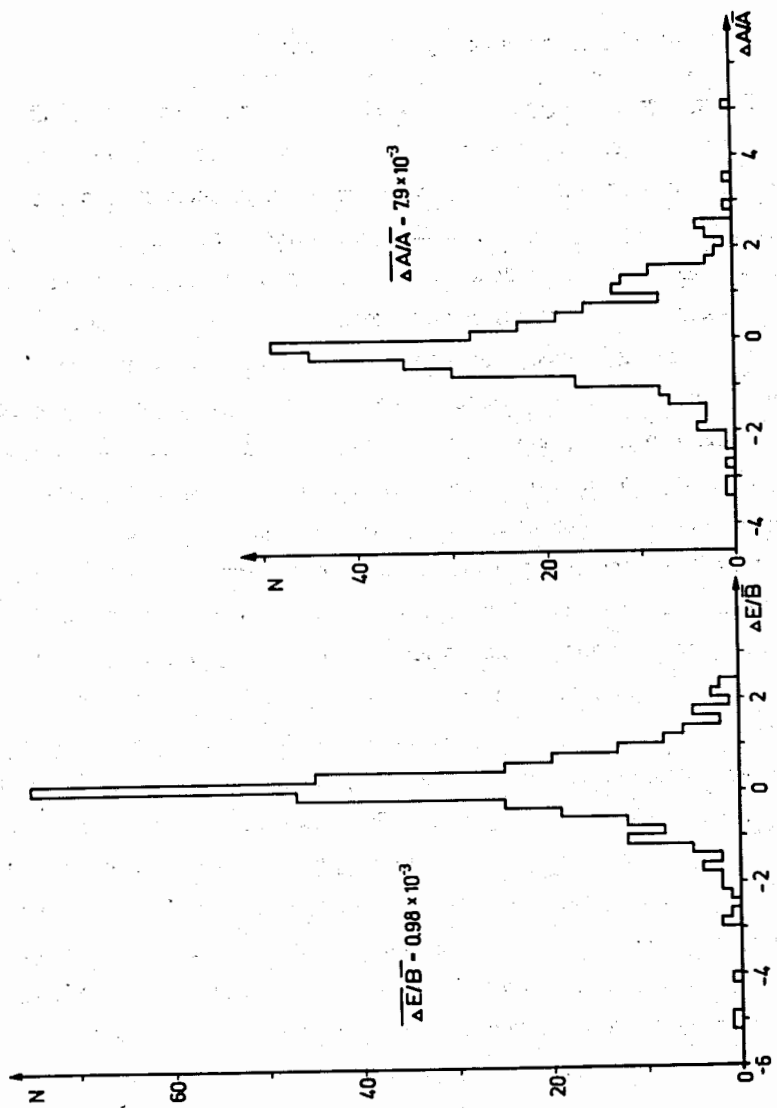


Рис. 3

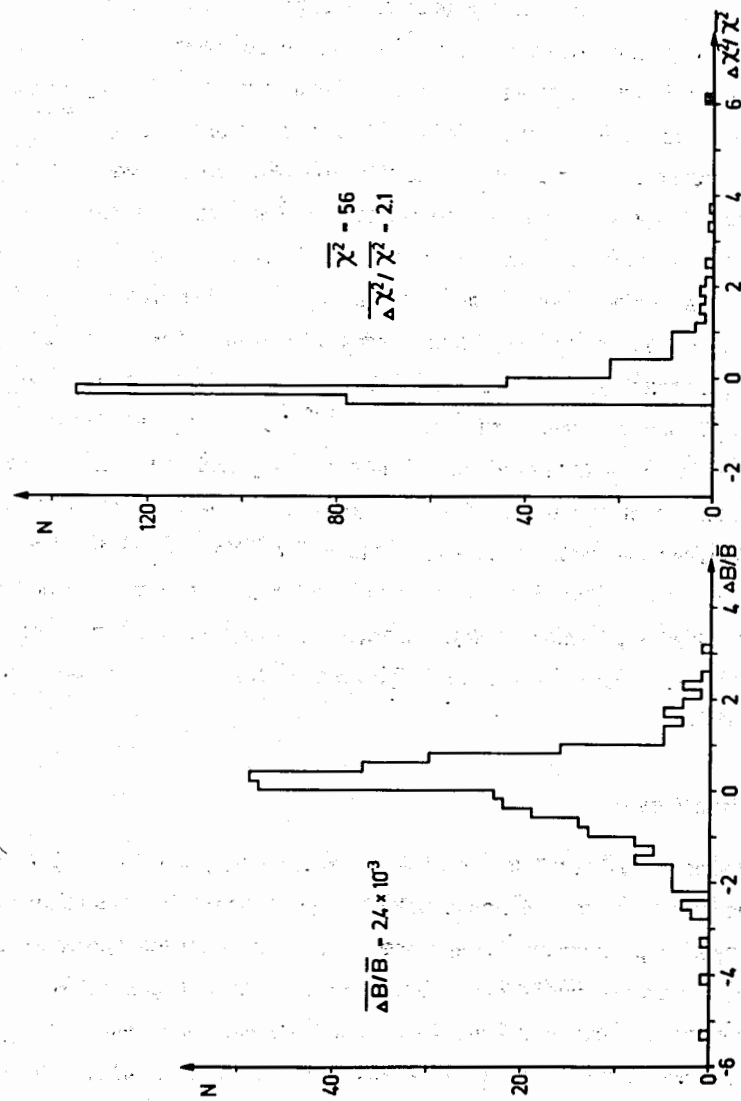


Рис. 4

значения, отстоящие от средних на четыре единицы и более. Особенно затянута распределение для  $\langle \chi^2 \rangle$ , для него имеются значения, превышающие 6 относительных среднеквадратичных отклонений (на гистограмме рис. 4 число таких отклонений заштриховано). Путём специального подбора чисел  $z$  число плохих подгонок можно существенно сократить, однако это не может полностью исключить возможности случайного появления плохих результатов. Поэтому, чтобы гарантировать точность результатов после подгонки не хуже, чем точность, содержащаяся в эксперименте, надо исключить случаи, дающие вклад в край распределений для каждого параметра на рис. 3 и 4. Для этого достаточно обрезать край распределения  $\langle \chi^2 \rangle$ , что достигается сейчас введением следующего условия: подгонка прекращается по достижению граничного значения  $\langle \chi^2 \rangle = \text{crit}$  (условие I5 на блок-схеме рис. 16). Асимметрия распределений для амплитуды и полуширины связана с тем, что эти два параметра сильно взаимосвязаны. При одном и том же значении  $\langle \chi^2 \rangle$ , если амплитуда найдена в результате подгонки несколько завышенной, то значение полуширины будет несколько занижено, и наоборот.

#### 5. Заключение.

Для оптимальной и надёжной работы процедуры подгонки, необходимо знать ожидаемый после подгонки результат. Высокая скорость работы процедуры подгонки позволяет систематически исследовать этот вопрос с помощью многократного моделирования спектров с квазиэкспериментальной статистикой. Такое исследование было выполнено и его предварительные результаты (формулы для  $mfe, mfa, mfb$ ) уже используются для управления. В дальнейшем процедура подгонки будет содержать определённый при таком анализе граничный параметр  $\text{crit}$  вместо

используемой сейчас его оценки ( $aa/4$ ), и функция (IV), определяющая начальный диапазон изменения параметров, будет соответствующим образом уточнена. В этом случае содержащаяся в экспериментальных результатах точность достигается за минимальное время.

#### ЛИТЕРАТУРА.

1. Г. Элер, П.М. Гопыч, Г.В. Винель, В.Хабенихт, Л.А. Вылова, Р10-6818, Дубна, 1972.
2. Г. Элер, Р 11-6816, Дубна, 1972.
3. Г. Элер, П.М. Гопыч, Г.В. Винель, В.Хабенихт, Л.А. Вылова, Р 10-6817, Дубна, 1972.
4. Г. Элер, П.М. Гопыч, Г.В. Винель, В.Хабенихт, Л.А. Вылова, Р 10-7365, Дубна, 1973.
5. Г. Элер, П.М. Гопыч, Г.В. Винель, В.Хабенихт, Л.А. Вылова, Р 10-6819, Дубна, 1972.

Рукопись поступила в издательский отдел  
27 июля 1973 года.