

УДК 519.254, 543.545

© А. Л. Буляница, Г. Е. Рудницкая, А. А. Евстапов

## ИНФОРМАТИВНЫЕ СИГНАЛЫ ПРИ ЭЛЕКТРОФОРЕЗЕ НА МИКРОЧИПЕ: МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ОЦЕНКИ

Рассмотрены возможные направления развития алгоритмов обработки и оценивания аналитических сигналов электрофоретических приборов на основе микрофлюидных чипов: моделирование информативных сигналов (пиков) и неинформативных составляющих (дрейф базовой линии и высокочастотная шумовая помеха). На основе этой информации могут быть установлены источники помех, определены наилучшие алгоритмы оценивания сигнала и модифицированы критерии разделения соседних пиков. На примере электрофореграмм, полученных при разделении смеси фрагментов ДНК на микрофлюидном чипе, показана эффективность традиционных экономичных методов первичной обработки сигналов — компенсация трендов и фильтрация, а также состоятельность новой методики оценивания площади пика и применимость метода главных компонент при многоканальных измерениях в случаях сложности локализации границ пиков.

### ВВЕДЕНИЕ

Развитие микрочиповых технологий для аналитического приборостроения является междисциплинарной задачей, включающей в себя следующие направления: 1) разработку и расчет топологий (архитектур) для микрочипов, обеспечивающих реализацию заданной методики анализа; 2) математическое моделирование процессов в микрочиповых устройствах и приборах с целью последующей адекватной интерпретации аналитического сигнала; 3) создание и отработку технологии изготовления (в том числе выбор и синтез материалов, формирование функциональных микро- и наноструктур с заданными свойствами, герметизацию каналов, реакторов и сосудов и т. п.); 4) постановку методики анализа (включая подтверждение получаемых результатов референтными методами); 5) разработку процедур и методов обработки результатов измерений, оценивания информативных сигналов, интерпретации получаемых результатов и т. д. Несмотря на очевидную автономность и самостоятельность, в рамках которых допустимо независимое развитие указанных направлений, успешная реализация микрочиповых технологий в аналитических приборах неизбежно требует согласованной разработки всех представленных направлений. Особо следует подчеркнуть, что развитие микрочиповых технологий непосредственно связано с процессами миниатюризации аналитических приборов, но даже при бурном прогрессе микропроцессорной вычислительной техники приборы на микрочиповой платформе требуют разработки эффективных, быстрых и простых методов обработки результатов измерений и оценивания информационных сигналов. Это в зна-

чительной степени определяет актуальность данной работы.

Информативные сигналы при электрофорезе на микрочипе, как и для макроаналога — капиллярного электрофореза, представляют собой совокупность пиков. Однако автоматический перенос разработанных для макромасштаба вычислительных алгоритмов на микроаналоги может приводить к значительным погрешностям измерения, а в ряде случаев — к ошибкам. Основные специфические особенности информативных сигналов при микроанализе таковы: а) снижение отношения сигнал/шум как следствие значимого (более чем на 2 порядка) уменьшения измеряемых количеств компонентов пробы; б) абсолютное и относительное сужение пиков (как следствие изменения соотношения между конвективной и диффузионной составляющими массопереноса, по существу величины характеристического числа Пекле); в) изменение вкладов различных источников помех, приводящее к изменению математических моделей формы пиков, в частности негауссовый их характер. Эти особенности требуют анализа погрешностей, что может позволить выявить их источник и как следствие изменить условия измерения или/и предложить оптимальный метод оценивания и компенсации. Также, естественно, новые математические модели формы пиков могут потребовать новых методов оценивания их параметров (высота, ширина, площадь), а также и новых критериев разделения.

### ИСХОДНЫЕ ДАННЫЕ

В Лаборатории информационно-измерительных био- и хемосенсорных микросистем Института

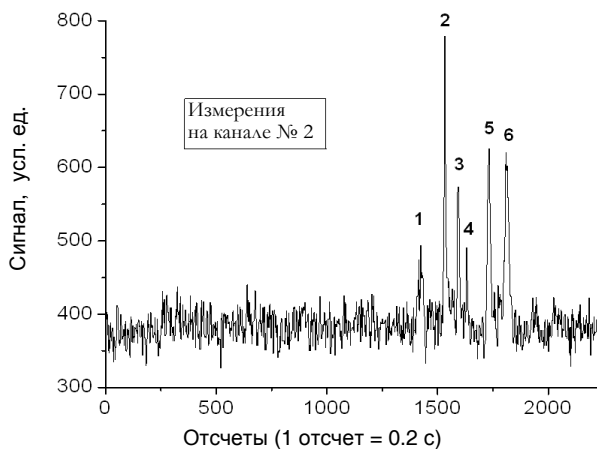


Рис. 1. Электрофоретическое разделение смеси фрагментов ДНК на микрофлюидном чипе

аналитического приборостроения РАН (Санкт-Петербург) созданы макеты микрофлюидных аналитических систем на основе электрофоретического метода разделения компонентов пробы на микрочипе с детектором лазер-индуцированной флуоресценции [1, 2]. На стеклянных микрофлюидных чипах с одноканальной топологией (изготовленных совместно с ЗАО "Светлана-полупроводники", Санкт-Петербург) [3] проведено электрофоретическое разделение смеси фрагментов ДНК 150-200-250-300-400-450 пар оснований, меченых флуоресцентной меткой (ЗАО "Синтол", Москва). Электрофореграмма разделения представлена на рис. 1. Получены воспроизводимые результаты анализов, а соответствующие информативные сигналы использованы в качестве иллюстраций для решения различных задач при разработке методов обработки результатов измерений, оценивания информативных сигналов и интерпретации получаемых результатов.

Сигналы флуоресценции, измеренные при сканировании одной и той же пространственной области сепарационного канала микрофлюидного чипа, сдвинуты во времени на 0.2 такта (0.04 с) по каждому измерительному каналу. Измерения осуществлялись последовательно по 6 идентичным каналам. При этом требовалось оценить а) степень вариабельности измерений по каналам, б) влияние предварительной обработки — фильтрации и компенсации базовой линии как линейного тренда (ЛТ) малого порядка.

Близость измерений была оценена по количественному совпадению:

- 1) временного положения и амплитуды пиков,
- 2) оценок величин площади пика (две методики), а также

3) интегрально, по величине коэффициента корреляции.

На рис. 2 представлен информативный фрагмент электрофореграммы, содержащей 6 пиков различной выраженности и формы (нумерация пиков слева направо), измеренных на 1-м и 4-м каналах. Времена выхода компонент соответственно от 2 до 3.5 минут. Визуально наблюдается хорошее совпадение измерений для каналов в областях явно выраженных пиков 2, 3, 5 и 6 и существенное отличие измерений как в области отсутствия пиков (базовая линия), так и в области относительно слабо выраженных пиков 1 и 4.

### БАЗОВЫЕ ПОЛОЖЕНИЯ МОДЕЛЕЙ ПИКА И НЕИНФОРМАТИВНЫХ СОСТАВЛЯЮЩИХ СИГНАЛА

**Положение 1.** Аппроксимация вершины пика параболической кривой.

Использованы 3 точки — точка максимума (такт 0) и 2 соседних измерения с абсциссами  $\pm 1$ ; их ординаты —  $y_1$ ,  $y_2$  и  $y_3$  соответственно. Вершина пика будет

$$x_{\max} = \frac{y_3 - y_1}{2(2y_2 - y_1 - y_3)}; \quad y_{\max} = y_2 + \frac{(y_3 - y_1)^2}{8(2y_2 - y_1 - y_3)}.$$

Если точки лежат на одной прямой, то за координату вершины пика принимается центральная точка  $(0, y_2)$ .

**Положение 2.** Модель случайного высокочастотного шума неизвестна.

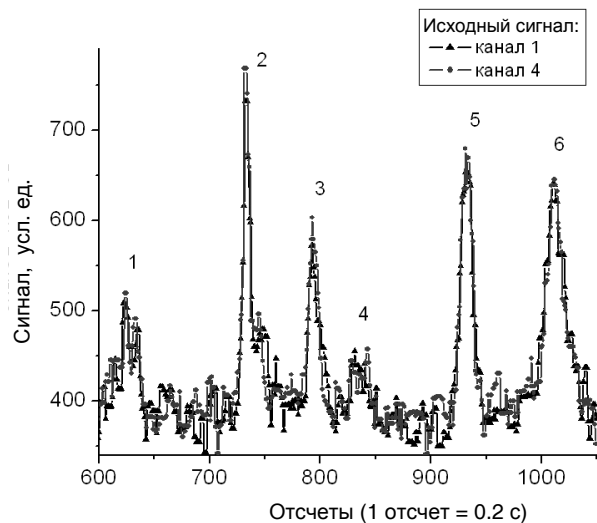


Рис. 2. Фрагмент электрофореграммы смеси фрагментов ДНК 150-200-250-300-400-450 оснований

Таким образом, одной из задач явилось выявление возможного закона распределения случайной высокочастотной составляющей.

**Положение 3.** Базовая линия аппроксимирована линейными трендами от нулевого до 4-го порядка.

Далее будут использованы обозначения ЛТ0 и ЛТ4 соответственно.

**Положение 4.** Форма пика может быть отлична от гауссовой.

Моделирование формы также является предметом исследования. При этом априорно предполагаем, что передний и задний фронты пика могут быть адекватно аппроксимированы ЛТ1 с большим тангенсом угла наклона.

**Положение 5.** Для цифровой фильтрации высокочастотного шума использован фильтр нижних частот Баттерворта (тангенсный, второго порядка).

Как было показано ранее, его использование при выборе относительной частоты среза 0.05–0.10 обеспечивает приемлемое уменьшение дисперсии шума при сохранении удовлетворительной точности, хорошего временного разрешения и простоты реализации [4–6].

#### РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ ИНФОРМАТИВНЫХ И НЕИНФОРМАТИВНЫХ СОСТАВЛЯЮЩИХ СИГНАЛА

Оценивание параметров шума по начальному фрагменту 300 тактов, заведомо не содержащему пиков, при измерениях по 6-ти каналам дает схожие результаты: близкие средние значения (376–382 кванта), среднеквадратичные отклонения измерений (17–22 кванта) и малая значимость дрейфа базовой линии (ЛТ1). Основные характеристики приведены далее в табл. 1.

Вывод о малой значимости вклада линейно

изменяющегося слагаемого можно сделать, сопоставив стандартные отклонения до и после компенсации ЛТ1 (столбцы 3 и 5 табл. 1), в соответствии с критерием Фишера [7]. Последующие компенсации ЛТ2–ЛТ4 также свидетельствует о малости вклада нелинейных слагаемых в дрейф базовой линии.

После компенсации ЛТ2 был проведен анализ возможного закона распределения случайной составляющей помехи на основе 300 измерений по каналу 5. При расчете моментов распределения получено: математическое ожидание  $m = 386$ , стандартное отклонение  $\sigma = 20.1$ , коэффициент асимметрии (нормированный центральный момент 3-го порядка)  $A = 0.069$  и коэффициент эксцесса (нормированный центральный момент 4-го порядка)  $\varepsilon = 2.572$ .

В качестве гипотетического теоретического закона распределения рассматривались: а) нормальное с  $A = 0$ ,  $\varepsilon = 3 \Rightarrow N(386.00; 404.01)$ ; б) треугольное (Симпсона) с  $A = 0$ ,  $\varepsilon = 2.4 \Rightarrow S(386.00; 49.23)$  (второй параметр — полуразмах помехи  $\sigma\sqrt{6}$ ) и в) смесь двух гауссиан, представленных в нормированных величинах как  $(1-\alpha)N(0,1) + \alpha N(k,1)$ . Т. о., последний закон соответствует сумме двух гауссовых кривых с одинаковыми дисперсиями и математическими ожиданиями, смещенными на  $k\sigma$ .

По  $(A, \varepsilon)$  можно оценить нормированные параметры  $(\alpha, k)$ , а по  $(m, \sigma)$  пересчитать в размерные (естественные) параметры по схеме:

$$\sigma^2 = 1 + \alpha(1-\alpha)k^2,$$

$$\mu_3 = \alpha(1-\alpha)(2\alpha-1)k^3 \Rightarrow A = \mu_3 / \sigma^3,$$

$$\mu_4 =$$

$$= 3 + \alpha(1-\alpha)k^2 \cdot (6 + (3\alpha^2 - 3\alpha + 1)k^2) \Rightarrow \varepsilon = \mu_4 / \sigma^4.$$

**Табл. 1.** Интегральные характеристики (математическое ожидание, стандартное отклонение) неинформативных составляющих сигнала

№ канала	Среднее	Ст. откл.	ЛТ1 (дрейф базовой линии)	Ст. откл.*
1	381.75	17.64	381.03 – 0.0048·i	17.65
2	376.48	19.02	367.95 + 0.0566·i	18.41
3	380.91	17.19	382.50 – 0.0105·i	17.19
4	380.97	17.97	371.77 + 0.0605·i	17.32
5	378.75	21.88	370.48 + 0.0549·i	21.39
6	377.95	17.94	377.49 + 0.0031·i	17.91

**Примечание.** \* После компенсации дрейфа базовой линии как ЛТ1.

**Табл. 2.** Распределение экспериментального (гистограмма) и гипотетического числа точек по интервалам

Интервал	1	2	3	4	5	6	7	8	9	$\chi^2_6$
Гистограмма	15	32.5	37.5	61	54.5	39	29	23	8.5	—
Закон а)	16.4	24.4	41.9	55.1	58.3	48.0	31.0	15.7	6.3	10.14
Закон б)	18.3	27.6	39.9	52.3	56.7	44.8	32.5	20.1	7.7	4.79
Закон в)	17.0	28.0	41.8	50.2	52.8	50.5	35.5	17.3	5.5	11.15

Расчет значений нормированных параметров:  $\alpha = 0.55$  и  $k = 1.9 \Rightarrow$  при переходе к естественным параметрам получаем аддитивную смесь гауссовых кривых  $0.45N(370.74; 213.37) + 0.55N(398.50; 213.37)$ . Упомянутые гипотезы законов распределения далее названы закон а), б) и в).

Диапазон измерения сигнала от 324 до 434 квантов изначально разбит на 11 интервалов длиной 10 квантов, первые 3 интервала были объединены в один [324; 354] по причине малости числа измерений. Точки, попавшие на границу, распределяются поровну между соседними интервалами. Т. о., общее число точек 300, число интервалов 9 и число степеней свободы 6 (т. к. производится оценивание 2 параметров распределения). Необходимая информация для оценки доверительных вероятностей по критерию согласия Пирсона представлена в табл. 2.

Доверительные вероятности равны 12.5 %, 57 % и 8.7 % соответственно. Т. о., наиболее вероятной является гипотеза треугольного распределения Симпсона. Однако и первая гипотеза не может быть безоговорочно отвергнута.

Другим важным следствием полученных данных является оценка наличия пика. Полагаем, что для достаточно достоверного выявления пика на фоне шума высота последнего должна составлять по крайней мере  $(5 \div 6)\sigma$ , т. е., по нашим данным, не менее 90 квантов. Эта величина может являться нижним граничным значением, определяющим чувствительность измерений при данной приборной реализации.

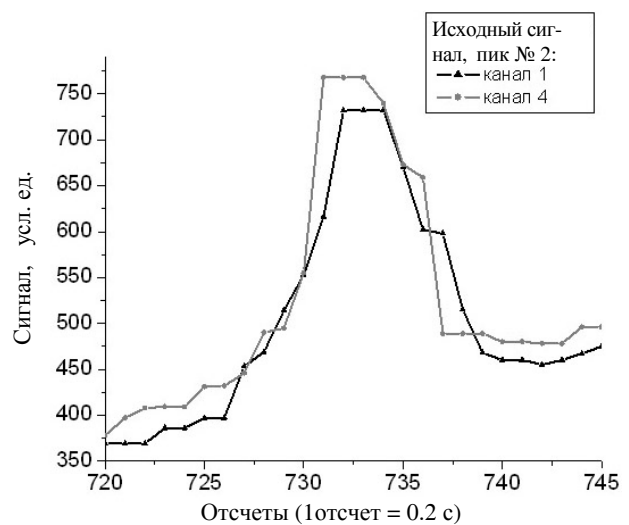
Математические модели формы пика также могут быть аналогичным образом исследованы на основе вероятностного подхода, в частности с использованием оценок первых 4 моментов.

### ОЦЕНИВАНИЕ ПАРАМЕТРОВ ПИКОВ РАЗЛИЧНОЙ ВЫРАЖЕННОСТИ

Для интерпретации результатов анализа необходимо получение следующих оценок наиболее важных параметров пика: а) временной координаты максимума пика, что при заданных условиях

анализа позволяет однозначно идентифицировать размер (длину) фрагмента ДНК; б) площадь пика, характеризующую количество (концентрацию) соответствующего компонента. Оценка амплитуды (высоты) пика в значительной степени является вспомогательной, т. к. она обычно используется для определения временного положения вершины пика. Поэтому представляется важным рассмотреть и сравнить оценки параметров пика, прежде всего его площади, полученные на основе различных алгоритмов.

На рис. 3 и 4 представлены фрагменты исходных электрофореграмм для пиков различной выраженности, измеренных на произвольно выбранных каналах. Из представленных данных следует, что пик № 2 достаточно ярко выражен. Компенсация дрейфа базовой линии как ЛТГ1 из-за его малости (см. данные табл. 1) достаточно слабо влияет на оценки его параметров. В табл. 3 представлены координаты вершины пика по 6 каналам измерения. Следует отметить очень хорошее совпадение временного положения максимумов пика по всем каналам, при отличающейся величине

**Рис. 3.** Фрагмент электрофореграммы, соответствующий пику № 2

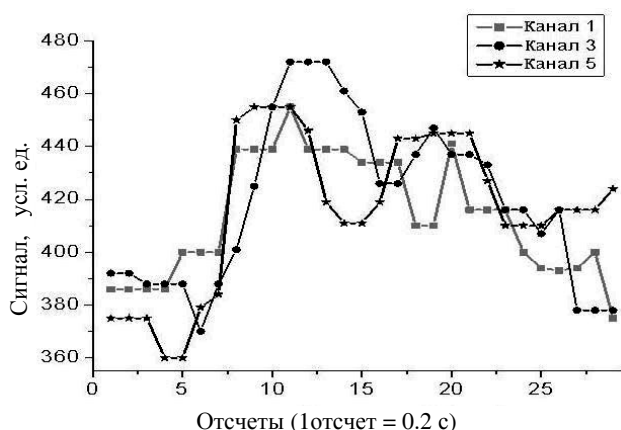


Рис. 4. Фрагмент электрофореграммы, соответствующий пику № 4

максимума (от 693.46 до 779.05). Примерно в таких пределах (около 90 единиц) меняется и высота пика.

При количественной оценке соответствующего компонента пробы необходимо определение площади пика. Оценки площади пика могут быть по-

лучены следующими методами: по приближенной формуле  $h\Delta_{1/2}$ , где  $h$  — высота пика и  $\Delta_{1/2}$  ширина на полувысоте (далее Площадь1) и традиционным интегрированием по области, ограниченной огибающей пика (далее Площадь2). Использование традиционной методики интегрирования (или суммирования) может быть неэффективным при малых отношениях сигнал/шум, поскольку в этих условиях затруднительна локализация границ пика. Сравнительные оценки площади пика № 2, вычисленные этими методами для 6 каналов измерения, приведены в табл. 4.

Коэффициенты корреляции между оценками площадей (по 4 расчетным методикам — выборочный, медианный, ранговый, знаковый коэффициенты корреляции (КК) [8]) равны соответственно 0.924; 0.918; 0.829; 0.733, т. е. наблюдается очень сильная корреляция, но характер связи между этими зависимостями достаточно далек от линейного. Последнее объясняется тем, что по-разному интерпретируется понятие "площади пика", а именно различный вклад измерений на различных фазах формирования пика. Однако в идеализированном случае треугольного пика, т. е. остроговершинного, с линейными фронтами, оценки площадей должны совпадать.

Табл. 3. Параметры пика (положение максимума и величина максимального сигнала) на 6 каналах измерения

Канал	1	2	3	4	5	6
Положение максимума, такт	733.15	733.03	732.64	732.38	732.36	733.50
Величина макс. сигнала	732.68	779.05	749.24	773.92	732.88	693.46

Табл. 4. Оценки площади пика № 2 для 6 каналов измерения

Канал	Высота	Ширина на полувысоте	Площадь1	Площадь2
1	309.41	6.56	2030	2603
2	354.84	6.02	2136	2955
3	335.03	6.32	2117	2792
4	334.29	5.92	1980	2600
5	316.10	6.75	2134	2882
6	265.96	7.65	2035	2584

В первом случае площадь пика есть  $2072 \pm 66$  (коэффициент вариации 3.2 %), во втором —  $2736 \pm 162$  (коэффициент вариации 5.9 %). Видно, что оценки площади по модифицированной методике (Площадь1) значительно меньше (на 24 % по средним значениям) соответствующих традиционных оценок (Площадь2). Это уменьшение оценок объясняется двумя особенностями пиков: а) фронты не являются идеальными ЛТ1, т. е. более адекватно аппроксимируются ЛТ2 или полиномом более высокого порядка; б) пики не являются островершинными (во всех случаях аппроксимацией вершины является либо парабола, либо ее вырожденный случай — горизонтальная прямая). Тем самым для гладковершинных пиков без существенной нелинейности фронтов формула площади должна быть уточнена как  $h(1 + \Delta_{1/2})$ . Тогда оценки площадей по обоим методикам существенно "сблизятся", и расхождение между средними оценками площадей составит 13 %. Вместо имеющихся данных Площадь1 в табл. 4 получатся измерения 2340, 2490, 2451, 2315, 2450 и 2301 соответственно ( $2391 \pm 82$ , коэффициент вариации 3.4 %).

Выделение фрагмента сигнала, содержащего пики компонент, позволяет оценить корреляционную матрицу, основанную на измерениях по 6 каналам. Величины выборочных КК для фрагментов сигналов, полученных на различных каналах измерений, соответствующих отдельному выраженному пику (например, № 3, 5 или 6), дают еще более сильную корреляцию — от 0.953 до 0.989.

Пик № 4 (см. рис. 4) выражен достаточно слабо. Более того, представляется обоснованной интерпретация данного объекта не как одного, а как двух слабо разрешенных пиков. Имеется несинхронность (малая корреляция) измерений по различным каналам 1–6 и, как следствие, очень малые значения элементов корреляционной матрицы (верхний диагональный ее фрагмент показан ниже):

1	0.677	0.797	0.601	0.687	0.639
	1	0.789	0.185	0.284	0.408
		1	0.465	0.636	0.723
			1	0.555	0.477
				1	0.601
					1

Элемент  $(i, j)$  представляет собой выборочный КК между измерениями на  $i$ -м и  $j$ -м каналах.

Применение метода главных компонент (или факторного анализа [8]) позволяет сформировать оптимальные (ортогональные друг другу) линейные комбинации измерений. Наибольшему собственному числу корреляционной матрицы

$\lambda_{\max} = 3.896$  соответствует собственный вектор весовых коэффициентов

$$[0.466; 0.354; 0.468; 0.338; 0.397; 0.408].$$

В главную компоненту, следовательно, войдут измерения, выполненные на каналах 1–6 с указанными весовыми коэффициентами.

Следующие собственные числа будут 0.973; 0.496; 0.386; 0.173 и 0.078 (сумма всех собственных чисел должна совпадать с размерностью матрицы 6 как первый инвариант или след матрицы). В случае близости большинства КК к 1 главная компонента строится на основе собственного вектора с равными координатами, и суммирование будет производиться с равными весами (т. е. практически осреднение). На рис. 5 представлен график первой главной компоненты.

Представленный график главной компоненты допускает следующую интерпретацию: а) высокая вероятность того, что рассматриваемый объект — единый пик; б) на основе экспертной (субъективной) оценки очевидна локализация как вершины пика (11-й отсчет), так и его границ (4-й и 27-й отсчеты соответственно). Следовательно, многократно упростился расчет амплитуды и временного положения пика, а также его площади по любому из двух предложенных алгоритмов.

### КОМПЕНСАЦИЯ ЛИНЕЙНЫХ ТРЕНДОВ ПО КЕНДАЛЛУ—СТЬЮАРТУ И ФИЛЬТРАЦИЯ

Компенсация ЛТ1 по Кендаллу—Стьюарту [9] представляет обработку по алгоритму  $y[n] = (x[n-1] + x[n] + x[n+1])/3$ , где  $x[n]$  — отсчеты исходного сигнала,  $y[n]$  — отсчеты обработанного сигнала, и позволяет решить три задачи:

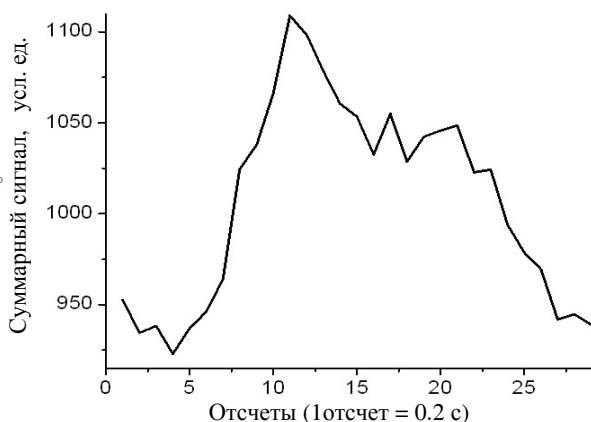
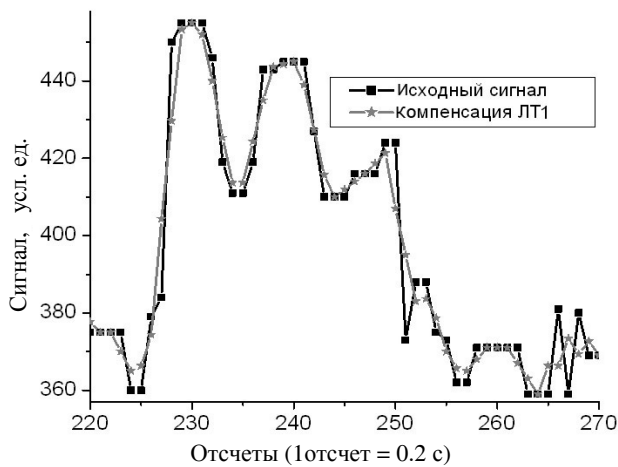
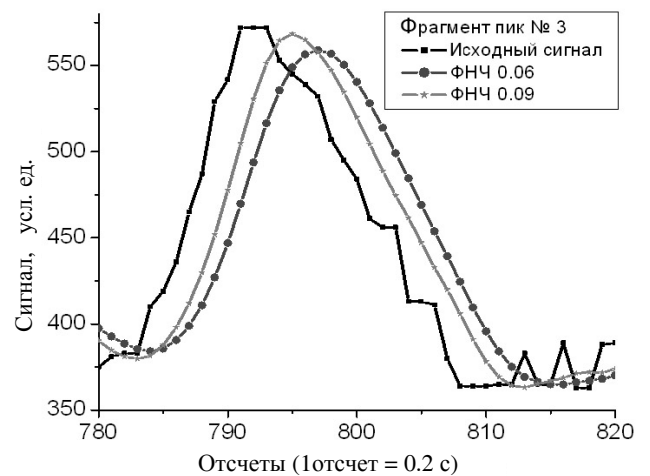


Рис. 5. Суммирование каналов 1–6 по методу главных компонент. Суммарный сигнал (главная компонента) пика № 4



**Рис. 6.** Эффект компенсации ЛТ1 по Кендаллу—Стьюарту на примере фрагмента электрофореграммы



**Рис. 7.** Преобразование исходного пика № 3 фильтром нижних частот Баттерворта

а) компенсировать линейно изменяющуюся составляющую ЛТ1; б) реализовать фильтр нижних частот; в) осуществить частичную нормализацию случайной составляющей сигнала (т. к. по центрально предельной теореме теории вероятности равновзвешенная сумма случайных величин асимптотически стремится к распределению Гаусса).

Разумеется, трех слагаемых может быть недостаточно для достижения существенного эффекта нормализации. Кроме того, на практике обработанный сигнал  $y[n]$  формируется не отсчетами с  $(n - 1)$  по  $(n + 1)$ , а с  $(n - 2)$  по  $n$ -й, исходя из принципа каузальности (т. е. невозможности использования будущего значения). Т. о., методически вносится временной сдвиг на 1 такт.

Поскольку линейные составляющие дрейфа базовой линии имеют малую значимость, априорно представляется влияние указанного алгоритма обработки незначительным. Это предположение подтверждено экспериментами, а незначительное влияние компенсации ЛТ на сигнал иллюстрируется на рис. 6.

Применение фильтра нижних частот (ФНЧ) Баттерворта (тангенсного, второго порядка) обусловлено простотой реализации и приемлемой эффективностью фильтрации высокочастотного шума. Его дискретная передаточная характеристика есть  $W(z) = \frac{k_n(1+z^{-1})^2}{1-\alpha_1 z^{-1} + \alpha_2 z^{-2}}$ . Коэффициенты

$\alpha_{1,2}$  определяются частотой среза; масштабный множитель  $k_n$  связан с этими коэффициентами, исходя из соображений полного пропуска сигнала с нулевой частотой (чему соответствует  $z = 1$ ). Тогда  $4k_n = 1 - \alpha_1 + \alpha_2$ .

Рассмотрены два варианта фильтрации с частотами среза 0.06 и 0.09 от частоты дискретизации по времени. Ранее [4, 6] нами отмечалось, что основными эффектами будут: а) сглаживание фронтов пика и значимое уменьшение дисперсии шума; б) сдвиг вершины пика и его расширение, что может привести к увеличению оценки площади пика. Указанные эффекты усиливаются по мере снижения частоты среза.

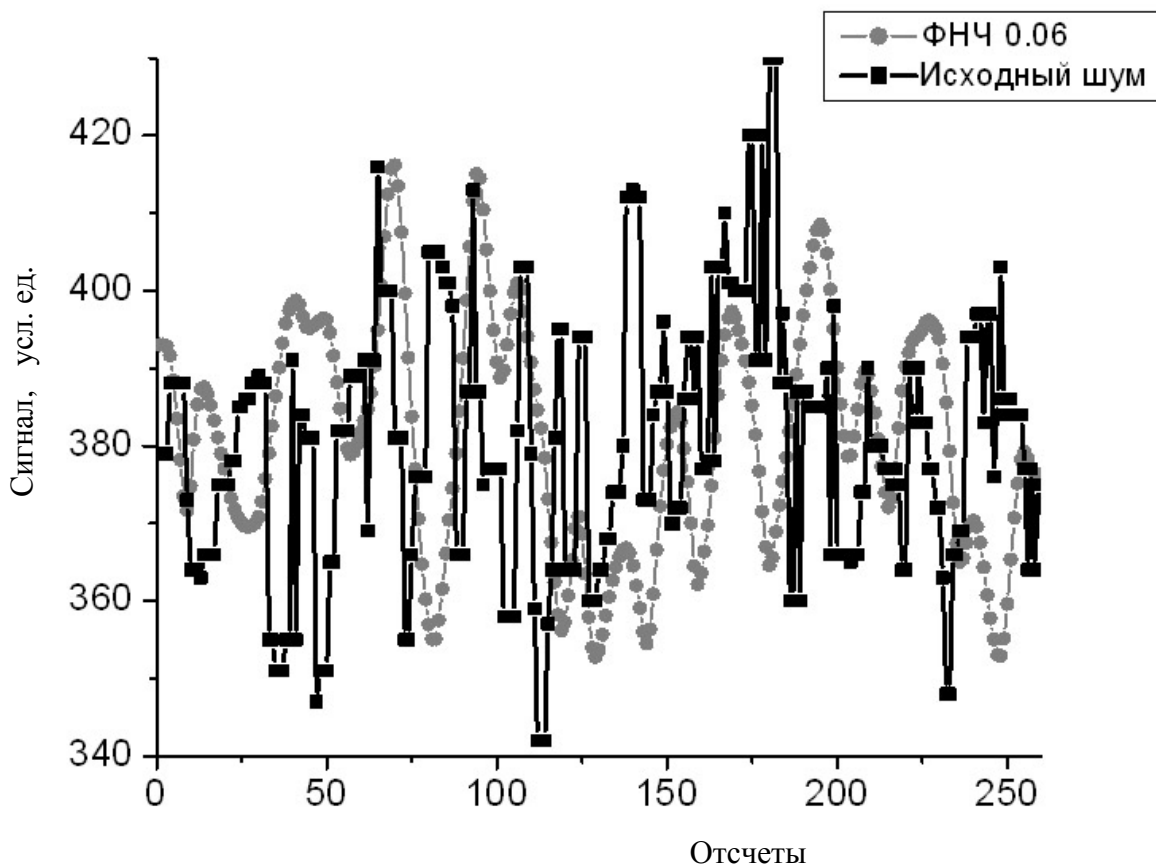
Возможна количественная оценка снижения амплитуды пика при упрощающем предположении, что передний и задний фронты являются идеальными ЛТ1. Тогда исходный (входной) сигнал  $x[k] = k$ , т. е. нормированный идеальный ЛТ1 преобразуется при фильтрации к виду:

$$y[k] = k + \frac{1 - \alpha_2}{2(1 - \alpha_1 + \alpha_2)} (\alpha_2^{k/2} \cos(\psi k) - 1) \quad \text{при}$$

$\psi = \arctg(\sqrt{4\alpha_2 / \alpha_1^2 - 1})$ . Нетрудно обосновать этой зависимостью эффекты более сильного смещения вершины пика при уменьшении частоты среза фильтра. Описанные выше эффекты приведены на рис. 7.

Степень уменьшения дисперсии высокочастотного шума зависит от частоты среза, т. е. от параметров фильтра, и вычисляется как  $k = \frac{k_n(3 + \alpha_1 - \alpha_2)}{2(1 - \alpha_2)}$ . При частоте среза фильтра 0.06

величина  $k$  примерно 0.146; т. о., стандартное отклонение шумовой составляющей около 6.5 квантов (при 17 для исходного сигнала). Влияние цифровой фильтрации на шумовую составляющую сигнала показано на рис. 8.



**Рис. 8.** Влияния цифровой фильтрации на неинформативную (шумовую) составляющую сигнала

Применение ФНЧ приводит к эффекту увеличения площади для всех пиков. Однако степень увеличения различна — от 4 до 46 %. Наиболее сильной трансформации подвергаются узкие пики с малой площадью. Кроме того, традиционная интегральная оценка площади более устойчива к последствиям фильтрации. По-видимому, значимость точной оценки высоты пика в этом случае не столь велика, как при применении модифицированной формулы (Площадь<sub>1</sub>). Т. е. при фильтрации эффект уширения пика более значителен, чем уменьшение его амплитуды (высоты).

### ВЫВОДЫ

1. Неинформативные составляющие сигнала — дрейф базовой линии и высокочастотная шумовая помеха обладают следующими свойствами.

а) Наклон базовой линии в большинстве случаев очень мал, и вклад линейно изменяющихся и нелинейных слагаемых весьма мал. Тем самым применительно к отдельным пикам базовую

линию можно считать "недрейфующей".

б) Высокочастотная (шумовая) составляющая сигнала имеет достаточно большое среднеквадратичное отклонение (в условиях измерения 17–22 кванта). В качестве закона распределения помехи по критерию согласия Пирсона могут быть приняты как нормальный (гауссовый) закон с доверительной вероятностью 12 %, так и треугольный закон Симпсона с вероятностью 57 %. Однако оба этих закона принципиально по-разному трактуют возможные источники помехи: в первом случае предполагается множественность аддитивных помех, во втором — доминирование равномерной помехи и построение ее первой разности.

2. Предложенная методика приближенной оценки площади пика (Площадь<sub>1</sub>), основанная на вычислении высоты пика и его ширины на полувысоте, в целом показала свою состоятельность и большую устойчивость по сравнению с традиционным интегрированием под огибающей пика.

3. Экономичные алгоритмы обработки информативных сигналов — компенсация ЛТ1 по Кендаллу—Стьюарту и фильтрация (ФНЧ Баттервор-



та 2-го порядка с относительными частотами среза (0.05–0.10) оказались достаточно эффективными, обеспечивая существенное уменьшение амплитуды высокочастотного шума без существенной неконтролируемой деформации пика. В приближении линейных фронтов эффекты деформации могут быть рассчитаны аналитически.

4. В случае слабой выраженности пика и, как следствие, неясной интерпретации его границ эффективен метод главных компонент, использующий в качестве исходных данных измерения на различных каналах. Достоинством в этом случае является отсутствие необходимости решения полной проблемы собственных чисел для корреляционной матрицы, т. к. требуется нахождение лишь первого (максимального) вещественного собственного числа.

#### **Благодарности**

Авторы благодарят участников семинара "Микрочиповые технологии в аналитической химии", 23 марта 2007 г. (Санкт-Петербург), за сделанные в ходе обсуждения доклада критические замечания и рекомендации.

#### **СПИСОК ЛИТЕРАТУРЫ**

1. *Евстратов А.А. и др.* Микрофлюидные аналитические системы на основе методов капиллярного электрофореза и микрочиповых технологий // *Аллергия, астма и клиническая иммунология*. 2003. Т. 7, № 9. С. 205–211.
2. *Евстратов А.А. и др.* Экспресс-анализ олигонуклеотидов на планарном микрофлюидном чипе // *Журнал аналитической химии*. 2004. Т. 59, № 6. С. 587–594.
3. *Евстратов А.А., Рудницкая Г.Е., Петухова Н.А.* Микрочиповые технологии в биологических исследованиях. Экспресс-анализ ДНК: сепарационные матрицы для разделения ДНК // *Научное приборостроение*. 2005. Т. 15, № 2. С. 27–40.
4. *Евстратов А.А. и др.* Особенности применения алгоритмов цифровой фильтрации электрофореграмм при анализе веществ на микрочипе // *Научное приборостроение*. 2003. Т. 13, № 2. С. 57–63.
5. *Bulianitsa A.L. et al.* Estimation of data reduction effects on informative parameters of analytical signals, their consideration and compensation // *Abstr. program. "100 Years of Chromatography" 3<sup>rd</sup> Int. Symp. on Separations in BioSciences (SBS), 13–18 May 2003, Moscow, Russia*. P. 184.
6. *Буляница А.Л. и др.* Оптимизация алгоритмов обработки электрофореграмм при анализе биологических веществ на микрочипе // *Каталог рефератов и статей: Межд. форум "Аналитика и аналитики", Воронеж, Россия, 2–6 июня 2003 г.* Т. II. С. 592.
7. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики. М.: Наука, 1983. 416 с.
8. *Буляница А.Л., Курочкин В.Е., Кноп И.С.* Методы статистической обработки экологической информации: дискриминантный, корреляционный и регрессионный анализ. СПб.: СПГУАП-ИАНП РАН, 2005. 48 с.
9. *Кендалл М.Дж., Стьюарт А.* Многомерный статистический анализ и временные ряды. М.: Наука, 1976. 736 с.

*Институт аналитического приборостроения РАН, Санкт-Петербург*

Материал поступил в редакцию 7.05.2007.

## **INFORMATIVE SIGNALS IN MICROCHIP-BASED ELECTROPHORESIS: MATHEMATICAL MODELS AND ESTIMATES**

**A. L. Bulianitsa, G. E. Rudnitskaja, A. A. Evstrapov**

*Institute for Analytical Instrumentation RAS, Saint-Petersburg*

The paper considers possible lines of developing algorithms for processing and estimating analytical signals of microchip electrophoretic devices: simulation of informative signals (peaks) and non-informative components (base line drift and high-frequency noise). Based on the information obtained, sources of interference can be revealed, the optimal algorithms for the signal estimation have been found, and criteria for neighboring peaks separation have been modified. Efficiency of conventional economical methods for signal pre-processing, namely, trend compensation and a filtration, as well as validity of the new technique for peak area estimation and applicability of the main component method in multi-channel measurements when the peak boundaries can hardly be localized, have been shown for electrofореграмм obtained by separating the DNA fragment mix on a microfluidic chip.