

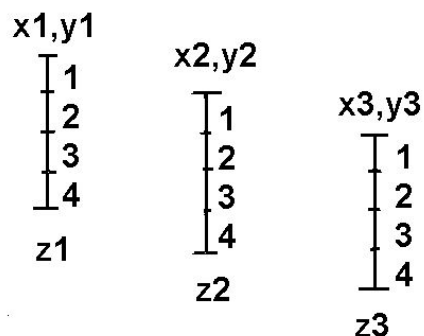
*Раздел I*

**НАЧАЛЬНЫЕ ПОНЯТИЯ  
И ПРЕДПОСЫЛКИ**

*Лекции 4-6*

# Лекция 4. ПОКАЗАТЕЛИ ПРОИЗВОДИТЕЛЬНОСТИ ПАРАЛЛЕЛЬНЫХ СИСТЕМ

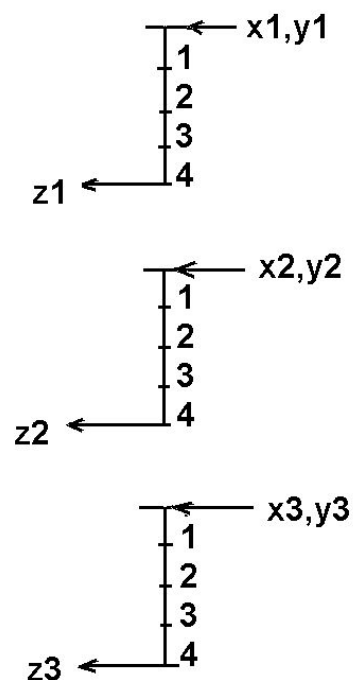
## КОНВЕЙЕРНАЯ



1 ТАКТ / РЕЗУЛЬТАТ

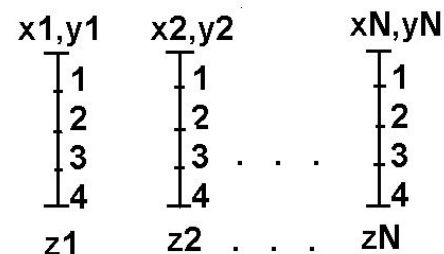
**STAR - 100**  
**CRAY - 1**  
**CYBER - 205**

## ПОСЛЕДОВАТЕЛЬНАЯ



4 ТАКТА / РЕЗУЛЬТАТ

## МАТРИЧНАЯ



4 ТАКТА / N РЕЗУЛЬТАТОВ

**ILLIAC IV**  
**ICL DAP**  
**STARAN**

# АБСТРАКТНЫЕ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ

- **Время выполнения векторной арифметической операции**

$$t = b + cn.$$

Или

$$t = c \left( n + \frac{b}{c} \right) = r_{\infty}^{-1} (n + n_{1/2})$$

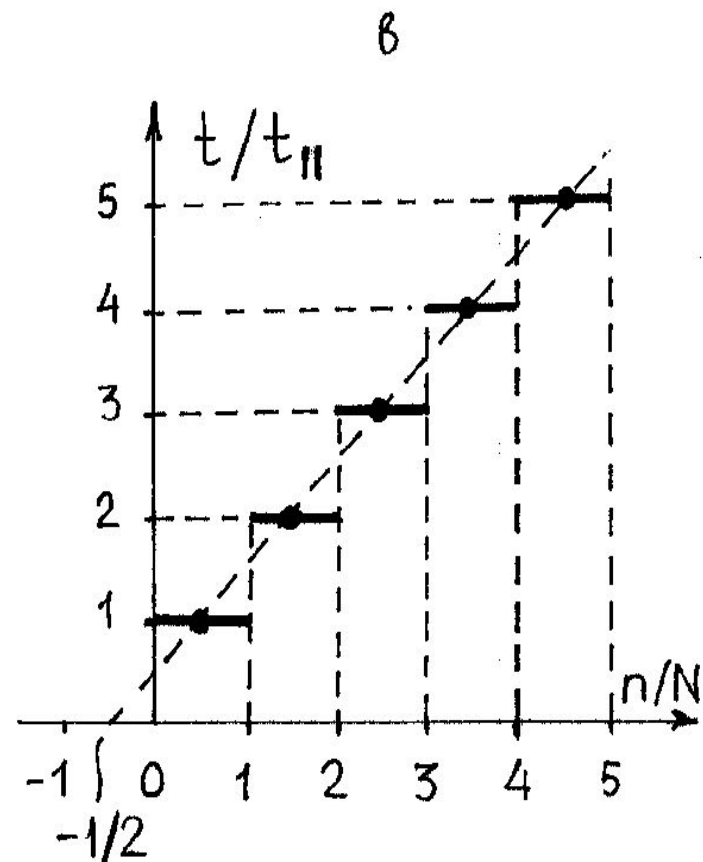
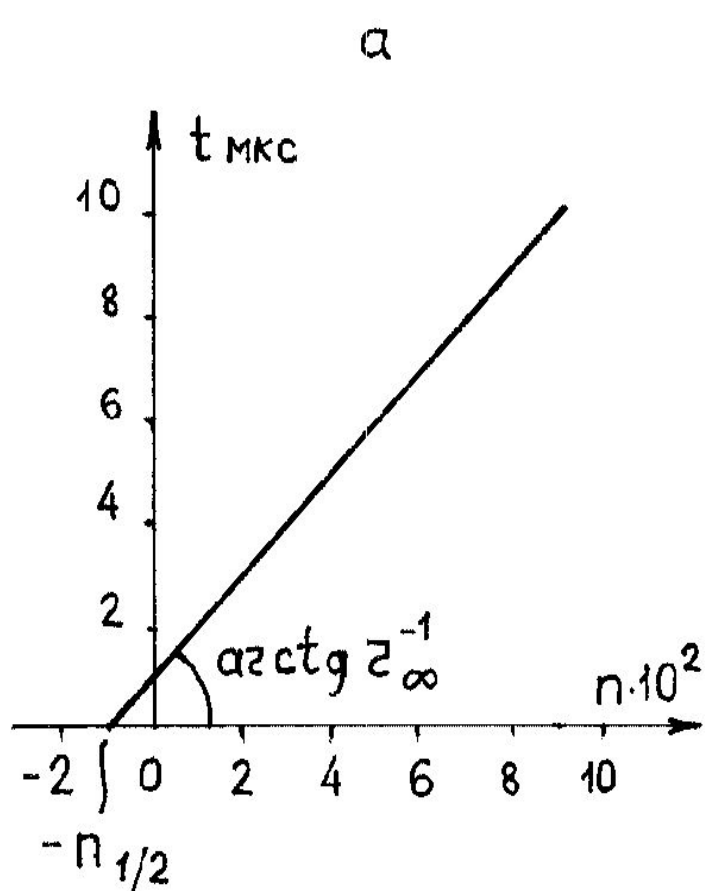
$$r_{\infty} = \frac{1}{c} = \lim_{n \rightarrow \infty} \frac{n}{t}$$

– **пиковая**, или асимптотическая **производительность**. Измеряется числом эквивалентных скалярных операций в секунду (**MFLOPS**).

$$n_{1/2} = b/c$$

– **длина полупроизводительности**, т. е. длина вектора, при которой достигается половина пиковой производительности.

# ЭКСПЕРИМЕНТАЛЬНОЕ ОПРЕДЕЛЕНИЕ ПАРАМЕТРОВ



## ОЦЕНКИ ВЕКТОРНОЙ ПРОИЗВОДИТЕЛЬНОСТИ

КОМПЬЮТЕР	$n_{1/2}$	$r_{\infty}$ MFLOPS
64' CRAY-1	10–20	80
48' BSP	25–150	50
2-конв. CDC CYBER 205	100	100
64' CDC STAR 100	150	25
32' (64x64) ICL DAP	2048	16



# СИСТЕМНАЯ ПРОИЗВОДИТЕЛЬНОСТЬ

## ОСНОВНЫЕ НЕДОСТАТКИ АБСТРАКТНОЙ ОЦЕНКИ:

### 1. Игнорирование времени на выполнение скалярных операций– организацию циклов, переходов и др.

Ускорение  $S$  для  $N$  процессоров (закон Амдала)

$$S \leq \frac{1}{f + \frac{1-f}{N}}$$

$f$  – доля трудозатрат на выполнение скалярных операций. Для получения заданного  $S$ , требуется разработка специальных параллельных алгоритмов с  $f < 1/S$ .

### 2. Игнорирование потерь на маршрутизацию.

Время маршрутизации (пересылки данных между процессорами) с ростом  $N$  может доминировать.

$$x_j = \sum_{k=1}^j d_k, \quad j = \overline{1, n}$$

**Пример.** Вычисление последовательных сумм можно реализовать как  $(n - 1)$  сложений:  $x_j = x_{j-1} + d_j, x_0 = 0$ .



На  $l$ -уровне выполняется сдвиг на  $2^l$  позиций. Уровень  $l = \log_2 n$  дает искомую сумму в крайнем правом аккумуляторе. Так что число операций суммирования составляет теперь уже

$$\log_2 n < n - 1.$$

В случае связи “к ближайшему соседу” на уровне  $l$  потребуется  $2^{l-1}$  единичных операций маршрутизации, что дает в целом

$$1 + 2 + \dots + n/2 = n - 1$$

таких операций со степенью параллелизма ‘ $n$ ’.

Пусть  $r$  – отношение времени выполнения одной операции суммирования ко времени единичной операции маршрутизации. Тогда отношение суммарных времен равно

$$r \log_2 n / (n-1),$$

т.е. с ростом ‘ $n$ ’ влияние маршрутизации растет.

При переходе от рекурсии к каскадному суммированию число эквивалентных скалярных арифметических операций растет с  $(n - 1)$  до  $n \log_2 n$ .

Поэтому **эмуляция параллельного алгоритма на последовательной ЭВМ будет всегда малоэффективной.**



# ОТНОСИТЕЛЬНАЯ ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ

- Параллельный процессор – **внешний акселератор Host ЭВМ.**
- **Длительности тактов** обеих компонент комплекса **одинаковы.**

Для каждого элемента множества **W** представительных программ находятся модельные оценки производительности **главной ЭВМ и комплекса в целом.** *Усредненный коэффициент ускорения S на множестве W дает искомую оценку.*

*Условие равенства длительности тактов делает такую оценку технологически независимой.* Поэтому она может быть использована *для сравнения различных систем одного предметного направления.*

**$D > 1$**  – коэффициент роста аппаратных затрат комплекса в сравнении с Host ЭВМ. Его **эффективность**

$$H = S / D.$$

Чем она выше, тем перспективнее система для данной предметной области.

# ТЕСТОВЫЕ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ

**Эти оценки наиболее объективны.** Разработкой и стандартизацией тестов оценки производительности современных компьютеров занимается ряд корпораций и комитетов (советов): **SPEC, LINPAC, AIM, TPC** и др.

Специальные тесты для определения производительности:

- только **процессора** (пример **LINPACK**),
- только **файловой системы** (пример **Bonnie**),
- только **коммутационной сети** (пример **MPI-тестов**),
- **комбинированные** тесты (пример **AIM**).

По результатам **LINPAC**-тестирования ведется список **TOP-500**, в котором ежегодно ранжируются **500** наиболее производительных в мире систем.

Тесты **TPC A, B, C, D, E** (**TPC – Transaction Processing Performance Council** – Совет по производительности обработки транзакций, основан в 1988г.) разработаны специально для оценки производительности **СУБД**.



# Лекции 5-6. ПРЕДМЕТНЫЕ ПРЕДПОСЫЛКИ ПАРАЛЛЕЛИЗМА

## ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ

**Обработка сигналов** – преобразование формы или частотного спектра электрических, речевых и видео сигналов.

**Электрические и речевые сигналы** – некоторые временные функции  $f(t)$ . Время  $t$  и значения  $f(t)$  непрерывны – аналоговые величины.

**Сигналы изображения** – функции  $f(X,Y)$  или  $f(X,Y,Z)$  изменения цветности, яркости и др. в 2- или 3-мерной системе координат.

**Цифровая обработка сигналов (ЦОС)** широко применяется при передаче речевых сигналов и сигналов изображений, при распознавании и синтезе речи, используется в медицине, метеорологии, сейсмологии и др.

Исходные аналоговые сигналы путем **дискретизации** во времени и **квантования** по амплитуде преобразуются в последовательность цифровых данных. Далее она подвергается обработке, основанной на

- **дискретном преобразовании Фурье (ДПФ),**
- **быстром преобразовании Фурье (БПФ),**
- **преобразовании Адамара и др.,**
- **преобразованиях частотного спектра (цифровая фильтрация).**

*Дискретизация во времени* – представление  $f(t)$  как последовательности значений  $f(nT)$  в моменты времени, кратные некоторому интервалу  $T$ . Если частотный спектр  $f(t)$  ограничен значением  $F$ , то  $T = 1/(2F)$  – теорема Котельникова.

*Дискретизация в пространстве* – вопрос более сложный.

*Квантование* – двоичное представление (разрядностью  $n$  бит) выборочных аналоговых значений одним из  $2^n$  уровней с максимальной ошибкой квантования  $1/2^n$ .

При обработке *эл. сигналов*  $n = 12 \dots 18$ ,  
*речи* –  $n = 8 \dots 14$ ,  
*изображений* –  $n = 4 \dots 9$ .



## ДИСКРЕТНОЕ И БЫСТРОЕ ПРЕОБРАЗОВАНИЯ ФУРЬЕ

- **ДПФ** комплекс. последовательности  $x_n, n=1 \dots N$   
Число операций –  $N^2$  парных операций умножение-сложение комплексных чисел
- **БПФ**. Понятие *расщепления*  
Достижимое ускорение –  $S = 2$  ( $N^2/2$  умножений-сложений)
- **Алгоритм Кули-Тьюки** – многократное применение алгоритма расщепления:  
Фазы *упорядочения* (последовательные разбиения) и *перемешивания*  
Фаза *перемешивания*: число этапов –  $m = \log_2 N$   
число операций –  $(N/2) \log_2 N$  компл. умножений,  
 $+ N \log_2 N$  компл. сложений

Дискретным преобразованием Фурье комплексной последовательности  $x_n$ ,  $n \in \overline{1, N}$ , называется преобразование

$$X_k = \sum_{n=1}^N x_n \exp(-ikn \frac{2\pi}{N}). \quad (1)$$

Ему обратным является преобразование

$$x_n = \frac{1}{N} \sum_{k=1}^N X_k \exp(ikn \frac{2\pi}{N}).$$

Величины  $X_k$  и  $x_n$  – это элементы упорядоченных множеств  $\{ X_k \}$  и  $\{ x_n \}$  одинаковой мощности  $N$ .

В основе всех алгоритмов БПФ лежит понятие *расщепления*. Пусть  $N$  четно. Разбиваем последовательность  $x_n$  на две  $N/2$ -точечные последовательности  $y_n$  и  $z_n$ :

$$y_n = x_{2n}, \quad z_n = x_{2n-1}, \quad n = 1, 2, \dots, N/2. \quad (2)$$

Заменяя в определении изображения  $N$  на  $N/2$  и  $\exp(i2\pi/N)$  на  $w$ , получаем:

$$Y_k = \sum_{n=1}^{N/2} y_n w^{-2kn}, \quad Z_k = \sum_{n=1}^{N/2} z_n w^{-2kn}, \quad k = 1, 2, \dots, N/2. \quad (3)$$

С учетом (3), окончательно

$$X_k = Y_k + w^k Z_k, \quad k = 1, 2, \dots, N/2. \quad (4)$$

$$X_{k+N/2} = Y_k - w^k Z_k. \quad (5)$$

Таким образом, *алгоритм расщепления* состоит в следующем.

1. Согласно (2), определяем последовательности  $Y_n$  и  $Z_n$ .
2. По формулам (3) вычисляем преобразования  $Y_k$  и  $Z_k$  этих последовательностей.
3. По формулам (4) и (5) вычисляем преобразования  $X_k$  исходной последовательности  $X_n$ .

Аналогичная процедура применима и к преобразованиям  $Y_k$  и  $Z_k$ . Если  $N$  кратно 4, то каждое из них может быть представлено двумя  $N/4$ -точечными преобразованиями. Если же  $N = 2^m$  для некоторого целого  $m$  (что обычно имеет место), то процедура может быть повторена  $m$  раз, пока не будет получено  $N$  преобразований длины 1.

## ИЛЛЮСТРАЦИЯ АЛГОРИТМА КУЛИ-ТЬЮКИ

Этапы преобразования	1	2	3	4	5	6	7
	–	–	–	0,07	–	–	–
Элементы преобразуемой	–	–	0,07	–	0,25	–	–
последовательности	–	–	0,32	0,32	0,39	–	–
	–	0,07	–	–	–	$-0,62+0,25i$	–
	0,07	0,91	–	0,91	–	$0,81+0,00i$	$-0,05-0,10i$
	0,40	0,32	0,91	–	$-0,62$	$-0,62-0,25i$	$-0,03+0,81i$
	0,91	0,29	0,29	0,29	1,20	$1,59+0,00i$	$1,19-0,42i$
	0,18	–	–	–	–	–	$0,30+0,00i$
	0,32	–	–	–	–	–	$1,19+0,42i$
	0,56	0,40	0,40	0,40	0,18	$0,57+0,16i$	$-0,03-0,81i$
	0,29	0,18	0,56	–	0,96	$-0,03+0,00i$	$-0,05-0,10i$
	0,75	0,56	–	0,56	–	$0,57-0,16i$	$3,48+0,00i$
	–	0,75	–	–	–	$1,89+0,00i$	–
	–	–	0,18	0,18	0,57	–	–
	–	–	0,75	–	0,93	–	–
	–	–	–	0,75	–	–	–



## ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ БПФ

- **Спектральный анализ** – используется при исследовании биотоков мозга, кардиограмм, сейсмических колебаний и т.д.
- **Цифровая фильтрация изображений** – затрагивает все составляющие спектра. Подавляя “ВЧ составляющие”, можно реализовать операцию “сглаживания” мелких деталей изображения.
- **Вычисление свертки** двух сигналов – прямые вычисления требуют  $N^2$  парных операций умножение-сложение. **Ускорение процесса**: сначала выполнить прямое преобразование, а затем – обратное.
- **Вычисление функции корреляции двух сигналов**

Рассмотрим выполнение операций *свертки* двух сигналов и вычисления их *корреляционной функции*.

*Свертка* двух рядов дискретных значений  $f_1(n)$  и  $f_2(n)$  определяется как

$$f_c(n) = \sum_{m=1}^N f_1(m) f_2(n - m), n = 1, 2, \dots, N.$$

Непосредственные вычисления требуют  $N^2$  парных операций умножение-сложение. Ускорить процесс в данном случае можно, если сначала выполнить прямое преобразование

$$F_c(k) = F_1(k) F_2(k),$$

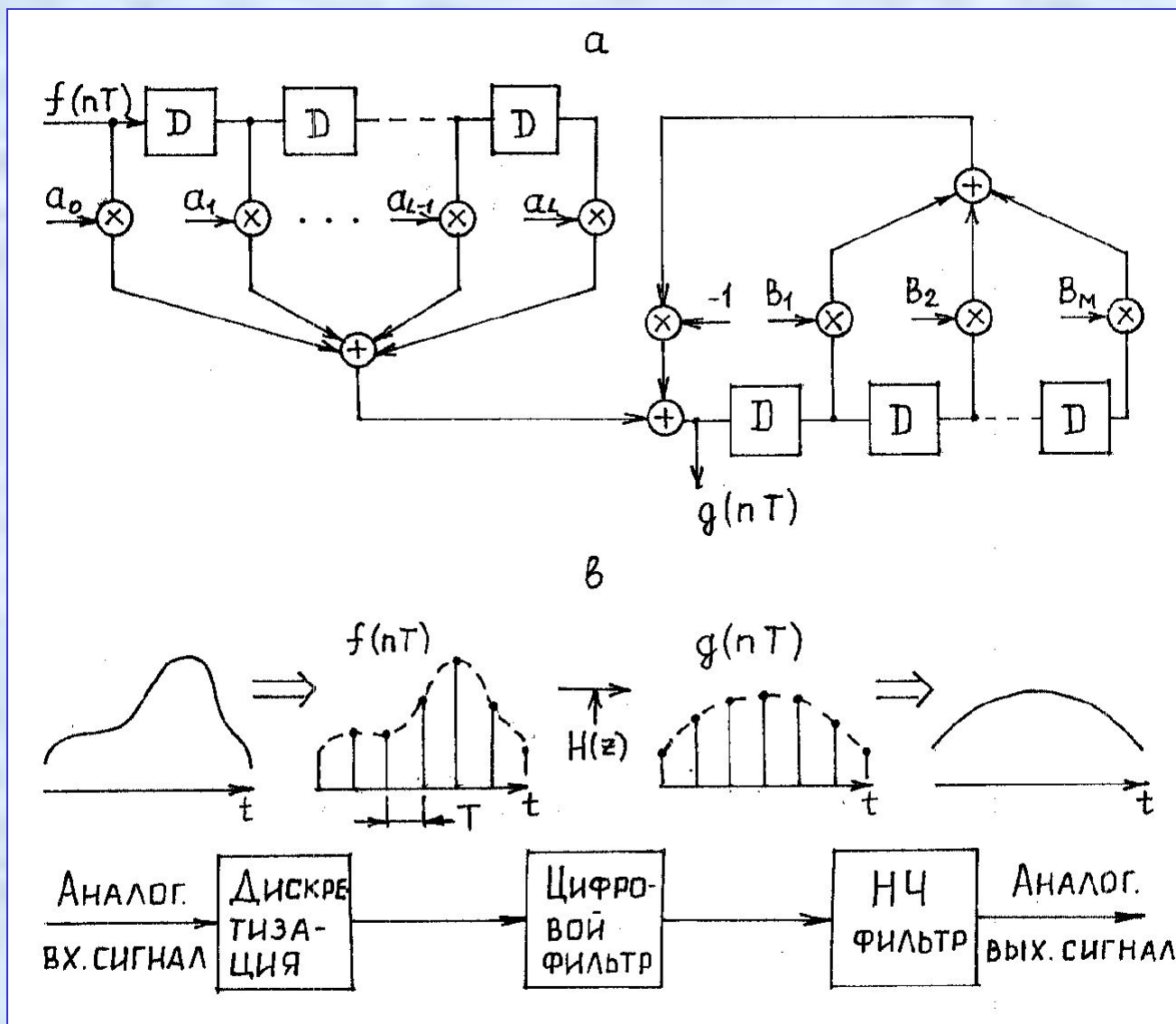
а затем – обратное.

*Функция корреляции*

$$f_{\text{кор}}(n) = \sum_{m=1}^N f_1(m) f_2(n + m)$$

(при  $f_1 = f_2$  функция называется *автокорреляционной*). Ее преобразование  $F_{\text{кор}}(k) = \bar{F}_1(k) F_2(k)$ , где  $\bar{F}_1(k)$  комплексно сопряжено с  $F_1(k)$ . Вычисление  $f_{\text{кор}}(n)$  выполняется аналогично  $f_c(n)$ .

# ЦИФРОВАЯ ФИЛЬТРАЦИЯ И УСТРОЙСТВА ЦОС



В этой схеме входная последовательность  $f(nT)$ ,  $n=0,1,2,\dots$ , с помощью цепей  $D$  временной задержки на период  $T$  и схем умножения преобразуется в последовательность  $g(nT)$

$$g(nT) = \sum_{k=0}^L a_k f[(n-1)T] - \sum_{m=1}^M b_m g[(n-m)T],$$

где  $a_k$  и  $b_m$  – вещественные постоянные,  $L$  и  $M$  – число ступеней задержки (порядок фильтра).

**Синтез таких фильтров** выполняется с использованием  $Z$ -преобразования

$$Z\{f(nT)\} = F(z) = \sum_{n=0}^{\infty} f(nT) z^{-n},$$

где  $z$  – комплексное переменное. На основании формулы для  $g(nT)$  имеем

$$G(z) = H(z)F(z), \quad H(z) = \frac{\sum_{k=0}^L a_k z^{-k}}{1 + \sum_{m=1}^M b_m z^{-m}}.$$

Функция  $H(z)$  называется **передаточной функцией** цифрового фильтра. В определении этой функции и состоит основная задача его синтеза.

Теория цифровых фильтров в настоящее время интенсивно развивается. Процесс фильтрации иллюстрирует рис. 6. **Полоса пропускания** оконечного **НЧ фильтра** выбирается равной  $1/(2T)$ . Для фильтров высоких порядков при работе в реальном масштабе времени вычисления по формуле  $g(nT)$ , как и процедуру БПФ, целесообразно выполнять на параллельных системах.



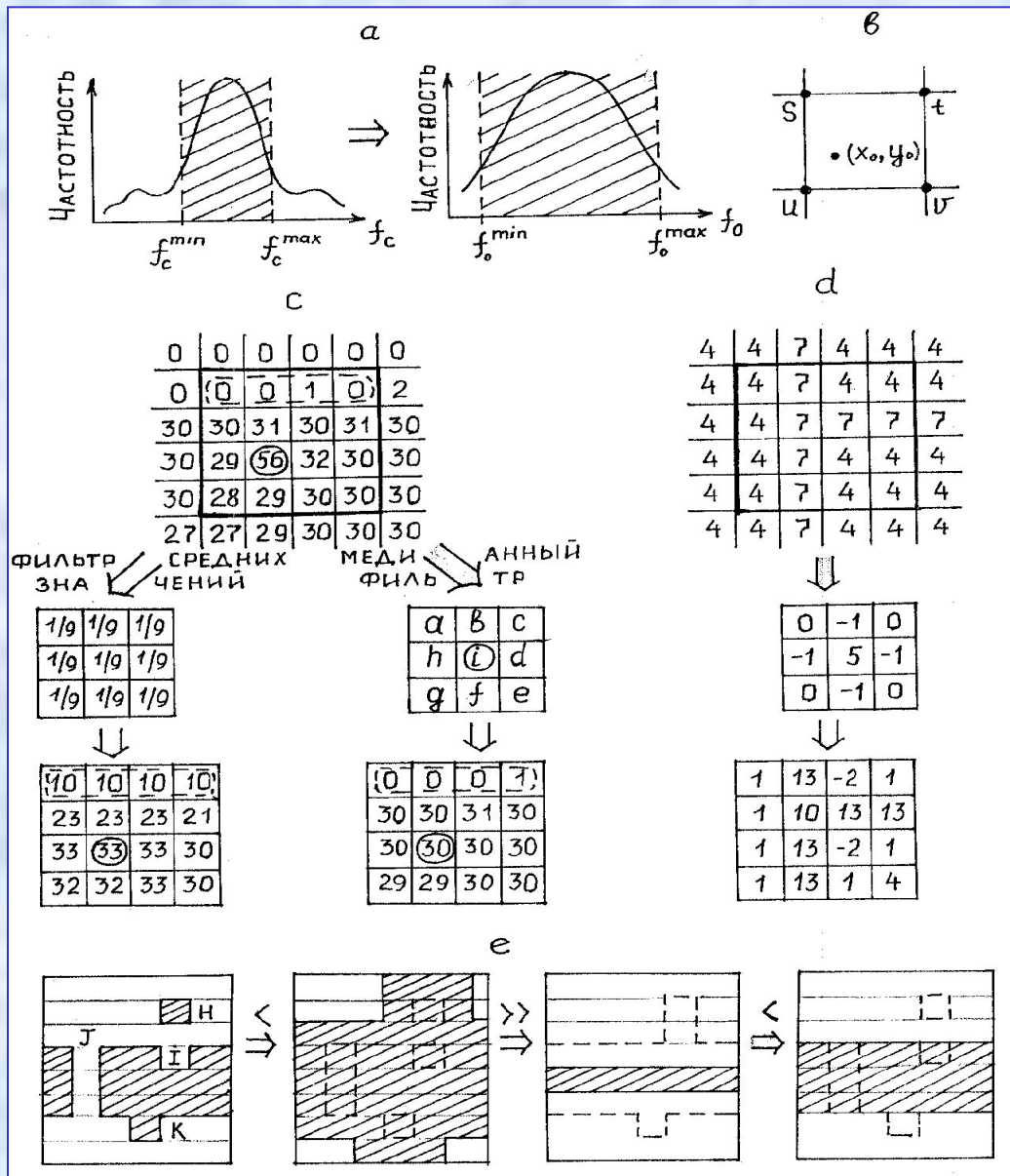
# ОБРАБОТКА ИЗОБРАЖЕНИЙ

- **Обработка изображений** – это выполнение различных операций над **многомерными сигналами**: телевизионные изображения, чертежи и рисунки, фотографии разведывательного характера, медицинские рентгенограммы, электронно-микроскопические фотографии молекул, радио- и звуколокационные карты, диаграммы сейсмических данных и др.
- **Основные виды обработки** – улучшение изображений, их эффективное кодирование, распознавание образов, машинная графика.
- **Области применения** – медицина, дистанционное зондирование, идентификация личности, промышленные измерения, информационная служба и т.д.
- **Данные изображения – пиксели** – элементы двумерного массива из  $m$  столбцов и  $n$  строк – **бинарные** (2 градации), **многоградационные** (например, 256 градаций) или **многоградационно-векторные** (256 градаций по каждой из составляющих – **красной**, **зеленой** и **синей**). Соответственно изображение – **бинарное**, **полутонное** или **спектральное**.
- $m, n$  – до  $10^7$  и более).
- **Одинаковые операции выполняются параллельно** по всему изображению, что адекватно использованию **процессорных матриц**. Применяют и специальные графические приставки к ПК.

# ОПРЕДЕЛЕНИЯ РАЗЛИЧНЫХ ВИДОВ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

- **Улучшение (реставрация) изображений** – *компенсация искажений*, вносимых при их формировании системами отображения.
- **Кодирование изображений** – *сокращение числа битов представления изображений*, при условии достоверности их воспроизведения. Сначала – преобразование изображения. Затем – кодирование результата преобразования.
- **Распознавание образов** – это и распознавание знаков, и средство медицинской диагностики, и составление карт земных ресурсов на основе фотографий, полученных со спутников (*дистанционное зондирование*), и др.
- **Машинная графика** – *ввод графической информации (чертежей и рисунков) в ЭВМ, ее обработка и вывод*. Основная задача такой обработки – *синтез и представление изображения*. Области применения: компьютерная мультипликация, машинное проектирование логических схем, выполнение дизайнерских проектов и др.

# УЛУЧШЕНИЕ ИЗОБРАЖЕНИЙ





**Улучшение контрастности** изображения дает градационное преобразование

$$f_0 = A f_c + B,$$

где  $f_c$  и  $f_0$  – значения видеоданных до и после преобразования (рис. а). Коэффициенты **A** и **B** определяют из условий перехода

$$f_c^{\max} \rightarrow f_0^{\max}, \quad f_c^{\min} \rightarrow f_0^{\min}.$$

**Коррекция геометрических искажений** (повороты или параллельные перемещения элементов) достигается преобразованием координат

$$\begin{vmatrix} x_0 \\ y_0 \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} x_c \\ y_c \end{vmatrix}$$

Значения параметров **a**, **b**, **c**, **d** определяются по степени искажений некоторых хорошо известных элементов. Процедура такова:

$$(x_c, y_c) \rightarrow (x_0, y_0); \quad f_0(x_0, y_0) := f_c(x_c, y_c).$$

Дробные координаты  $(x_0, y_0)$  округляют до ближайшей овокупности (рис. b), принимая

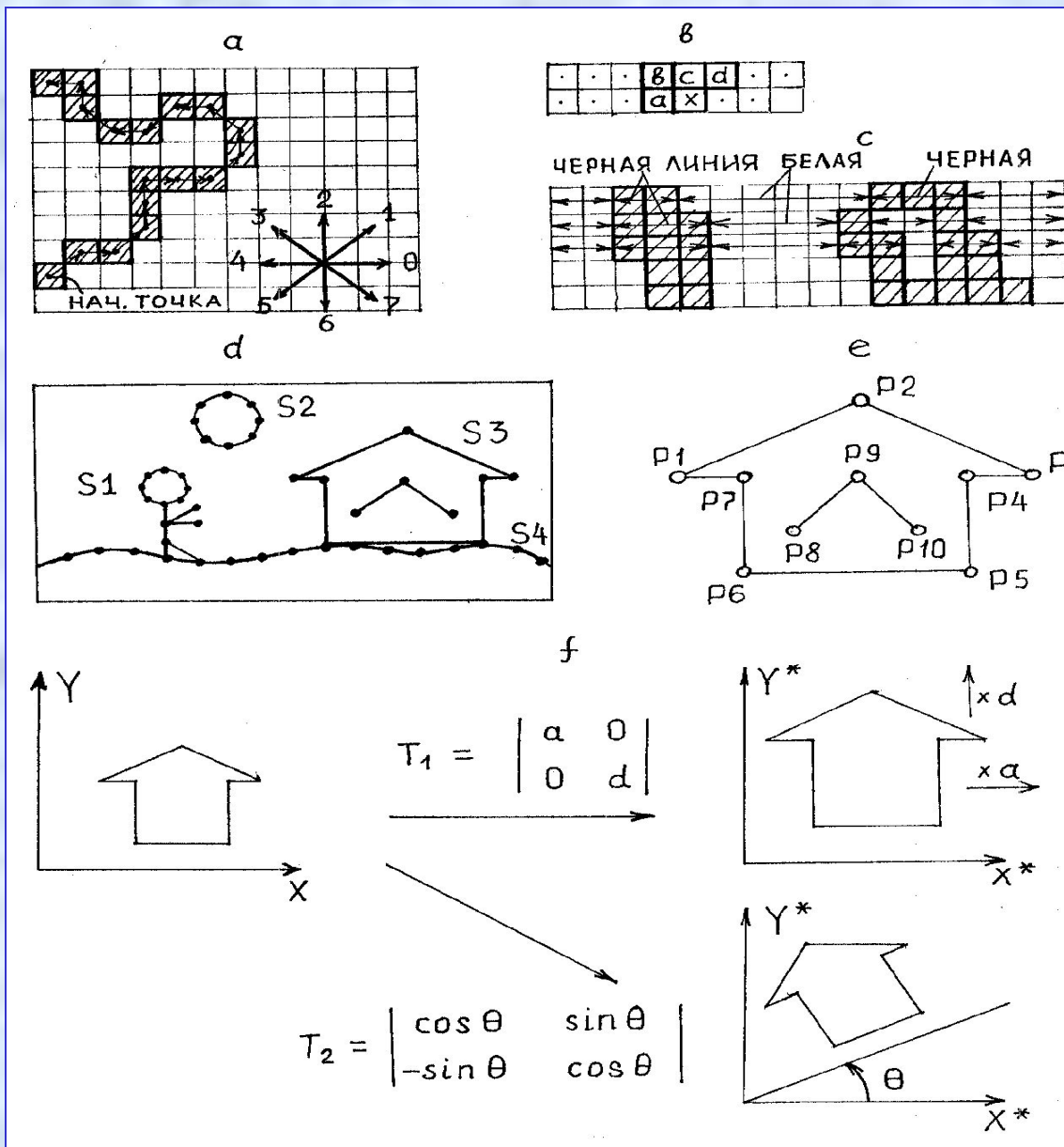
$$f_0(u) := f_0(x_0, y_0).$$

**Типичные шумы изображения** – **зернистый шум** и **пятна** на **полутонном изображении**, **отдельные шумы** и **обрывы линий** на **бинарном изображении**. Обычно эти шумы могут быть устранены проведением для **каждого элемента изображения** локальной фильтрации на окружающем его участке  $3 \times 3$  элементов, центром которого он является (рис. c, d, e).

**Для эффективной обработки изображений целесообразно применение процессорных матриц со связью между 8 соседями.**



# КОДИРОВАНИЕ ИЗОБРАЖЕНИЙ И ОБРАБОТКА ГРАФИКИ



# ОБРАБОТКА СИМВОЛОВ

- **Обработка символов** – связана с редактированием текстов, переводом с одного языка на другой, доказательством теорем, преобразованием математических формул, медицинской диагностикой и т.д. В целом – с созданием **искусственного интеллекта**.

## ОБРАБОТКА ЦЕПОЧЕК СИМВОЛОВ

- конкатенация (объединение нескольких цепочек),  
– сопоставление (сравнение двух цепочек),  
– замещение (замена одной цепочки на другую),  
– выборка (выборка части цепочки).
- **Конкатенация:**  $Z = XY$  либо  $Z = X \cdot Y$
- **Сопоставление:** (СРАВНИВАЕМАЯ ЦЕПОЧКА) (ОБРАЗЦОВАЯ ЦЕПОЧКА) В наилучшем случае –  $n(m-n)$  сравнений,  $m$  и  $n$  ( $< m$ ) – длины сравниваемой и образцовой цепочек.

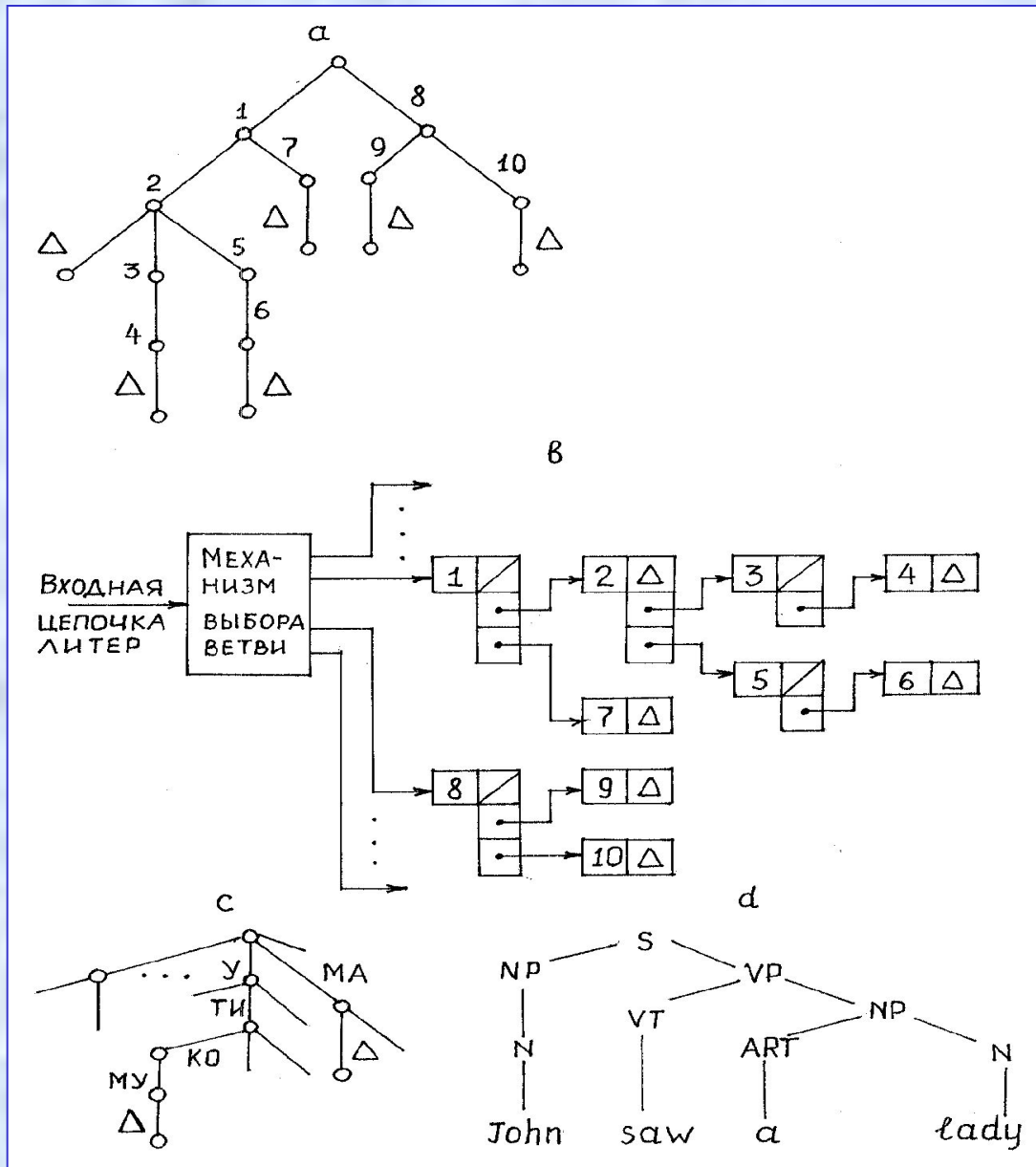
**Алгоритм КМП** – развит **Кнудом, Моррисом и Праттом**.

- **Замещение:**  $ZY = \text{'mosq'}$ ;  $Z = \text{'knpt'}$   $\cdot$   $\text{'alsvi'}$ ,  $Y = \text{'alsvi'}$ .  
 $Z = \text{'knpt'}$   $\cdot$   $\text{'mosq'}$
- **Операцию сравнения последовательностей литер целесообразно распараллелить.**

## ОБРАБОТКА ЕСТЕСТВЕННЫХ ЯЗЫКОВ

- **ЕСТЕСТВЕННЫЙ ЯЗЫК** – используемый в повседневной жизни.
- **ВИДЫ ОБРАБОТКИ ЕЯ :**
  - обработка слов (поиск в словаре, обработка морфем);
  - обработка предложений (синтаксическая, семантическая);
  - обработка текстов (обработка контекста).
- **ТЕРМИНОЛОГИЯ:**
  - Слово** – последовательность букв.
  - Словарь** – все слова данного текста должны находиться в словаре для этого текста.
  - Предложение** – ряд нескольких слов.
  - Морфема** – наименьшая языковая единица: слово, префикс, суффикс.
- **ПОИСК В СЛОВАРЕ, ОРГАНИЗОВАННОМ КАК TRIE-ДЕРЕВО**
  - Пример:** {1 2, 1 2 3 4, 1 2 5 6, 1 7, 8 9, 8 10}
  - Механизм выбора последовательности узлов при поиске.**
  - Обработка морфем.**

# ОБРАБОТКА СЛОВ И СИНТАКСИЧЕСКАЯ ОБРАБОТКА





## *СИНТАКСИЧЕСКАЯ ОБРАБОТКА, или ГРАММАТИЧЕСКИЙ РАЗБОР*

- **СТРУКТУРА ПРЕДЛОЖЕНИЯ** определяет связи между объектами:
- **S** – предложение,
- **N** – существительное,
- **VT** – глагол,
- **ART** – артикль,
- **PREP** – предлог,
- **NP** – существительное или существительное с артиклем,
- **VP** – глагол и существительное (без или с артиклем) или существительное с артиклем и с предлогом,
- **PP** – существительное (без или с артиклем) с предлогом.
- **ПРАВИЛА CFG** (context-free grammar – контекстно-свободная грамматика) для структуры предложения в английском языке:
- **S NP VP;**
- **NP ART N** либо **NP N;**
- **VP VT NP** либо **VP VT PP;**
- **PP PREP NP.**
- Согласно этим правилам, для предложения **'John saw a lady'** получаем дерево грамматического разбора рис.d.