

ОПТИЧЕСКИЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 681.51 : 681.3

А. А. Акаев, А. А. Кутанов, С. З. Дордоев, Б. Д. Абдрисаев,
И. А. Снимщиков*(Бишкек, Кыргызстан)*ОПТИКО-ЭЛЕКТРОННАЯ НЕЙРОННАЯ СЕТЬ
НА ОСНОВЕ ГОЛОГРАФИЧЕСКОГО
ФОТОТЕРМОПЛАСТИЧЕСКОГО ДИСКА

Рассмотрены основные характеристики оптико-электронной нейронной сети на основе голографического фототермопластического диска. Показано, что использование матрицы фотоприемников в такой структуре при реализации операции суммирования электронным способом позволяет снизить влияние флуктуаций дифракционной эффективности фазовых фурье-голограмм на точность вычислений в сети и упростить оптическую схему. Исследованы возможности рассматриваемого устройства для обработки изображений большой размерности.

В последние годы наблюдается повышенный интерес к созданию интеллектуальных вычислительных систем для решения задач, свойственных мозгу человека. Появление модели Хопфилда для описания нейронной сети [1] стимулировало рост работ по созданию нейрокompьютеров. Первоначально были предложены электронные реализации нейрокompьютеров. Однако такие программно-аппаратные модели и нейрочипы по числу связей и быстродействию не могут удовлетворять требованиям задач искусственного интеллекта, так как их предельные параметры ограничены: 10^7 — 10^8 связей между нейронами с быстродействием 10^9 — 10^{10} переключений в секунду.

Оптика имеет свойства, привлекательные для моделирования нейронных ансамблей: естественный параллелизм, возможность создания связей в свободном пространстве и большой объем памяти запоминающих устройств. Наиболее привлекательными являются оптико-электронные нейронные сети, сочетающие в себе легкость создания глобальных оптических связей и реализацию нелинейных и логических операций с помощью электроники [2—5]. Были предложены оптико-электронные модели нейросетей на основе пространственного модулятора света, высокоразрешающей ЭЛТ [3, 4], оптического диска [5].

Возможности параллельной обработки информации и создания массовых межсоединений, а также большой объем памяти оптических дисковых систем могут быть использованы в нейросетевых системах. Однако последовательный побитовый способ регистрации информации на оптическом диске приводит к ряду сложностей и ограничений при записи-считывании двумерных изображений: необходимости предварительного расчета разложения элементов изображения по трекам и исправления аберраций, обусловленных кривизной треков оптического диска, сложности изготовления головки считывания и др. Поэтому использование дисковых ЗУ с голографическим способом записи информации имеет неоспоримые преимущества вследствие естественного параллелизма записи-чтения двумерных данных, возможности восстановления изображения с голограммы без искажений идентично исходному, избы-

точности и инвариантности к сдвигу фурье-голограмм, высокой помехозащищенности, а также низких требований к точности систем записи-считывания (десятки мкм) [6—8].

Успех в области создания систем оптической обработки информации в значительной степени связан с разработкой регистрирующих сред и методов записи информации. Среди реверсивных сред выделяются фототермопластические (ФТП) материалы, обладающие высокой светочувствительностью (10^{-7} Дж/см²), разрешающей способностью (1000 мм^{-1}) и фазовым характером. В [9] был разработан оригинальный одновременный метод локальной записи голограмм на ФТП-носитель при ИК-лазерном нагреве, устраняющий влияние гауссовского распределения интенсивности излучения ИК-лазера на однородность проявления голограммы вдоль ее поверхности и обеспечивающий качественную запись фурье-голограмм аналоговой и цифровой информации при одновременном ФТП-процессе. На основе данного подхода и разработанной методики автоматической записи двумерных фурье-голограмм с однородными характеристиками по всей поверхности фототермопластического диска с управлением процессом регистрации от персонального компьютера был получен ФТП-диск с голографической записью. На рис. 1 приведен голографический фототермопластический диск. Экспериментально продемонстрирована плотность записи информации разработанным методом 2×10^5 бит/мм², что с использованием наложения записей позволяет обеспечить емкость голографического ФТП-диска диаметром 133 мм, равную 1 Гбит. Дифракционная эффективность, полученная для фурье-голограмм цифровой информации диаметром $d_H = 1 \text{ мм}$, равнялась $\eta = 5 \%$ при отношении сигнал/помеха не менее 70 : 1.

Наиболее перспективным, на наш взгляд, является применение голографического ФТП-диска в системах оптической обработки информации. В [10, 11] был предложен вариант оптико-электронной нейронной сети на основе голографического диска (ОНС—ГД). В данной схеме операция векторно-матричного умножения реализуется оптическим путем (рис. 2). Особенность схемы — расположение входного модулятора за устройством, формирующим матрицу весов межсоединений (МВМ), что отличается от традиционной реализации векторно-матричного умножения [2].

Лазерный луч восстанавливает изображение МВМ с двумерной фурье-голограммы (предварительно записанной на ФТП-диск ГД), которое проходит через фурье-объектив L_1 и далее через одномерный входной транспарант ПМС1 с прямоугольными ячейками, вытянутыми в вертикальном направлении. Размер рабочей зоны одномерного управляемого транспаранта совпадает с размерами МВМ. При этом производится умножение интенсивности для каждой информационной точки МВМ на коэффициент пропускания одномерного транспаранта, кодируемого данными входного вектора. Далее с помощью цилиндрической линзы L_2 осуществляются фокусировка и суммирование све-

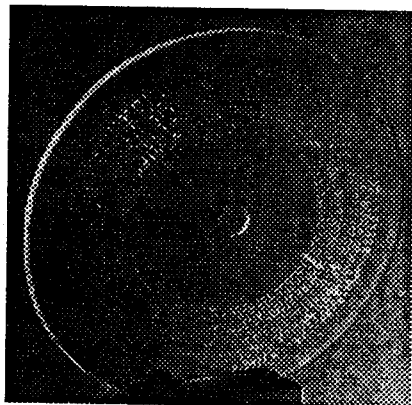


Рис. 1

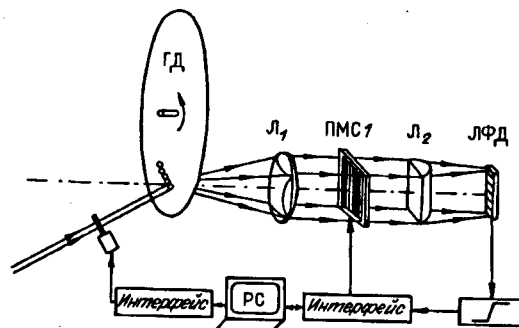


Рис. 2

та на вертикально расположенной линейке фотодиодов (ЛФД). С выхода линейки фотоприемников сигналы поступают через блок электронной обработки в ЭВМ, где выполняются операция сложения результатов умножения на положительную и отрицательную части МВМ, операция пороговой нелинейности и организуется итерационный процесс вычислений в сети. Через интерфейсы осуществляется управление от ЭВМ работой одномерного модулятора света.

Проанализируем точность обработки сигнала в оптической системе ОНС—ГД. Основная часть вычислений производится в аналоговом канале, в который входят полностью оптическая система, включая лазер (Л), голографический диск (ГД), модулятор (М), формирующую оптику и фотоприемник (ФП), а также электронная часть до АЦП. В аналоговых схемах важнейшим вопросом является оценка точности вычислений.

Голограмма, на которой записана положительная либо отрицательная часть матрицы весов межсоединений $T = \{t_{ij}\}$ размерностью $N \times N$, считывается лучом лазера Л мощностью $P_{л}$. Оптическая энергия импульса, несущего МВМ, равна

$$E_n = \eta P_{л} t_n = \eta E_n,$$

где η — дифракционная эффективность голограммы, выражаемая в относительных единицах; E_n — энергия импульса лазера; $t_n = d_n / \pi D \nu$ — продолжительность импульса считывания голограммы; d_n — диаметр голограммы; D — диаметр диска; ν — частота вращения диска. Тогда энергия E_0 , приходящаяся на один бит информации, определится как

$$E_0 = E_n / N^2.$$

Рассмотрим часть аналогового канала для одного элемента фотоприемника, осуществляющего умножение строки матрицы T на входной вектор x (рис. 3). Необходимо, чтобы динамический диапазон фотоприемника, на котором происходит суммирование N оптических сигналов, соответствовал диапазону изменения произведения строки матрицы T на входной вектор, т. е. числу N .

При переходе от одной голограммы к другой происходит изменение дифракционной эффективности. Эксперименты показывают, что η меняется в пределах $\pm 25\%$ [9]. В этом случае при восстановлении изображения матрицы с голограммы оптический сигнал, несущий строку МВМ, равен

$$S_r = (\eta + \Delta\eta) E_0 t_{ij}, \quad j = 1, \dots, N,$$

где t_{ij} — элемент МВМ; $\Delta\eta$ — величина изменения дифракционной эффективности.

После прохождения линейки модуляторов, где производится операция умножения, сигнал становится равным:

$$S_m = (\eta + \Delta\eta) E_0 t_{ij} \alpha x_j, \quad j = 1, \dots, N,$$

где α — коэффициент пропускания модулятора; $x_j, j = 1, \dots, N$ — элемент входного вектора.

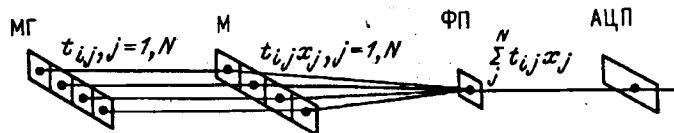


Рис. 3

В фотоприемнике происходит сложение N оптических сигналов:

$$S_{\text{ФП}} = k_{\text{ФП}} \alpha \eta E_0 (I + \Delta\eta/\eta) \sum_j^N t_{ij} x_j,$$

где $k_{\text{ФП}}$ — коэффициент преобразования фотоприемника. Его можно подобрать таким образом, чтобы компенсировать потери в оптическом канале $k_{\text{ФП}} = 1/\alpha\eta$. Тогда

$$S_{\text{ФП}} = E_0 (1 + \Delta\eta/\eta) \sum t_{ij} x_j.$$

Такой сигнал, дискретизованный по тактам работы ОНС, приходит на вход АЦП, где выполняются квантование и кодирование аналогового сигнала. Известно, что квантование приводит к возникновению методической погрешности, называемой шумом квантования. Максимальное значение этой погрешности равно по абсолютной величине половине единицы младшего разряда [12].

Примем за единицу квантования величину E_0 . Перепишем последнее выражение в виде

$$S_{\text{ФП}} = E_0 \left(\sum t_{ij} x_j + \Delta\eta/\eta \sum t_{ij} x_j \right).$$

Второе слагаемое, обусловленное изменением дифракционной эффективности, приводит к погрешности. Исходя из характеристики квантования должно выполняться следующее условие:

$$E_0 \Delta\eta/\eta \sum t_{ij} x_j < 0,5 E_0;$$

$$\Delta\eta < 0,5 \eta / \sum t_{ij} x_j.$$

Например, если в результирующей строке будет хотя бы половина значащих пикселей, т. е. $\sum t_{ij} x_j = N/2$, то

$$\Delta\eta < \eta/N.$$

При $N = 64$ $\Delta\eta < 0,0157\eta$. Таким образом, допустимое изменение дифракционной эффективности должно быть не более 1,5%. На практике данное требование трудно реализуемо из-за неоднородности характеристик регистрирующего материала, влияния окружающей среды и т. д. [9].

В связи с этим предлагается изменить оптическую схему, используя матрицу фотоприемников. При этом в оптической системе будет выполняться

только операция умножения, а суммирование — в электронной части. В этом случае структура ОНС—ГД будет иметь вид, представленный на рис. 4. Современные возможности интегральной технологии позволяют изготовить одномерный модулятор света и фотоматрицу в одном устройстве, а узел считывания изображений матриц весов межсоединений с голограмм может быть реализован в виде компактной подвижной головки. На вход каждого элемента фото-

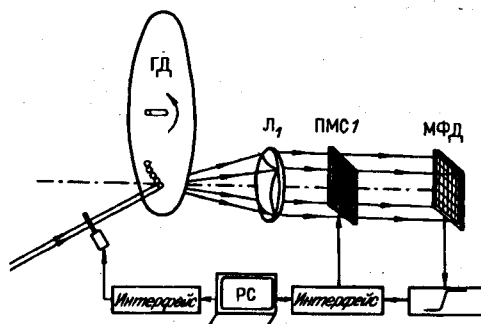


Рис. 4

приемной матрицы размерностью $N \times N$ подается двоичный сигнал:

$$S_{\Phi\Pi} = E_0\eta\alpha(1 + \Delta\eta/\eta)t_{ij}x_j.$$

Тогда, настроив порог чувствительности ФП на величину

$$\mu = E_0\eta\alpha \pm E_0\Delta\eta\alpha,$$

Такое условие легко выполнимо по существующей технологии записи-воспроизведения голограмм на фототермопластический материал.

Однако в структуре с матрицей ФП увеличивается время вычислений из-за выполнения операций суммирования электронными схемами. Если принять время выполнения одной операции сложения $t_{\text{сл}} = 0,1$ мкс [14], то общее время суммирования N чисел

$$t_{\Sigma} = t_{\text{сл}}(N - 1).$$

Так, при $N = 64$ $t_{\Sigma} = 6,3$ мкс. Это существенно меньше времени перехода с одной голограммой на другую на голографическом диске (сотни микросекунд).

Одна из важнейших характеристик нейронной сети — число нейронов N и размерность матрицы весов межсоединений $N \times N$, характеризующей число связей сети. С увеличением этого числа возрастает информационная емкость сети, улучшается устойчивость к помехам, увеличивается пространственная полоса пропускания системы. Поэтому далее рассмотрим вопросы обработки изображений большой размерности на ОНС—ГД. В оптико-электронной нейронной сети на основе голографического диска входной вектор формируется одномерным модулятором света размерностью, равной числу нейронов в сети N . Матрица связей размерностью $N \times N$ записывается на голограмму на диске. Таким образом, предельные значения числа нейронов N и числа связей $N \times N$ определяются возможностями одномерного ПВМС и плотностью записи на голографическом диске. Что касается линейки модуляторов, то уже существуют матрицы ПВМС размерностью 512×512 [15]. Это означает, что в пределе на одной фурье-голограмме может быть записана матрица размерностью $N \times N = 512 \times 512$ В [16], максимальная плотность записи рассчитана как 10^6 бит/мм². Однако практически удается записать информацию с плотностью 2×10^5 бит/мм² [17, 18]. Таким образом, предельное число нейронов в ОНС—ГД ограничивается разрешением $R = 512$, а число связей — величиной 2×10^5 .

Применительно к задачам распознавания образов и обработки изображений при числе нейронов в сети $N = 512$ допустимая размерность изображения оказывается менее 23×23 . В практических задачах обрабатывают изображения намного большей размерности. Поэтому возникает необходимость решения задач большой размерности на небольшой нейронной сети.

Пусть N — требуемая размерность обрабатываемых векторов; R — разрешение голограммы; L — размерность матрицы ПВМС и фотоприемников, причем $N > R$, $R = L$. Определим $D = \text{int}(N/L)$, где int — функция выделения целого. Разделим входной вектор размерностью N на D составляющих размерностью L , а матрицу весов межсоединений размерностью $N \times N$ на $D \times D$ размерностью $L \times L$ каждая (рис. 5, а, б).

Результирующий вектор Y размерностью N определяется по формуле

$$Y = f(TX),$$

где X — входной вектор размерностью N , T — МВМ размерностью $N \times N$, f — пороговая функция.

После пространственного разделения МВМ и входного вектора вектор Y будет вычисляться по частям, подобно тому как был разделен входной вектор X . Вектор, являющийся k -й частью вектора Y , определяется по следующему выражению:

$$Y_k = f\left(\sum_{i=1}^D T_i^{k,i} x_i\right), \quad k = 1, \dots, D.$$

При этом элементы вектора Y размерностью N можно вычислить по формуле

$$y_i = f\left(\sum_{k=0}^{D-1} \sum_{j=1}^{(k+1)L} t_{ij} x_j\right),$$

$$i = \{(kL + 1) \dots (k + 1)L\}, \quad k = 0 \dots D - 1\}.$$

Здесь k — индекс части вектора, а выражение в квадратных скобках — изменение индекса внутри каждой части.

Таким образом, рассчитанная МВМ размерностью $N \times N$ записывается в $D \times D$ голограммах на диске. Входной вектор подается на матрицу ПВМС последовательно, по частям. Соответственно на диске осуществляется последовательный просмотр D голограмм с субматрицами T_{11}, \dots, T_{1D} . Полученные в результате работы ОНС—ГД D субвекторов размерностью L суммируются для определения D -й части вектора Y . После этого процесс повторяется до получения полного вектора Y .

Нетрудно заметить, что время вычислений увеличивается в D^2 раз. Время на выполнение дополнительных $D(D - 1)$ операций суммирования не затрачивается, поскольку они выполняются во время перехода от одной субматрицы к другой на голографическом диске. Например, при $N = 1024$ (изображение размерностью 32×32), $R = 512$, $L = 256$ параметр D будет равен 4. Значит, время вычислений увеличивается в $4^2 = 16$ раз. Однако оно, конечно, намного меньше времени, которое затрачивается в электронном нейрокompьютере.

Оценим быстродействие работы ОНС—ГД. Общее время выполнения вычислений в ОНС—ГД t_2 складывается из следующих составных частей:

— время выполнения операции умножения элементов входного вектора на элементы МВМ параллельно, оптическим путем, не включая операции суммирования t_y ;

— время выполнения операции суммирования t_c ;

— время выполнения операции вычитания результата умножения на отрицательную часть МВМ из результата умножения на положительную часть МВМ t_b ;

— время выполнения пороговой операции t_n ;

— время выполнения операции сравнения результата данной итерации с результатом предыдущей итерации t_w .

Поскольку векторно-матричное умножение выполняется отдельно для положительной и отрицательной части

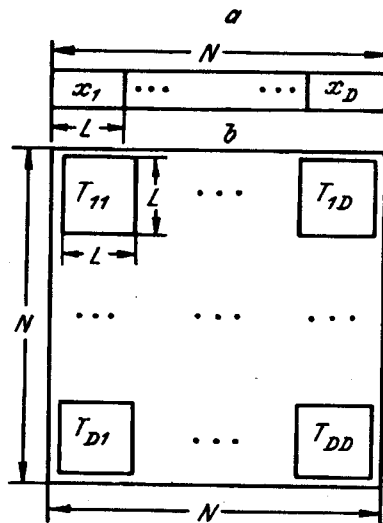


Рис. 5

МВМ, то t_y и t_c входят в уравнение с коэффициентом 2:

$$t_{\Sigma} = 2t_y + 2t_c + t_b + t_n + t_n.$$

Операции суммирования, вычитания, сравнения и пороговая операция, являющаяся также операцией сравнения, выполняются в ЭВМ. Эти операции одного ранга и выполняются за одно и то же время $t_{оп}$. Тогда

$$t_c = t_{оп}(N - 1);$$

$$t_b = t_{оп}N;$$

$$t_n = t_{оп}N;$$

$$t_c = t_{оп}N,$$

где N — размерность вектора.

Операция умножения осуществляется оптическим путем параллельно. Время выполнения этой операции определяется временем прохождения луча лазера через оптическую схему:

$$t_y = S/c,$$

где S — длина схемы, c — скорость света.

Если принять $S = 0,3$ м, то $t_y = 10^{-9}$ с. Эта величина на несколько порядков меньше других составляющих общего времени вычислений, поэтому ею можно пренебречь. Тогда

$$t_{\Sigma} = 2t_{оп}(N - 1) + t_{оп}N + t_{оп}N + T_{оп}N$$

или

$$t_{\Sigma} = 5Nt_{оп}.$$

Например, для $N = 256$, $t_{оп} = 0,1$ мкс общее время выполнения вычислений в нейронной сети $t_{\Sigma} = 128$ мкс.

Особенностью ОНС—ГД является то, что выполнение операций суммирования и сравнения в ЭВМ происходит одновременно с переключением от одной МВМ к другой. Время переключения, в свою очередь, определяется скоростью вращения ФТП-диска:

$$t_d = d/\pi D\nu,$$

где d — расстояние между голограммами; D — диаметр диска; ν — частота вращения диска.

Если принять $d = 1$ мм, $D = 160$ мм, $\nu = 1000$ об./мин = 16,67 об./с, то $t_d = 120$ мкс.

Таким образом, время перехода от одной голограммы к другой на ГД и время выполнения операций в ЭВМ являются примерно равными величинами. Другими словами, пока происходит переход от одной матрицы связей к другой, ЭВМ успевает выполнить все необходимые вычисления. Поэтому быстродействие ОНС—ГД полностью определяется временем смены голограммы на ФТП-диске.

Быстродействие нейронной сети определяется числом переключений связей за единицу времени. Поскольку на голограмме можно хранить до 10^5 весов связей, а время переключения связей составляет ~ 120 мкс, быстродействие

ОНС—ГД будет $8,3 \times 10^8$ переключений в секунду.

Важным достоинством голографического диска является большая емкость памяти. При плотности записи цифровой информации 10^5 бит/мм² емкость ГД составляет 0,6 Гбит и может быть увеличена за счет наложения голограмм до

Нейросеть	Число связей	Число переключений связей, с (быстродействию)
Многопроцессорный вычислительный комплекс СМ-2 (64К ПЭ)	64×10^6	13×10^6
Электронный нейрочип	$10^2 - 10^3$	25×10^8
ОНС—ГД	2×10^5	8×10^8

1,2 Гбит, что позволит хранить до 6000 матриц связей размерностью 512×512 .

Сопоставление с лучшими представителями электронных нейрокомпьютеров показывает (см. таблицу), что ОНС—ГД по числу связей вплотную приближается к многопроцессорному вычислительному комплексу "Connection Machine-2" с 64000 процессорными элементами, а по быстродействию превосходит этот доро-

стоящий комплекс. По сравнению с разрабатываемыми сейчас электронными нейрочипами ОНС—ГД на несколько порядков превосходит их по числу связей при одинаковом быстродействии.

Таким образом, оптико-электронная нейросеть на основе голографического ФТП-диска может иметь лучшие показатели по сравнению с электронными аналогами.

СПИСОК ЛИТЕРАТУРЫ

- Hopfield J. J. Neural networks and physical systems with emergent collective computational abilities // Proc. Nat. Acad. Sci. US.—1982.—79.—P. 2554.
- Farhat N. H., Psaltis D., Prata A., Park E. Optical implementation of the Hopfield model // Appl. Opt.—1985.—24.—P. 1469.
- Yu F. T. S., Lu T., Yang X., Gregory A. Compact size optical neural network using liquid crystal televisions // Proc. SPIE.—1990.—P. 1215.
- Wu S., Lu T., Xu X., Yu F. T. S. An adaptive optical neural network using a high resolution video monitor // Microwav. and Opt. Techn. Lett.—1989.—2, N 7.—P. 252.
- Psaltis D., Neifeld M., Jamamura A. et al. Optical memory disks in optical information processing // Appl. Opt.—1990.—29, N 14.—P. 2038.
- Микаэлян А. Л. Оптические методы в информатике: запись, обработка и передача информации.—М.: Наука, 1990.
- Блок А. А., Домбровский В. А., Домбровский С. А., Пен Е. Ф. Экспериментальные исследования достоверности считывания данных в голографических ЗУ // Автометрия.—1984.—№ 3.
- Kutanov A. A., Abdrisaev B. D. Holographic disk with thermoplastic recording for optical information processing // Opt. Comput. and Process.—1991.—1, N 4.—P. 315.
- Акаев А. А., Абдрисаев Б. Д., Кутанов А. А. и др. Некоторые принципы локальной записи микроголограмм на движущийся фототермопластический диск.—Фрунзе, 1990.—(Препр. АН Кирг. ССР, Ин-т физики).
- Акаев А. А., Кутанов А. А., Абдрисаев Б. Д., Дордоев С. З. Оптико-электронная реализация модели Хопфилда для нейронной сети на основе голографического фототермопластического диска // ДАН СССР.—1991.—321, № 2.
- Kutanov A. A., Abdrisaev B. D., Dordoev S. Z. Holographic disk based optical neural network // Opt. Lett.—1992.—17, N 13.—P. 952.
- Бахтияров Г. Д., Малинин В. В., Школин В. П. Аналого-цифровые преобразователи / Под ред. Г. Д. Бахтиярова.—М.: Сов. радио, 1980.
- Аксененко М. Д., Баранчиков М. Л. Приемники оптического излучения: Справочник.—М.: Радио и связь, 1987.
- Долтон Р., Мюллер С. Персональные ЭВМ семейства IBM PS/2: Пер. сангл.—М.: Мир, 1991.
- Нефф Дж. А., Аткале Р. А., Ли С. Х. Двумерные пространственные модуляторы света: Методический обзор // ТИИЭР.—1990.—78, № 5.
- Акаев А. А., Майоров С. А. Когерентные оптические вычислительные машины.—Л.: Машиностроение, 1977.
- Акаев А. А., Жумалиев К. М., Кутанов А. А. Запись микроголограмм на фототермопластические носители при проявлении излучением CO₂-лазера // Квантовая электрон.—1983.—10, № 7.
- Блок А. А., Домбровский В. А., Домбровский С. А. и др. Практический предел плотности записи данных в голографических ЗУ на плоских носителях // Автометрия.—1989.—№ 5.

Поступила в редакцию 18 марта 1993 г.