

## Содержание

1	Список обозначений и сокращений	3
2	Введение	4

# 1 Список обозначений и сокращений

CD (Critical Dimension) - размер минимального элемента на слое чипа

RET (Resolution Enhancement Technique) - прием улучшения разрешения на микрочипе

Q - излучение, падающее на голограмму (маску) - поле подсветки

Г - плоскость голограммы

b - размер голограммы

E - комплекснозначная функция пропускания маски или голограммы ("комплексная голограмма")

T - вещественнозначная функция пропускания маски или голограммы ("серая голограмма")

B - бинарная функция пропускания голограммы ("бинарная голограмма")

W - поле подсветки после прохождения через голограмму (маску)

d - расстояние между голограммой и плоскостью микросхемы

I - плоскость изображения

a - размер изображения

$(\xi_C, \eta_C)$  - центр микросхемы

$h(\xi, \eta)$  - интенсивность излучения в плоскости фоторезиста

$p(\xi, \eta)$  - форма канавок в фоторезисте после экспонирования

$q(\xi, \eta)$  - профиль дорожек в слое металла

$q_0(\xi, \eta)$  - желаемая топология чипа (желаемый профиль дорожек в слое металла)

## 2 Введение

В последние десятилетия методы литографии получили широчайшее распространение. Одно из главных применений микролитографии (создание структур менее 10 мкм) и нанолитографии (создание структур менее 100 нм) - это создание слоев микрочипов. За последние десятилетия технология создания микропроцессоров интенсивно развивалась в соответствии с т.н. законом Мура (рис. 1). В 1965 году инженер Мур высказал предположение, что число транзисторов на кристалле будет удваиваться каждые 24 месяца. При этом характерный размер элементов серийно выпускаемых чипов уменьшился с 10 мкм в 1972 году (процессор Intel 8008) до 45 нм в 2008 году (Intel i7, ядро Nehalem). Все это время основным способом создания микроструктур слоев процессора была фотолитография - когда желаемая топология создается на трафарете (маске) и с помощью осветительной уменьшающей системы переносится в плоскость микрочипа (рис. 3).

На слой кремния, находящийся рядом с фокусом оптической проекционной системы, нанесен слой фоторезиста - вещества, которое в процессе засветки меняет свои свойства в местах, на которые попало излучение. После облучения фоторезист проявляется - в местах, которые были засвечены толщина слоя уменьшается, возможно, до самой подложки из кремния. Затем фоторезист закрепляется и начинается процесс травления металла. В результате в местах, которые были освещены на слое металла образуются канавки и в кремнии формируется требуемая топология из элементарных элементов (прямоугольников со сторонами параллельными осям координат либо прямоугольников, наклоненных на 45 градусов) отвечающих проводникам либо элементам транзисторов. Далее наносится следующий слой металла и процесс повторяется. Таким образом формируется требуемая трехмерная топология микропроцессора (рис. 2).

Со времени создания первых образцов микрочипов технологический процесс фотолитографии существенно усложнился. Стала использоваться иммерсия (когда между проекционным объективом и фоторезистом помещается жидкость с высоким коэффициентом преломления), изменилось освещение (сейчас применяют квадрупольную, кольцевую и др. внеосевые подсветки), с появлением мощной вычислительной техники широко применяется метод оптических коррекций (ОРС) - когда в маску добавляются некоторые

## CPU Transistor Counts 1971-2008 & Moore's Law

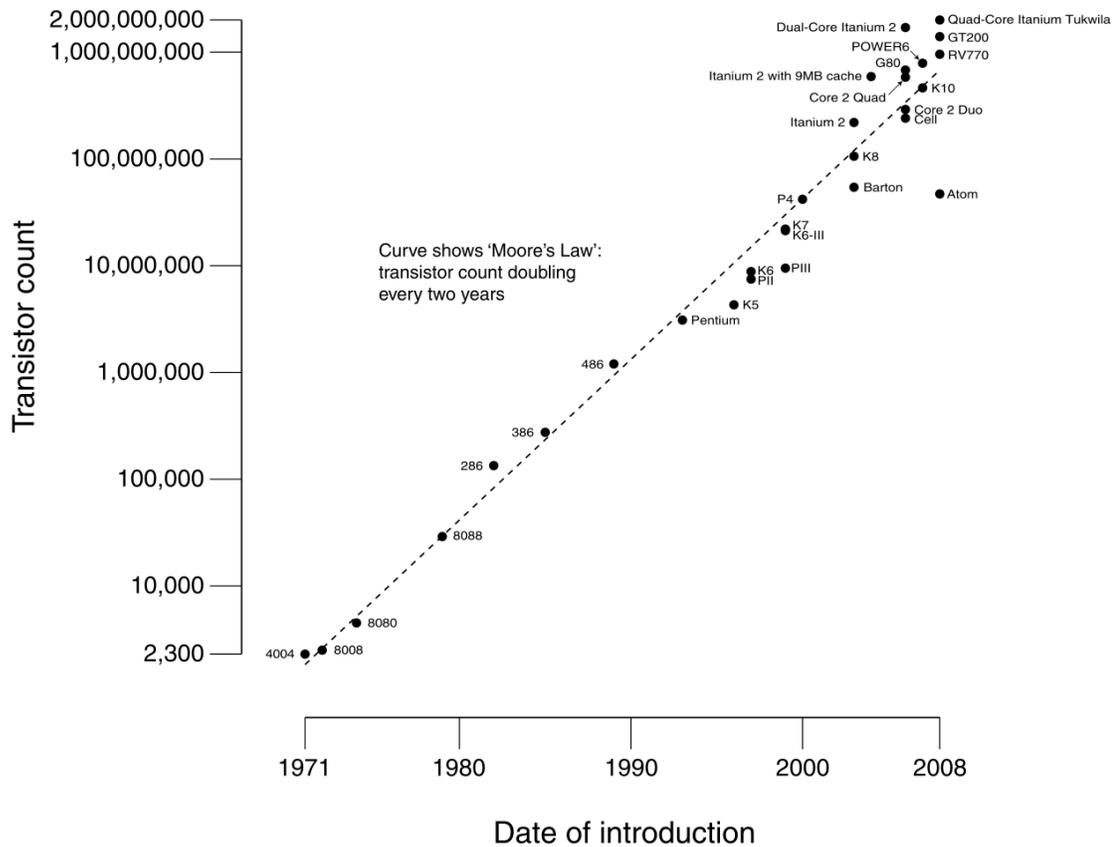


Рис. 1: Рост количества транзисторов на процессорах в 1971-2008 годах

корректирующие элементы (прямоугольники вдоль сторон и квадратики близ углов исходной топологии). С уменьшением характерных размеров структур проявляются новые физические эффекты, усложняющие дальнейшее развитие проекционной фотолитографической технологии. Одним из новейших методов фотолитографии является предложенная в 2008 году технология фазовых масок [1].

Кроме фотолитографии применяются либо активно разрабатываются следующие нанолитографические технологии: электронно-лучевая литография, интерференционная литография, рентгеновская литография, жестко-ультрафиолетовая литография, дифракционная литография. Еще в 1968 году была известна идея голографической литографии. В [6] была предложена оптическая схема без уменьшающего объектива, где

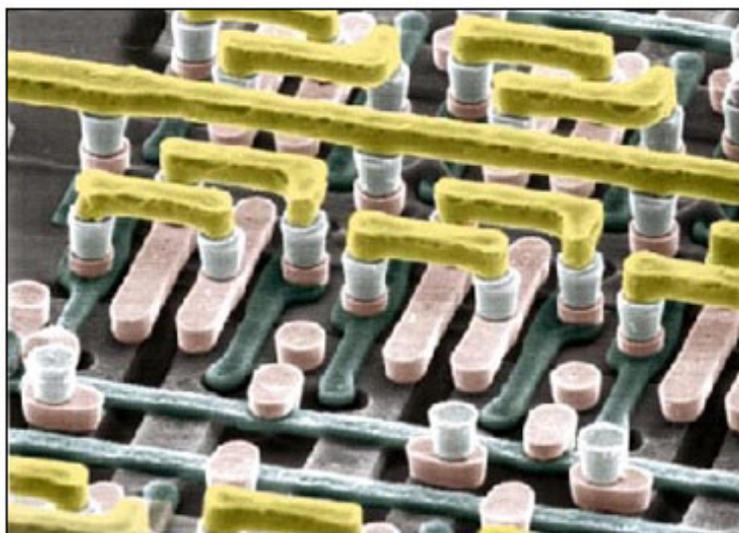


Рис. 2: Несколько слоев микропроцессора без диэлектрика

вместо маски используется голограмма, предварительно записанная на фотопластинке по методу Габора [15],[16] (рис. 4). Однако в то время голографическая литография развития не получила, так как получаемые с ее помощью изображения на фоторезисте имели неудовлетворительное качество. Действительно, при создании голограммы физическим путем, в изображении на границах элементов обязательно возникают волнообразные всплески - подобно эффекту Гиббса в рядах Фурье [8]. В настоящей работе показано, что с использованием оптимизационных алгоритмов может быть создана цифровая голограмма, дающая требуемое технологами качество изображения в плоскости чипа.

Идея использовать уже цифровую голографическую литографию была предложена Раховским В.И в (...) году. Голограмма, позволяющая создать слой современного процессора (примерно 3x4 см) должна содержать триллионы отверстий. Расчет таких голограмм ранее не представлялся возможным. Однако, как показано в настоящей работе, в последние годы мощность суперкомпьютерных вычислителей достигла уровня, позволяющего по заданной топологии слоя микросхемы рассчитать требуемую голограмму.

Таким образом, **целью настоящей диссертационной работы** являлось математическое описание метода голографической литографии, разработка и реализация на супер-ЭВМ алгоритмов расчета требуемых голограмм и алгоритмов моделирования процесса фотоэкспонирования. Во-первых,

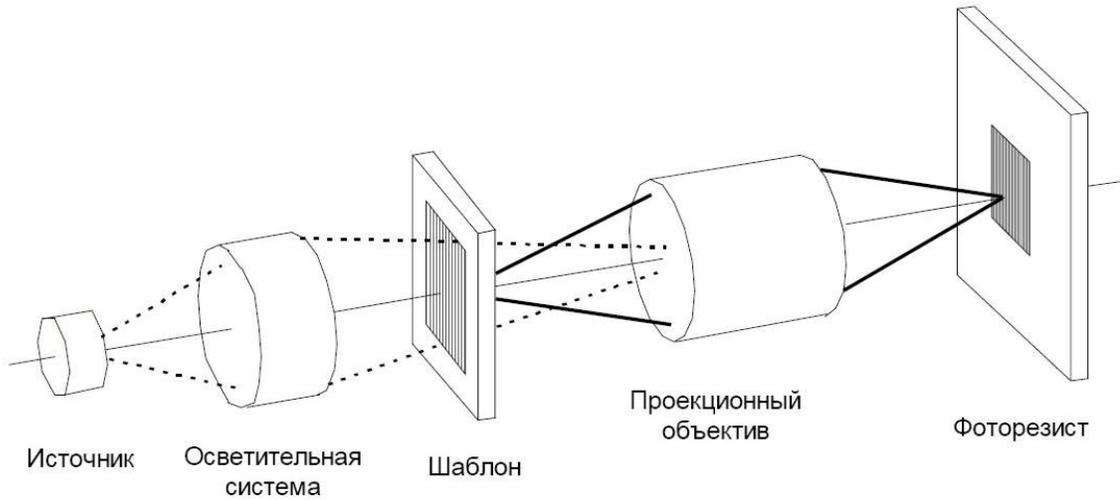


Рис. 3: Схема фотолитографической установки

нужно уметь находить такое месторасположение элементарных отверстий на голограмме, что после процессов засветки и проявления фоторезиста и травления кремния получится слой чипа с топологией максимально близкой к заданной. Во-вторых с целью уменьшить количество дорогих экспериментов нужно иметь программный комплекс позволяющий проводить численное моделирование литографического процесса.

Обозначим через  $Q$  поле, падающее на голограмму размера  $b \times b$ , находящуюся в координатной плоскости  $xOy$  - плоскости  $\Gamma$ . Голограмму, представляющую из себя набор отверстий ("бинарную голограмму") обозначим через  $B(x, y)$ . Технологически именно такую голограмму можно изготовить.

$$B(x, y) : \left[-\frac{b}{2}, \frac{b}{2}\right] \times \left[-\frac{b}{2}, \frac{b}{2}\right] \longrightarrow \{0, 1\}$$

где значение 0 отвечает отсутствию покрытия пластинки голографической маски - отверстию, а 1 - непрозрачной области голограммы.

Для постановки и решения задачи расчета бинарной голограммы удобно использовать следующие непрерывные функции, описывающие пропускание голограммы:

$$T(x, y) : \left[-\frac{b}{2}, \frac{b}{2}\right] \times \left[-\frac{b}{2}, \frac{b}{2}\right] \longrightarrow [0, 1]$$

$$E(x, y) : \left[-\frac{b}{2}, \frac{b}{2}\right] \times \left[-\frac{b}{2}, \frac{b}{2}\right] \longrightarrow \{g | g \in \mathbb{C}, |g| \leq 1\}$$

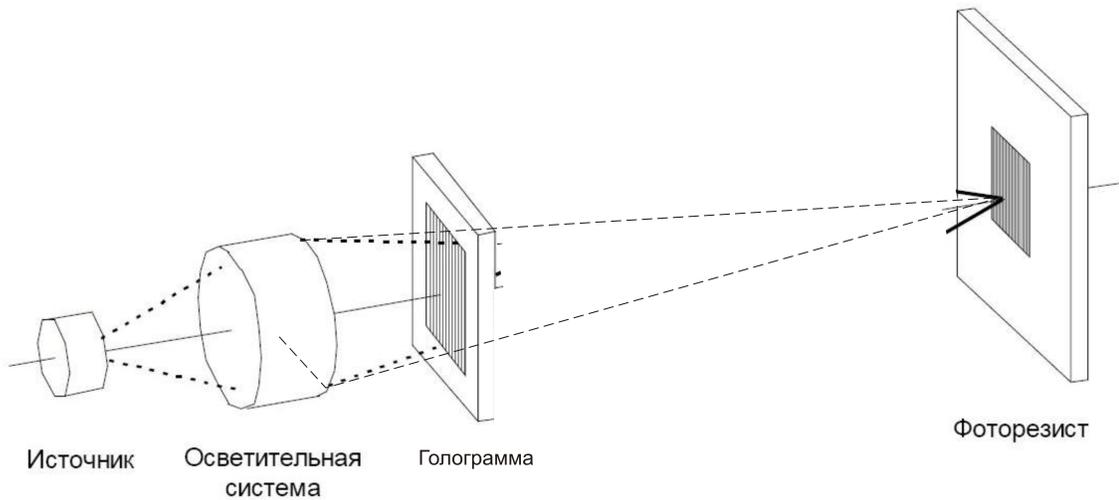


Рис. 4: Схема голографической литографической установки

Эти функции будем называть "серая" и "комплексная" голограмма соответственно.

Поле, сформировавшееся после прохождения голограммы, обозначим за  $W(x, y)$ . На расстоянии  $d$  от голограммы находится плоскость изображения И. Здесь в области размером  $a \times a$  происходит регистрация изображения. В голографическом степпере эта область должна попадать на фоторезист обрабатываемого в данный момент слоя; и в плоскости фоторезиста должно формироваться требуемое распределение интенсивности излучения.

Если это не будет приводить к недоразумениям, те же обозначения будут использоваться и при описании проекционной литографии.

При моделировании процесса распространения световой волны в проекционной и голографической литографии используются три основные модели: когерентная, некогерентная и частично когерентная.

В когерентной модели излучение от поля  $W(x, y)$  в плоскости голограммы Г в плоскости изображения И создаст интенсивность

$$h_{coh}(\xi, \eta) = \left| \int_{-\frac{b}{2}}^{\frac{b}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} W(x, y) K(\xi - x, \eta - y) dx dy \right|^2 \quad (1)$$

где  $K(x, y)$  - функция источника в голографической литографии или некоторая функция, зависящая от проецирующей оптической системы для

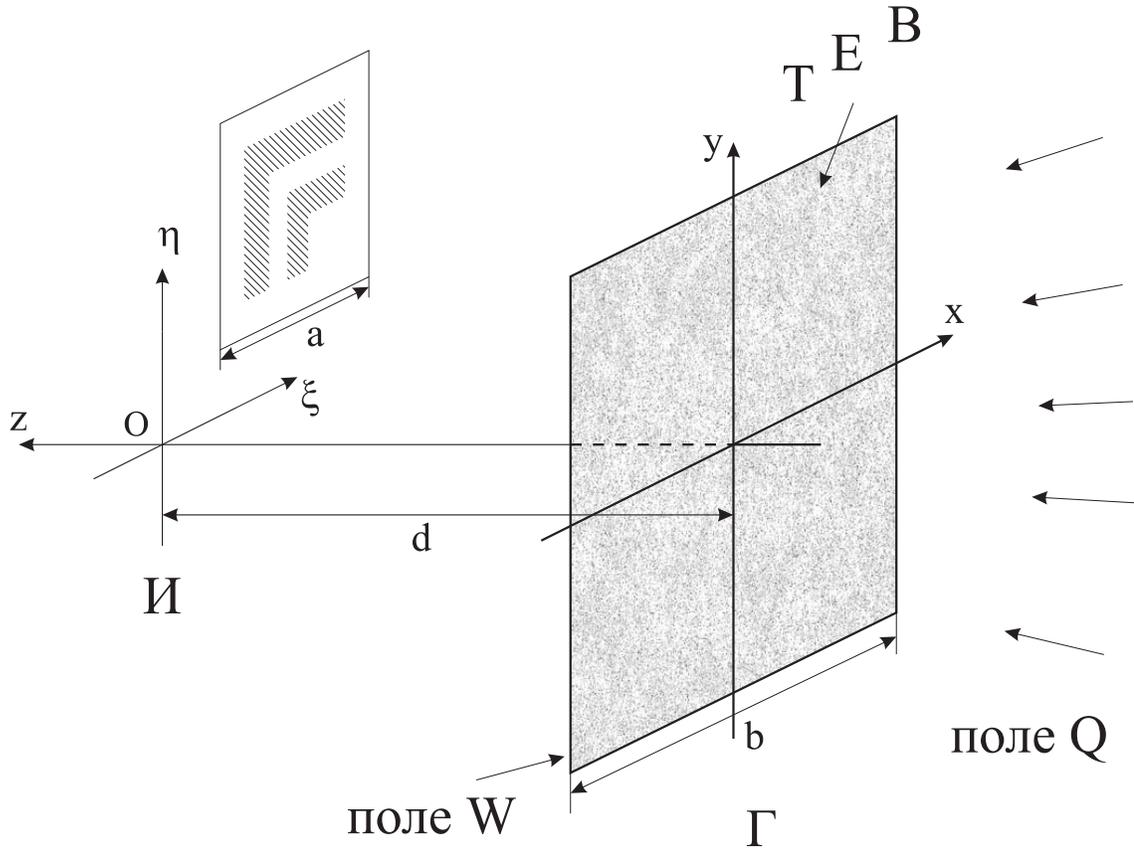


Рис. 5: Схема процесса экспонирования фоторезиста

случая проекционной литографии. Например, в [11]

$$K(x, y) = \frac{J_1(2\pi r NA/\lambda)}{2\pi r NA/\lambda} \quad (2)$$

где  $J_1$  - функция Бесселя первого рода.

Если подсветка некогерентна, получаем формулу:

$$h_{incoh}(\xi, \eta) = \int_{-\frac{b}{2}}^{\frac{b}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} |W(x, y)|^2 |K(\xi - x, \eta - y)|^2 dx dy \quad (3)$$

где  $K$  - из (2).

Интенсивность от частично-когерентной подсветки считается как сумма  $s$  когерентных:

$$h_{pcoh}(\xi, \eta) = \sum_{j=1}^{j=s} \sigma_j \left| \int_{-\frac{b}{2}}^{\frac{b}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} W(x, y) K_j(\xi - x, \eta - y) dx dy \right|^2 \quad (4)$$

Излучение, сформированное источником, пройдя через осветительную систему попадает на маску. Маска - это пластина, обычно кварцевая, на которую наносится увеличенная желаемая топология слоя чипа в случае обычной проекционной литографии либо голограмма  $B(x, y)$  для голографической литографии. В настоящей работе рассматривались голографические маски с двумя возможными уровнями пропускания в каждой точке: 0 - где голограмма абсолютно непрозрачна и 1 - где голограмма полностью пропускает все излучение. В настоящее время широко используют еще несколько типов масок (см. таблицу из [11] на рис. 6).

RET	Allowable transmission values
OPC	0 or +1
6% Attenuated PSM	-0.2449 or +1
18% Attenuated PSM	-0.4243 or +1
Strong PSM (100% transmission)	-1 or +1
Strong PSM (With chrome)	-1 or 0 or +1

Рис. 6: Коэффициенты пропускания различных типов масок [11]

OPC - обычная маска. PSM (Phase Shift Mask) - это фазосдвигающие маски. Strong PSM со 100% пропусканием - это маска на которой есть области двух типов: либо полное пропускание либо полное пропускание но со сдвигом фазы на  $180^\circ$ . Attenuated PSM маски отличаются тем, что области сдвигающие фазу также ослабляют интенсивность прошедшего через них излучения. Наконец, Strong PSM маску можно в некоторых местах покрыть непрозрачным хромом. Тогда получим три значения пропускания: -1, 0 и +1.

В 2008 году были представлены Pixelated Phase Masks (PPM) - маски на которых возможны области с любым сдвигом фазы [1],[2]. На стекле вытравливаются очень маленькие гребенки (рис. 7). От частоты этих гребенок и зависит сдвиг фазы прошедшего через такую маску излучения. Пока были изготовлены двухтоновые (со сдвигом фазы  $0^\circ$  и  $180^\circ$ ) PPM маски для слоя процессора 65нм [2], однако предложенная технология позволяет создавать до

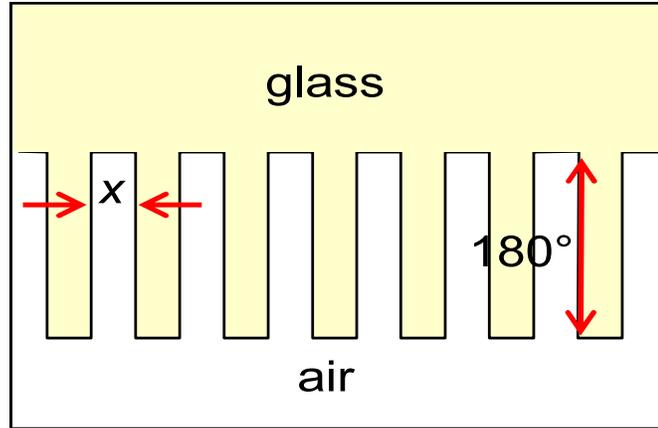


Рис. 7: Схема гребенки на маске PPM [1]

4-х и более различных значений сдвигов фаз на одной маске.

В проекционной литографии проецирующая система представляет собой очень дорогостоящий и сложный набор оптических элементов, позволяющий достичь 4-х кратного уменьшения топологии, нанесенной на маску. В голографической литографии проекционная система отсутствует и поле  $W(x,y)$ , сформированное голограммой свободно распространяется до плоскости фоторезиста И.

В плоскости чипа слой за слоем формируется будущая микросхема.

На уже сформированный слой металла с диэлектриком  $M_i$  наносится слой металла  $M_{i+1}$  и фоторезист (рис. 8). После засветки излучением  $W$ , происходит обработка фоторезиста специальным раствором. В результате фотоэкспонирования материал фоторезиста в местах, подверженных действию излучения  $W$  изменил свои свойства и, например, негативный фоторезист вымывается там где он был засвечен. В плоскости изображения интенсивность падающего излучения описывается функцией  $h(\xi, \eta)$ .

$$h(\xi, \eta) : \left[-\frac{a}{2}, \frac{a}{2}\right] \times \left[-\frac{a}{2}, \frac{a}{2}\right] \longrightarrow [0, 1]$$

Функция формы канавок на фоторезисте

$$p(\xi, \eta) : \left[-\frac{a}{2}, \frac{a}{2}\right] \times \left[-\frac{a}{2}, \frac{a}{2}\right] \longrightarrow [0, 1]$$

Существуют сложные модели, описывающие зависимость толщины оставшегося фоторезиста  $p(\xi, \eta)$  от интенсивности излучения в плоскости

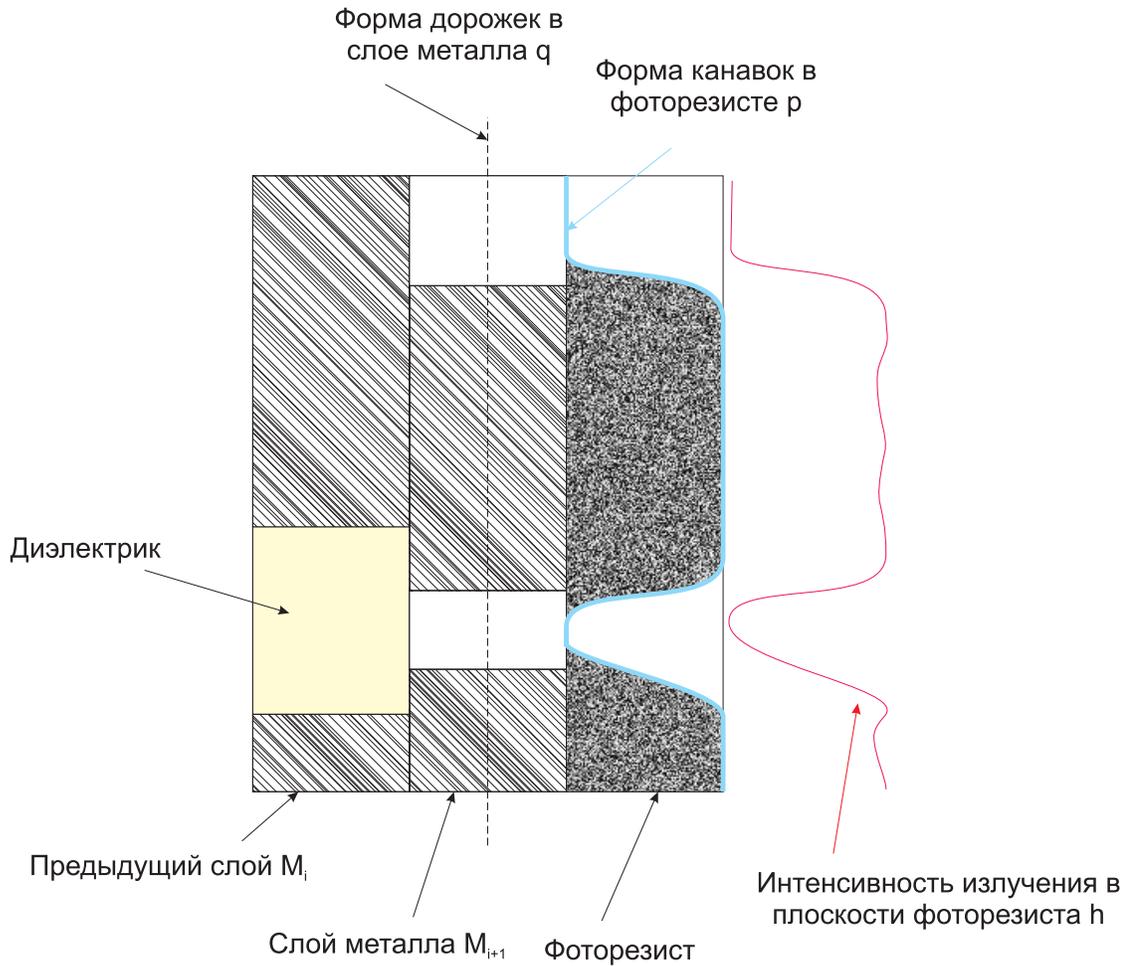


Рис. 8: Схема слоев металла и негативного фоторезиста в плоскости чипа

И  $h(\xi, \eta)$  и времени экспонирования [17]. Далее проводится травление металла  $M_{i+1}$ . В тех местах где фоторезист отсутствует или толщина его невелика, в металле образуются канавки с профилем  $q(\xi, \eta)$ .

$$q(\xi, \eta) : \left[-\frac{a}{2}, \frac{a}{2}\right] \times \left[-\frac{a}{2}, \frac{a}{2}\right] \longrightarrow \{0, 1\}$$

Эти канавки заполняются диэлектриком а остатки фоторезиста смываются, наносится следующий слой металла  $M_{i+2}$ , устанавливается соответствующая слою  $M_{i+2}$  новая маска и процесс обработки слоя повторяется.

Для расчетов применяются модели, связывающие функцию интенсивности излучения  $h(\xi, \eta)$  сразу с профилем дорожек на чипе  $q(\xi, \eta)$ . Самая популярная - это однопороговая модель [2],[11],[11],[14]:

$$q_{\Theta}(\xi, \eta) = \begin{cases} 1, & \text{если } h(\xi, \eta) > \Theta; \\ 0, & \text{если } h(\xi, \eta) \leq \Theta. \end{cases} \quad (5)$$

Существуют также более сложные двухпороговые модели [19] и модели, в которых  $q$  в точке  $\xi_0, \eta_0$  зависит от поведения функции интенсивности  $h$  в некоторой окрестности  $\xi_0, \eta_0$  [20]

$$q_h(\xi, \eta) = \mathbf{q}(h), \quad (6)$$

причем эта зависимость определяется во многом эмпирически.

Общая задача литографии в том числе проекционной и голографической может быть сформулирована следующим образом. Требуется подобрать такие источник, осветительную оптику, маску, проецирующую систему и так организовать процесс экспонирования фоторезиста и травления металла, чтобы в результате в слое металла получить заданный профиль дорожек  $q_0$ :

$$\mathbf{F} \circ \mathbf{P} \circ \mathbf{M} \circ \mathbf{L} \circ \mathbf{R} = q_0 \quad (7)$$

где  $\mathbf{R}$  - излучение источника,  $\mathbf{L}$  - оператор соответствующий преобразованию поля, задаваемому осветительной системой,  $\mathbf{M}$  - преобразование поля маской,  $\mathbf{P}$  - преобразование проецирующей системы,  $\mathbf{F}$  - оператор преобразующий поле падающее на фоторезист в профиль дорожек в металле  $q$ , получающийся после экспонирования фоторезиста и травления металла.

Для оценки размера минимального элемента на слое микросхемы (CD - Critical Dimension) в традиционной фотолитографии пользуются соотношением [11]

$$CD = \frac{k\lambda}{NA}.$$

Длина волны лазера  $\lambda$  на данный момент равна 193 нм и ее дальнейшее уменьшение потребует серьезного усложнения и удорожания технологии, т.е. источник  $\mathbf{R}$  в (7) улучшить нельзя. Текущее значение числовой апертуры оптической системы  $NA$  составляет около 0.9 при физическом пределе  $NA = 1$  для неиммерсионной (безжидкостной) литографии. Таким образом  $\mathbf{L}$  практически не поддается коррекции, а для улучшения проецирующей системы  $\mathbf{P}$  применяют иммерсионную литографию: когда проекционная установка помещается в жидкость с большим коэффициентом преломления  $n$  (рис. 9).

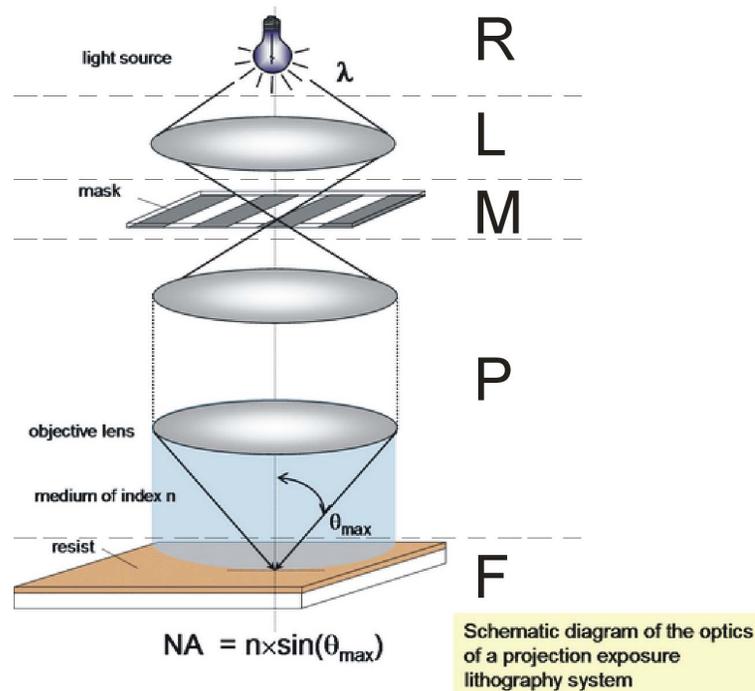


Рис. 9: Иммерсионная литография

Подробно химические процессы происходящие в фоторезисте и металле при экспонировании и травлении **F** в настоящей работе не рассматриваются, однако и там технологии не могут быть существенно улучшены и главной возможностью уменьшить минимальный размер элемента на чипе становится уменьшение константы процесса  $k$ . Для этого используются разнообразные приемы улучшения разрешения (RETs - Resolution Enhancement Techniques). Как правило, эти приемы корректируют волновой фронт (операторы **L** и **M** из (7)), а точнее, амплитуду фазу либо направление волны. Используется внеосевая подсветка - когда создается нужная форма освещения (квадрупольная, дипольная, кольцевая и другие подсветки) - в результате чего изменяется оператор **L**. Оптимизации оператора **M** - то есть проекционной маски, в последнее время является, пожалуй, одним из самых активно развивающихся областей исследований в литографии. Так в 1995 году Коббом и Захаром [18] был предложен метод оптических коррекций (OPC): топология чипа представляется как совокупность ребер и углов, после чего к трафарету полностью повторяющему исходную топологию добавляются корректирующие элементы - прямоугольники вдоль ребер и квадратики в вершинах.

Также при проектировании фазосдвигающих масок (PSM) используется

метод сдвига фазы (phase shift). Если соседние элементы на маске будут находиться в противофазе, это увеличит четкость результирующего изображения в объектной плоскости. Для объяснения этого эффекта можно воспользоваться следующим соображением: если функция принимает значение 1 на некотором элементе и -1 на соседнем, то между ними у нее где-то есть ноль, т.е. интенсивность в некоторой точке обнулится вследствие чего эти два соседних близко расположенных элемента хорошо разделятся (что очень важно, так как слияние соседних элементов-дорожек приведет к короткому замыканию в микрочипе). На практике сейчас используется сочетание методов OPC и phase shift (рис. 10).

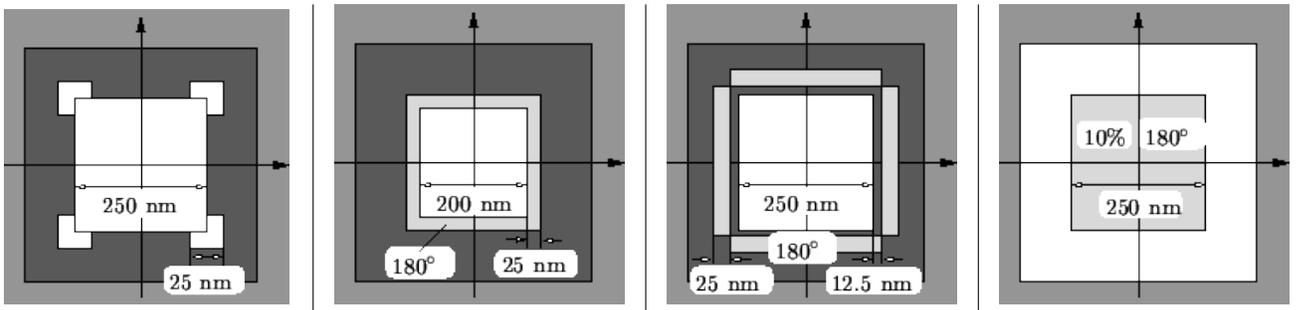


Рис. 10: Применения метода OPC для квадрата 250x250 нм

Ранее, когда разрешение CD было велико, литографическая установка работала в области геометрической оптики и маска просто гомотетично повторяла заданное распределение дорожек  $q_0$ . С уменьшением CD и развитием технологий фазосдвигающих (PSM) и пикселированных масок (PPM) масок особенно остро встает вопрос расчета таких масок. Теперь при построении моделей используется волновая оптика [1]-[3],[14],[11],[12] а рассчитанные по этим моделям маски порой совершенно не похожи на исходную топологию (рис. 11). При этом, поскольку, как правило, решить задачу (7) не представляется возможным, ставится задача

$$\rho(q_0 - \mathbf{q}(\mathbf{M})) \longrightarrow \min \quad (8)$$

где  $\rho$  - это некоторая функциональная норма. Рассмотрим модель "тонкой маски то есть не будем рассматривать эффекты, связанные с тем, что маска имеет ненулевую толщину. Тогда оператор задающий действие маски  $\mathbf{M}$  можно заменить матрицей пропускания  $B = \{b_{ij}\}$ :

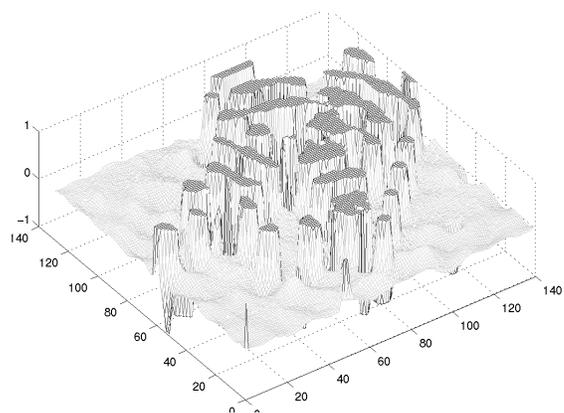
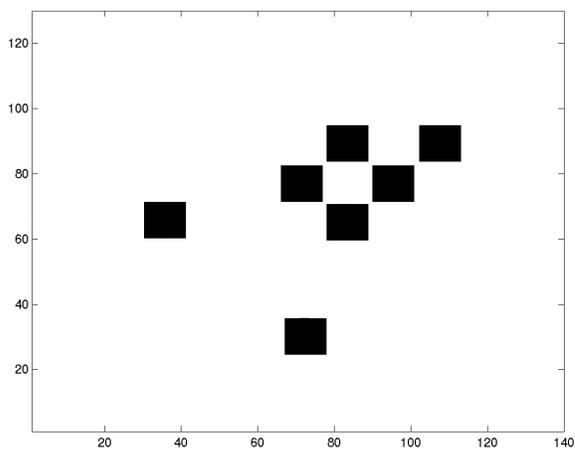
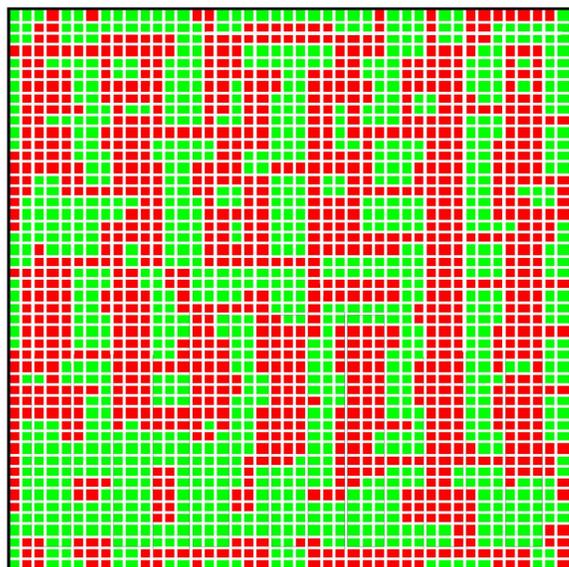
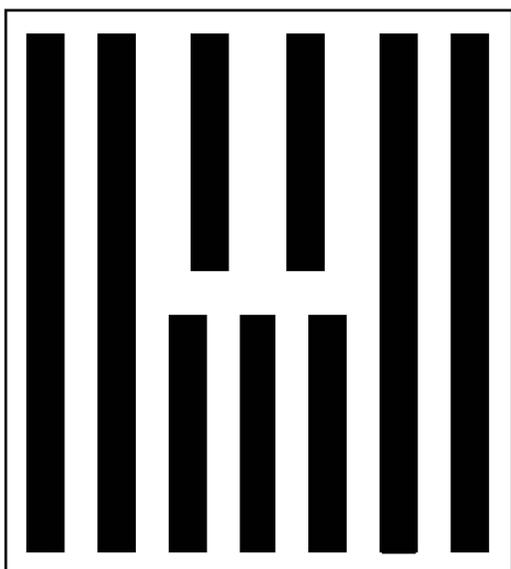


Рис. 11: Исходные топологии (слева) и итоговые маски (справа) из работ [2] и [14]

$$b_{ij} = B(\xi_i, \eta_j), \text{ где } \xi_i = -b/2 + (i-1)b/N, \eta_j = -b/2 + (j-1)b/N, i, j = 1, \dots, N \quad (9)$$

Если теперь маску представлять как набор из  $ij$  одинаковых квадратиков ( $i, j = 1, \dots, N$ ), на каждом из которых пропускание принимает некоторое значение  $b_{ij}$  из фиксированного для данного типа масок набора, то для матрицы пропускания  $B = b_{ij}$  задача (8) запишется следующим образом:

$$\rho(q_0 - \mathbf{q}(B)) \longrightarrow \min \quad (10)$$

Как правило, при решении (10) переходят к непрерывным коэффициентам пропускания  $m_{ij}$  и задача сводится к минимизации некоторой целевой функции  $J(\mathbf{m})$ :

$$J(\mathbf{m}) \longrightarrow \min \quad (11)$$

При решении задачи (11) используются градиентные методы [11][12], метод локальных вариаций [14] и другие оптимизационные подходы [3],[13].

В методе локальных вариаций [21] выбирается некоторая начальная голограмма (маска)  $T_0 = \mathbf{m}_0$ . Далее на  $k$ -ом шаге итерации в ячейке голограммы  $(i_k, j_k)$  рассматриваются две вариации функции пропускания  $\mathbf{m}$ :

1.  $\mathbf{m}_k^+$  - с увеличенным на  $\delta_{lv}$  по сравнению с  $\mathbf{m}_{k-1}$  значением пропускания в ячейке  $(i_k, j_k)$ ,
2.  $\mathbf{m}_k^-$  - с уменьшенным на  $\delta_{lv}$  по сравнению с  $\mathbf{m}_{k-1}$  значением пропускания в ячейке  $(i_k, j_k)$ .

Выбор  $k$ -го приближения  $\mathbf{m}_k$  будет осуществляться по следующему алгоритму:

$$\mathbf{m}_k = \begin{cases} \mathbf{m}_k^+, & \text{если } J(\mathbf{m}_k^+) < J(\mathbf{m}_{k-1}) \text{ и } J(\mathbf{m}_k^+) < J(\mathbf{m}_k^-) \\ \mathbf{m}_k^-, & \text{если } J(\mathbf{m}_k^-) < J(\mathbf{m}_{k-1}) \text{ и } J(\mathbf{m}_k^-) < J(\mathbf{m}_k^+) \end{cases} \quad (12)$$

В методе градиентного спуска [22] на каждом шаге  $\mathbf{m}_k$  выбирается следующим образом:

$$\mathbf{m}_k = \mathbf{m}_{k-1} - \delta_{gd} \nabla J(\mathbf{m}_{k-1}) \quad (13)$$

где  $\delta_{gd}$  - величина шага.

В [11] решается задача (10) для проекционной литографической установки и различных типов масок (см. таблицу на рис. 6). Моделируется литографический процесс фотоэкспонирования и травления резиста. В качестве модели фоторезиста принимается сглаженный вариант однопороговой модели (5):

$$q(h(\xi, \eta), \xi, \eta) = \frac{1}{1 + e^{-\alpha(h(\xi, \eta) - \theta)}} \quad (14)$$

Освещение считается когерентным и интенсивность  $h(\xi, \eta)$  моделируется по (1). Для того, чтобы оперировать только с гладкими функциями, авторы от коэффициентов пропускания  $b_{ij}$ , принимающих, например, для обычной ОРС маски значения 0 и +1, переходят сначала к непрерывным  $t_{ij}$ , а затем формулируют оптимизационную задачу для фиктивных коэффициентов  $m_{ij}$ , определенных на  $\mathbb{R}$ :

$$t_{ij} = \frac{1 + \cos(m_{ij})}{2}, \quad \text{где } i, j = 1, \dots, N \quad (15)$$

Окончательно

$$h(\xi_s, \eta_p, \mathbf{m}) = \left( \sum_{i,j=1}^N K_{ijsp} \frac{1 + \cos(m_{ij})}{2} \right)^2 \quad (16)$$

и профиль дорожек в металле

$$q(\xi_s, \eta_p, \mathbf{m}) = \frac{1}{1 + e^{-\alpha \left( \sum_{i,j=1}^N K_{ijsp} \frac{1 + \cos(m_{ij})}{2} \right)^2 + \alpha \theta}}, \quad s, p = 1, \dots, N^2 \quad (17)$$

В качестве целевой функции берется

$$J(\mathbf{m}) = F_{fid}(\mathbf{m}) + F_{aerial}(\mathbf{m}) + F_{bin}(\mathbf{m}) + F_c(\mathbf{m}) \quad (18)$$

Здесь  $F_{fid}(\mathbf{m})$  - мера отличия заданной топологии  $q_0$  от  $q$ :

$$F_{fid}(\mathbf{m}) = \sum_{i,j=1}^N (q_0(\xi_i, \eta_j) - q(\xi_i, \eta_j, \mathbf{m}))^2 \quad (19)$$

$F_{aerial}(\mathbf{m})$  - мера отличия интенсивности излучения  $h$  от топологии. Чем это отличие меньше, тем результирующее изображение будет устойчивее к отклонениям в фокусирующей системе и времени экспонирования фоторезиста.

$$F_{aerial}(\mathbf{m}) = \sum_{i,j=1}^N (q_0(\xi_i, \eta_j) - h(\xi_i, \eta_j, \mathbf{m}))^2 \quad (20)$$

$F_{bin}(\mathbf{m})$  - функция штрафа, отвечающая за бинаризуемость, то есть определяющая, какие будут потери в качестве изображения при замене непрерывной маски  $t_{ij}$  на бинарную  $b_{ij}$ . Так, для ОРС маски  $b_{ij}$  принимает значения 0, +1 и в качестве  $F_{bin}$  берется

$$F_{bin}(\mathbf{m}) = \sum_{i,j=1}^N [1 - (2t_{ij}(m_{ij}) - 1)^2] \quad (21)$$

$F_c(\mathbf{m})$  - функция штрафа за сложность топологии маски. Чем сложнее конфигурация отверстий на маске, тем значение этого функционала больше.

Оптимизационная задача

$$J(\mathbf{m}) \rightarrow \min \quad (22)$$

решается методом градиентного спуска (13). Результат для ОРС маски приведен на рис. 12.

Метод голографической литографии предлагает вместо проекционной маски использовать амплитудную голограмму, подсвеченную когерентной сферической сходящейся волной (рис. 5). Голограмма  $V(x,y)$  представляет собой огромное количество очень маленьких отверстий. Технологически она изготавливается методом электронно лучевой литографии на литографе. На вход литографу подается файл с координатами и, возможно (если отверстия неодинаковы), размерами отверстий.

В математической постановке задачи голограмма представляется либо как вещественная непрерывная функция пропускания  $T(x,y)$ , освещаемая волной  $Q$  с фиксированной длиной волны, - "серая голограмма либо как комплекснозначная непрерывная функция  $E(x,y)$  - "комплексная голограмма". После прохождения через плоскость  $\Gamma$  освещающее поле  $Q$  преобразуется в  $W(x,y)$ . Главная задача - это найти такую голограмму  $V(x,y)$ , что при освещении ее полем  $Q$  на фоторезисте будет сформирована такая

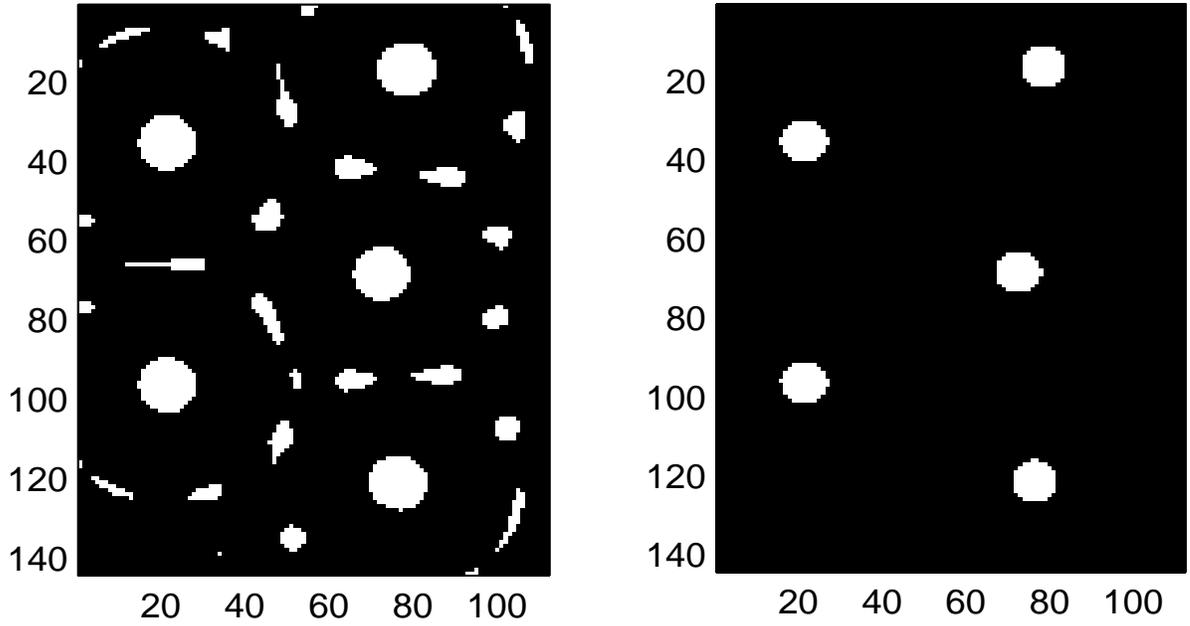


Рис. 12: OPC маска  $B(x,y)$ , рассчитанная градиентным методом (слева) и соответствующий профиль в металле  $q(\xi, \eta)$  из [11]

интенсивность  $h(\xi, \eta)$ , что после его обработки и травления, профиль дорожек в металле  $q(\xi, \eta)$  будет максимально близок к заданному  $q_0(\xi, \eta)$ :

$$\rho(q_0 - q(h(B, Q))) \rightarrow \min \quad (23)$$

Зависимость  $q(h)$  может быть получена экспериментально, либо взята из моделей фоторезиста (5)-(6). В настоящей работе на основе консультаций с технологами принято, что  $q(h)=h$  и основная задача:

$$\rho(q_0 - h(B, Q)) \rightarrow \min \quad (24)$$

Освещение в голографической голографии когерентно, поэтому для нахождения  $h$  используем (1):

$$J_\rho(B) \equiv \rho(q_0 - |[B(x, y)Q(x, y)] * K|^2) \rightarrow \min \quad (25)$$

Количество отверстий на реальных голограммах исчисляется триллионами, из-за этого решение задачи (25) потребовало бы огромных, несуществующих в настоящее время вычислительных ресурсов. Поэтому удобно переформулировать ее в терминах серой голограммы  $T(x,y)$ :

$$\rho(q_0 - |[T(x, y)Q(x, y)] * K|^2) \rightarrow \min \quad (26)$$

или

$$J_\rho(T) \equiv \rho(q_0(\xi, \eta) - \left| \int_{-\frac{b}{2}}^{\frac{b}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} T(x, y)Q(x, y)K(\xi - x, \eta - y)dx dy \right|^2) \rightarrow \min \quad (27)$$

Серая голограмма  $T(x, y)$  может быть преобразована в бинарную  $V(x, y)$  с помощью процедуры "бинаризации". Можно представить себе такую голограмму, которая в каждой точке нужным образом меняет и фазу и амплитуду. Она будет описываться комплекснозначной функцией  $E(x, y)$ . Тогда (24) примет вид:

$$\rho(q_0 - |[E(x, y)Q(x, y)] * K|^2) \rightarrow \min \quad (28)$$

или

$$J_\rho(E) \equiv \rho(q_0(\xi, \eta) - \left| \int_{-\frac{b}{2}}^{\frac{b}{2}} \int_{-\frac{b}{2}}^{\frac{b}{2}} E(x, y)Q(x, y)K(\xi - x, \eta - y)dx dy \right|^2) \rightarrow \min \quad (29)$$

Технологии изготовления таких комплексных голограмм пока не существует (хотя РРМ маски уже очень к ней близки), но в расчетах рассмотрение такой функции может быть полезно. Комплексную голограмму, полученную из (29) можно преобразовать в серую по формулам, вытекающим из метода Габора записи голограмм.

$$T(x, y) = \frac{1}{A_Q(x, y)}(CA_Q(x, y) + 2A_E(x, y)\cos(k(\Phi_E(x, y) - \Phi_Q(x, y)))) \quad (30)$$

где константа  $C \geq 2\max \left[ \frac{A_E(x, y)}{A_Q(x, y)} \right]$ .  $A_E(x, y)$ ,  $\Phi_E(x, y)$ ,  $A_Q(x, y)$ ,  $\Phi_Q(x, y)$  - амплитуды и фазы полей  $E$  и  $Q$  в плоскости  $\Gamma$ :

$$A(x, y) = A_E(x, y)e^{i\Phi_E(x, y)}, \quad Q(x, y) = A_Q(x, y)e^{i\Phi_Q(x, y)}$$

В результате такого преобразования результаты восстановления с серой и комплексной голограмм для участков изображения расположенных на достаточном расстоянии от оптической оси системы будут совпадать. После нахождения серой голограммы ставится задача преобразования непрерывной вещественной функции в искомую конфигурацию отверстий одинакового или разного размера - задача "бинаризации".

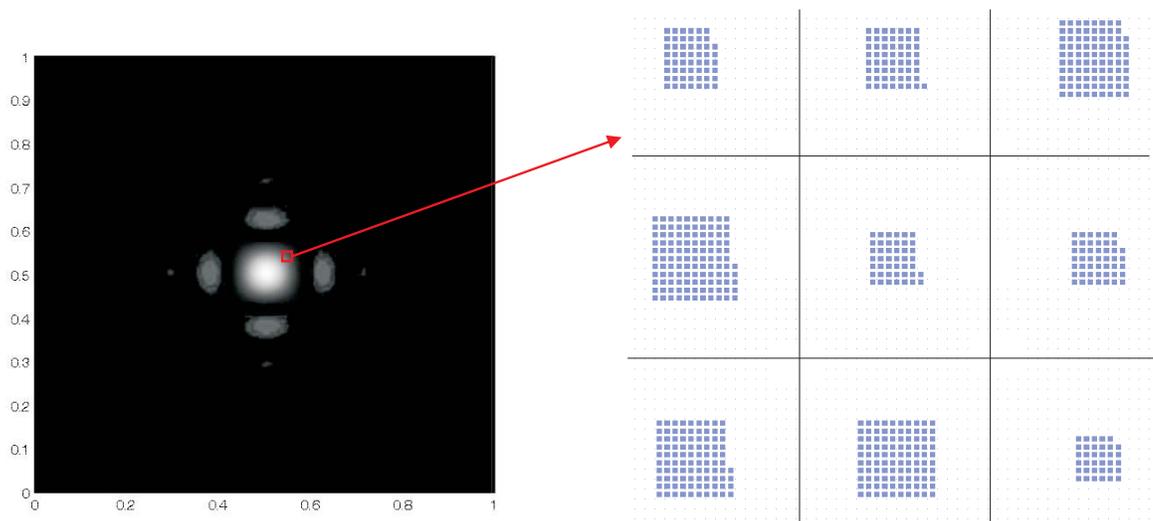


Рис. 13: Серая голограмма  $T(x,y)$  и участок соответствующей бинарной голограммы  $V(x,y)$

Разработанный алгоритм бинаризации разбивает серую голограммы на ячейки. Затем в ячейке на бинаризованной голограмме  $V(x,y)$  ставятся отверстия в количестве соответствующем средней яркости  $T(x,y)$ .

Качество изображения и бинаризуемость голограммы зависят от формы волны подсветки  $Q$ . Оказалось, что для параллельной волны качество восстановленного с нее изображения лучше, чем для сферической сходящейся в плоскости  $I$ , однако для успешной бинаризации гораздо лучше подходят голограммы, полученные для сферической волны  $Q$ .

Проводился численный эксперимент для выяснения при каком размере отверстий голограмма хорошо бинаризуется. На рис. 14 показано изображение квадрата  $10 \times 10$  мкм полученного с серой голограммы (сверху) и с бинаризованной (снизу). Бинарная голограмма  $V(x,y)$  состояли из 2'900'000 отверстий. Программа получения изображения по голограмме выполнялась на 128 процессорах суперкомпьютера МВС 100К МСЦ РАН, расчет занял 60 мин. В результате было показано, что при бинаризации одинаковыми

отверстиями размером  $1/2$  длины волны изображения  $h(\xi, \eta)$  полученные с серой и бинаризованной голограмм практически совпадают (рис. 14), при размере отверстия, равном длине волны отличаются, а при размере отверстия, равном двум длинам волн - отличаются очень сильно.

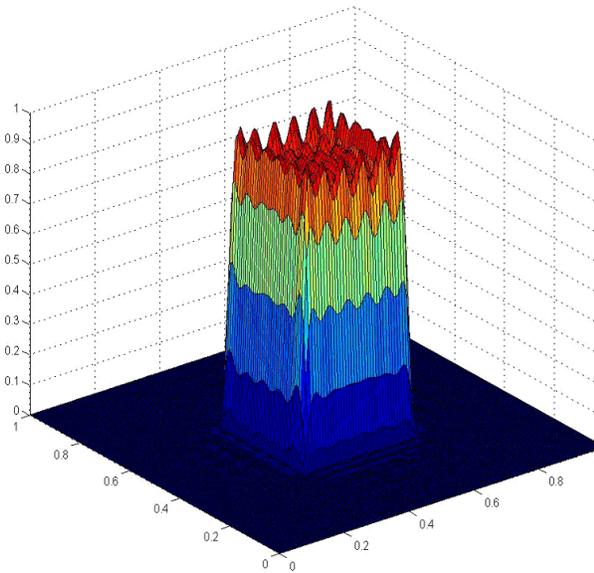
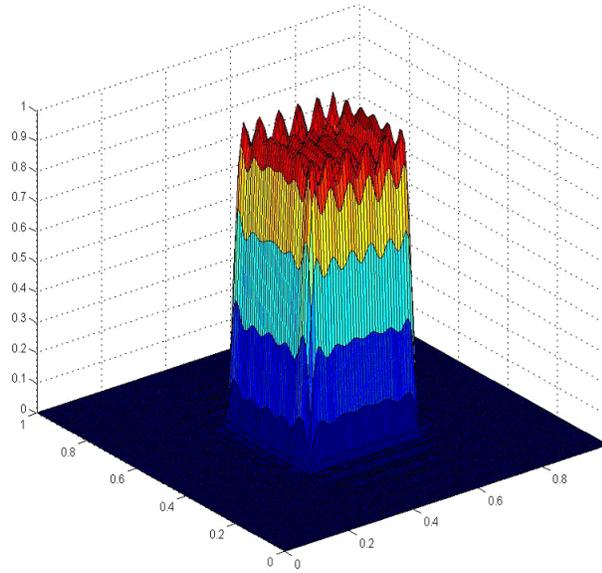


Рис. 14: Бинаризация отверстиями размером  $1/2$  длины волны

В работах [1]-[3] подробно описана предложенная в 2008 году методика создания и расчета фазовых масок. Маска здесь является уже не гомотетически подобным требуемой топологии трафаретом, а представляет собой огромное количество маленьких квадратиков (пикселей). Фаза на каждом таком пикселе постоянна и принимает значение из некоторого фиксированного набора. Ключевым моментом здесь является расчет необходимой конфигурации фазированных пикселей [2], [3]. В ходе расчета авторы вариационными методами, методом наименьших квадратов, либо градиентным методом решают задачу (11).

Исходная постановка существенно отличается от постановки задачи голографической литографии поскольку в проекционной литографии используется частично когерентное излучение (4), а в голографической - когерентное. На рис. 15 показаны результаты работы оптимизационных алгоритмов в зависимости от типов масок и начальных приближений. В первом ряду показаны начальные маски, во втором - маски, полученные в результате оптимизации, а в нижнем ряду идеальные профили в металле  $q_0$  с наложенными на них профилями  $q$ , получающимися при использовании соответствующих масок (выделены красным).

Как видно из рис. 15 результат оптимизации сильно зависит от выбора начальной конфигурации. В настоящем исследовании в качестве начального приближения берется голограмма, рассчитанная по формулам, полученным из метода Габора. Этот метод был предложен еще в 1947 г. [15],[16], однако получил распространение лишь после появления лазера в 1960 г., и только в 1971 г. Дэннис Габор получил Нобелевскую премию по физике "за изобретение и развитие голографического метода". В первых работах по голографической литографии для формирования ДОЭ на фотопластинке применялся именно этот способ [6]. Схема используемой при этом экспериментальной установки показана на рис. 16.

Идущее от лазера плоское излучение разделяется с помощью зеркала 3 на освещающую объект 4 волну 1 и опорную волну 2. Предметное поле 5, полученное после отражения волны 1 от объекта, и опорная волна 2 интерферируют в плоскости фотопластинки А, создавая на светочувствительном слое некоторую картину пропускания (серую голограмму). Если получившийся дифракционный оптический элемент А подсветить волной, сопряженной к опорной волне волной 6 (т.е. волной с тем

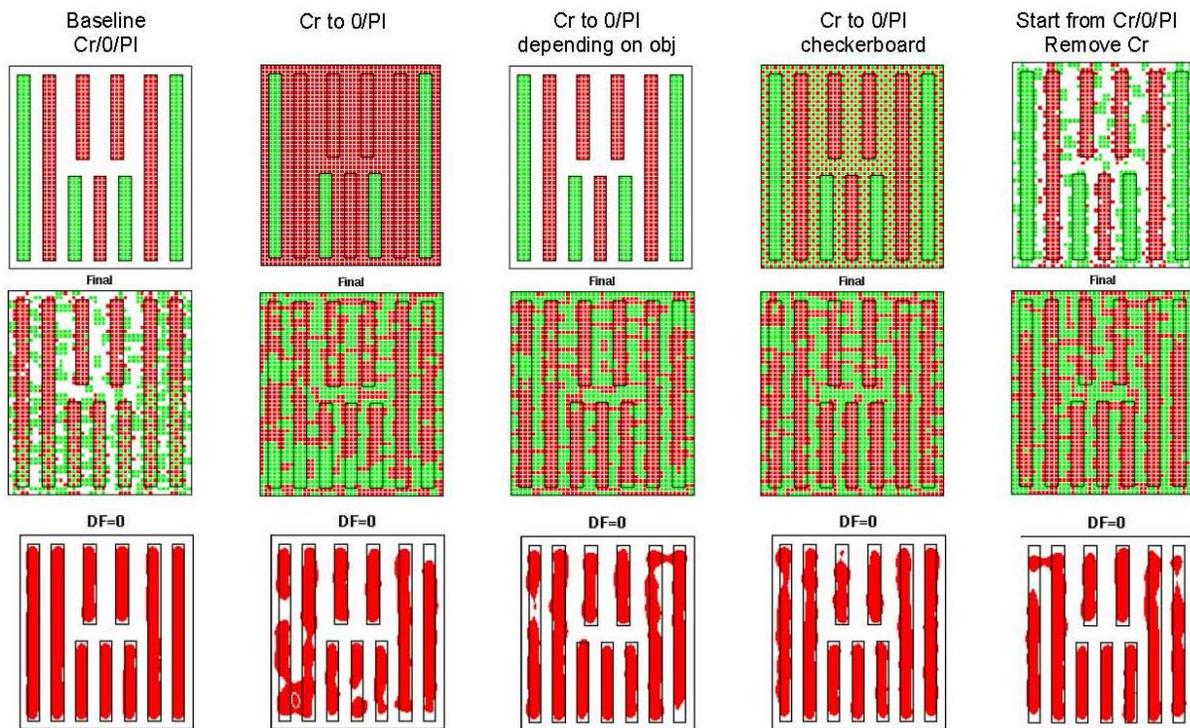


Рис. 15: Результаты оптимизации для различных масок

же ходом лучей, но распространяющейся в противоположную сторону), то сформируется несколько полей, одно из которых будет близко к полю 5 и на границе объекта создастся распределение интенсивности такое же, какое было там в момент записи голограммы.

В настоящей работе используется похожий подход. В качестве опорной берется сферическая сходящаяся в плоскости объекта волна. Такая форма фронта была выбрана с целью улучшить бинаризуемость серой голограммы: при использовании плоской подсветки теоретически результат восстановления получается лучше, но период колебаний функции пропуска на ДОЭ оказывается слишком маленьким, что приводит к невозможности передать функцию серости маленькими отверстиями одинакового размера.

Рассмотрим подробнее процесс расчета серой голограммы по методу Габора (рис. 17).

Пусть на слое металла в плоскости изображения И требуется создать профиль дорожек  $q_0$ . В качестве подсветки 1 возьмем плоскопараллельную волну R, принимающую в плоскости И значение  $R(\xi, \eta)$ . Зададим предметное поле V в плоскости И следующей комплексной амплитудой:  $V(x, y, z) |_{O=}$

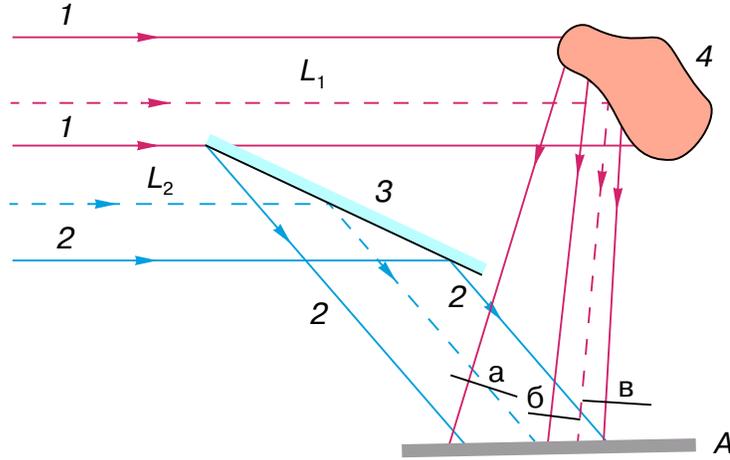


Рис. 16: Схема установки, записывающей голограмму по методу Габора

$A_V(\xi, \eta)e^{i\Phi_V(\xi, \eta)} = q_0(\xi, \eta)R(\xi, \eta)$ . (Это построение соответствует подсветке параллельным пучком лазера  $R$  трафарета топологии  $q_0$ , находящегося в плоскости  $O$ .) Предметное поле  $V$  в плоскости голограммы:

$$V(x, y) = \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{a}{2}}^{\frac{a}{2}} K(x - \xi, y - \eta, d) q_0(\xi, \eta) R(\xi, \eta) d\xi d\eta \quad (31)$$

Пусть предметное поле  $V$  в плоскости голограммы  $\Gamma$  имеет амплитуду  $A_V(x, y)$  и фазу  $\Phi_V(x, y)$ . В качестве комплексной голограммы можно взять

$$E(x, y) = A_V(x, y)e^{-i\Phi_V(x, y)} \quad (32)$$

Для создания серой голограммы Габора в качестве опорной волны  $Q^*$  берется сферическая волна с центром в точке пересечения плоскости  $O$  и оси  $z$ . Опорная волна в плоскости  $\Gamma$  имеет амплитуду  $A_Q(x, y)$  и фазу  $\Phi_Q(x, y)$ . Возьмем следующую функцию пропускания голограммы  $T(x, y)$ :

$$T(x, y) = \frac{1}{A_Q(x, y)} (CA_Q(x, y) + 2A_V(x, y)\cos(k(\Phi_V(x, y) - \Phi_Q(x, y)))) \quad (33)$$

где константа  $C$  выбирается так, чтобы выражение в скобках было неотрицательным ( $C \geq 2\max \left[ \frac{A_V(x, y)}{A_Q(x, y)} \right]$ ). После освещения такой голограммы

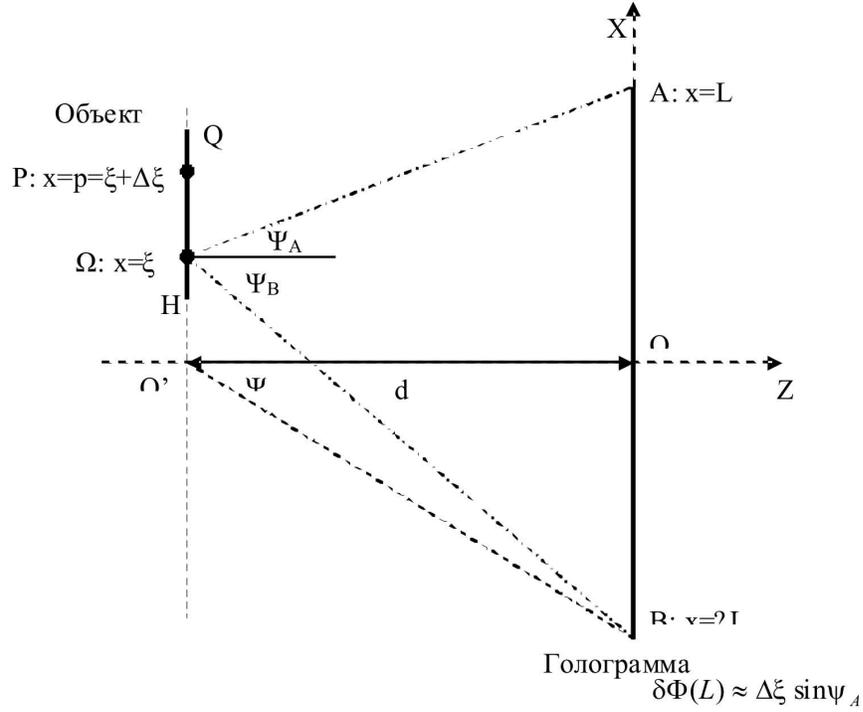


Рис. 17: Схема метода Габора в цифровой голографической литографии

сопряженной к  $Q^*$  восстанавливающей волной  $Q = A_Q(x, y)e^{-ik\Phi_Q(x, y)}$  на второй стороне голограммы получим поле  $W$ :

$$W = e^{-ik\Phi_Q(x, y)}(CA_Q(x, y) + 2A_V(x, y)\cos[\Phi_V(x, y) - \Phi_Q(x, y)]) \quad (34)$$

Поле  $W$  состоит из трех частей  $W_1$ ,  $W_2$  и  $W_3$ :

$$W_1 = A_V(x, y)e^{-ik\Phi_V(x, y)}$$

$$W_2 = CA_Q(x, y)e^{-ik\Phi_Q(x, y)}$$

$$W_3 = A_V(x, y)e^{ik(\Phi_V(x, y) - 2\Phi_Q(x, y))}$$

$W_1$  - волна комплексно сопряженная к предметному полю  $V$ ,  $W_2$  - паразитная волна, создающая в  $O$  крест с сильным пиком в начале координат,  $W_3$  - паразитное поле, которое при малых углах дифракции будет в плоскости  $O$  симметрично относительно начала координат повторять исходный объект.

Пусть объект имеет топологию, изображенную на рис. 18. Тогда после создания серой голограммы по формулам (31),(33) и ее освещения опорной волной  $Q$  в плоскости объекта  $O$  будем иметь поле в котором легко выделить элементы соответствующие полям  $W1, W2, W3$  на  $\Gamma$ . В левом верхнем углу рис. 19 в точности создается изображение как от комплексной голограммы (32), в правом нижнем углу - симметричное относительно оптической оси паразитное изображение, вдоль осей координат - паразитный крест с пиком в точке пересечения оптической оси и плоскости  $O$ .

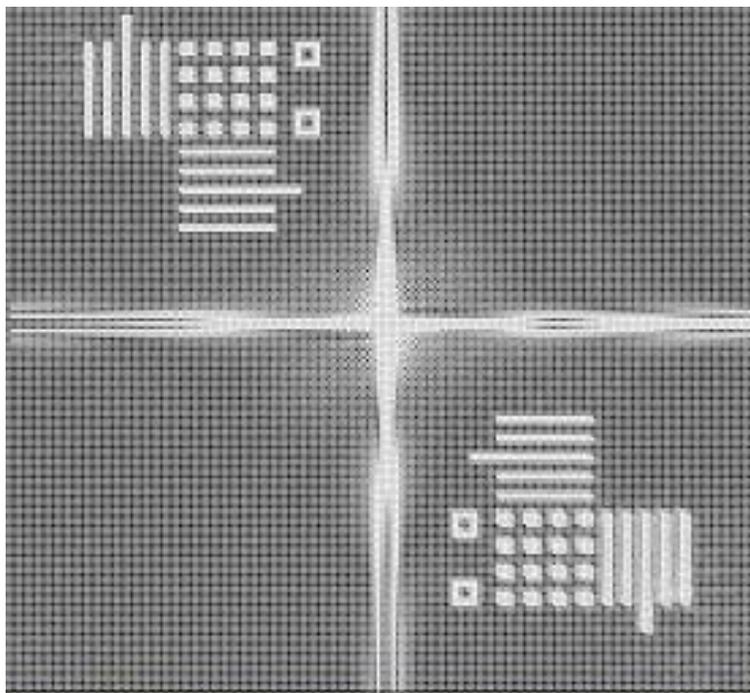


Рис. 19: восстановление с голограммы Габора

Програмный комплекс, реализованный в ходе проведения данного исследования кроме блока создания голограмм по формулам (31),(32),(33), имеет так же блок "восстановления" - когда по уже существующей бинарной, серой либо комплексной голограмме строится изображение - интенсивность электромагнитного излучения в плоскости фоторезиста  $h(\xi, \eta)$ . Так, для серой голограммы  $T(x,y)$ , подсвеченной волной  $Q$ , имеем:



Рис. 18: Объект

$$h(\xi, \eta) = \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{a}{2}}^{\frac{a}{2}} K(\xi - x, \eta - y, d) T(x, y) Q(x, y) dx dy \quad (35)$$

Аналогично для бинарных и комплексных голограмм:

$$h(\xi, \eta) = \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{a}{2}}^{\frac{a}{2}} K(\xi - x, \eta - y, d) B(x, y) Q(x, y) dx dy \quad (36)$$

$$h(\xi, \eta) = \int_{-\frac{a}{2}}^{\frac{a}{2}} \int_{-\frac{a}{2}}^{\frac{a}{2}} K(\xi - x, \eta - y, d) E(x, y) Q(x, y) dx dy \quad (37)$$

В качестве примера объекта средней по размеру топологии была взята используемая в микроэлектронике тестовая мира - топология, изображенная на рис. 20

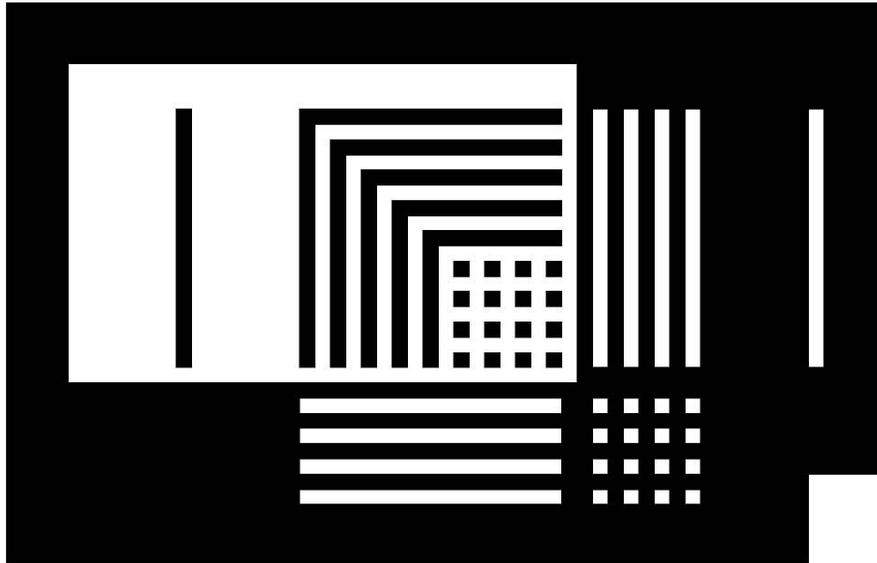


Рис. 20: Тестовая мира

Расчет серой голограммы (рис. 21) тестовой мира (рис. 20) по формулам (31),(33) проводился на 512 ядрах суперкомпьютера МВС100К МСЦ РАН и занял 3 секунды. Расчет восстановленного изображения по формуле (35) занял около 9 минут, результаты этого расчета приведены на рисунке 22.

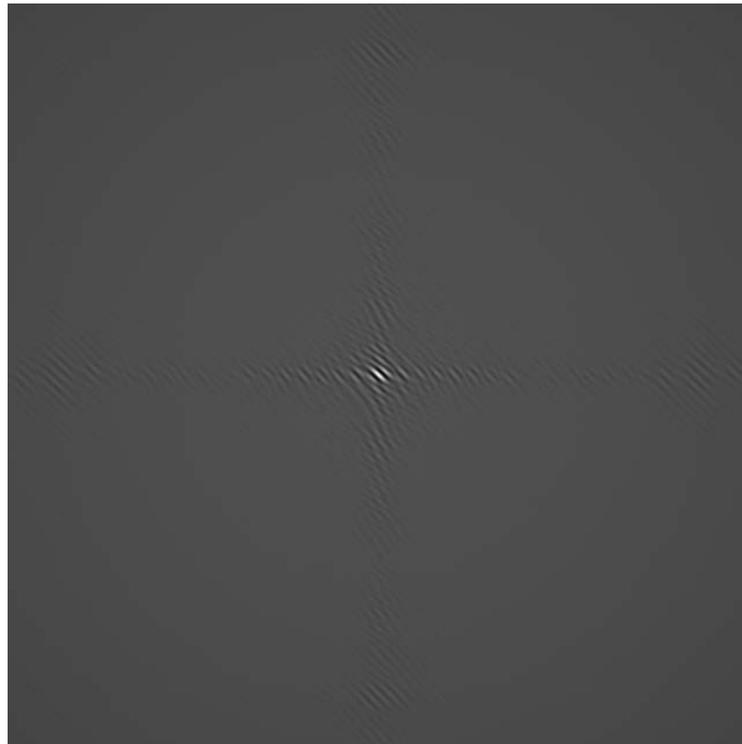
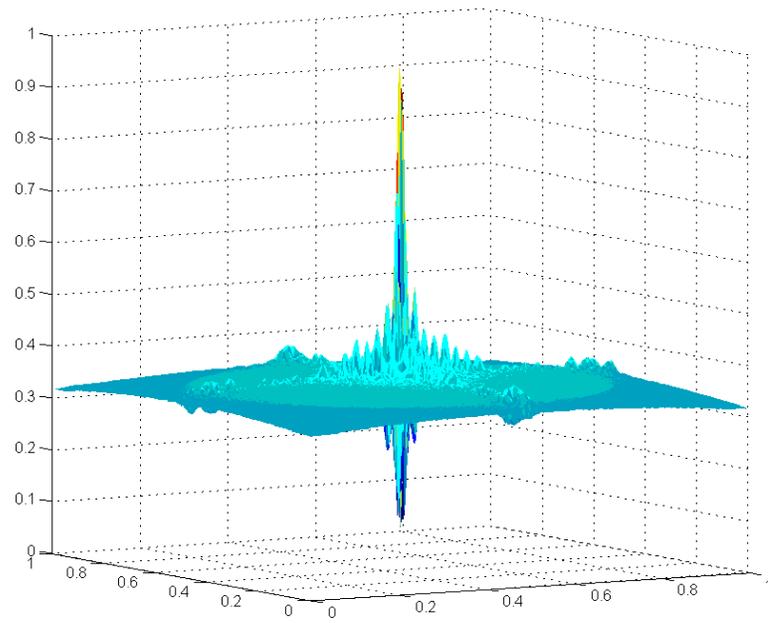


Рис. 21: Серая голограмма вид сверху и сбоку.

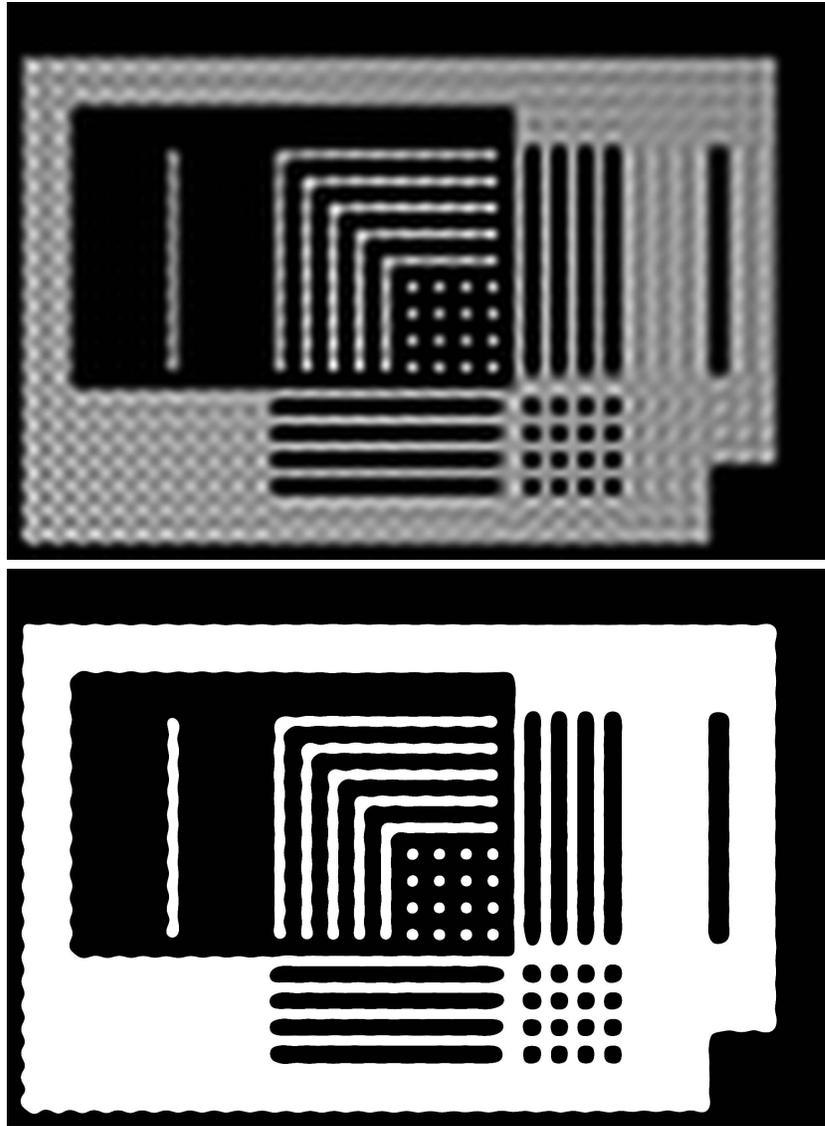


Рис. 22: Результат восстановления с серой голограммы рис. 21 (вверху) и соответствующий профиль дорожек в металле  $q_0$  (внизу). Использована однопороговая модель фоторезиста (5) с  $\theta=0.3$ .

Как видно из рис. 22 качество изображения, восстановленного с голограммы, созданной по методу Габора, оставляет желать лучшего. Поэтому серая голограмма Габора берется за первое приближение  $T_0(x, y)$  и далее методом локальных вариаций (12) решается задача (27), где норма  $\rho(f) = \|f\|_{L_1}$ . В качестве примера применения метода возьмем простой объект - квадрат  $10 \times 10$  мкм. Изображение, восстановленное с серой голограммы Габора (рис. 23) имеет колебания в белом поле более 10%, что превышает допустимый 5%-ный порог. Результат, полученный после 30-ти итераций оптимизационного алгоритма

приведен на рис. 24: неравномерность белого поле стала меньше 5%, амплитуда в черном поле осталось незначительной.

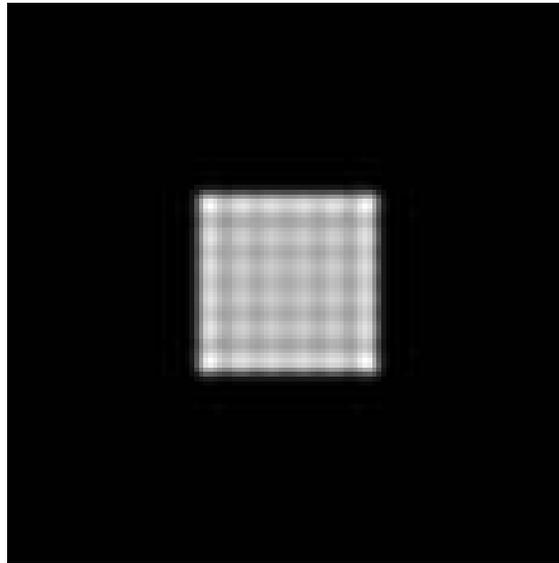
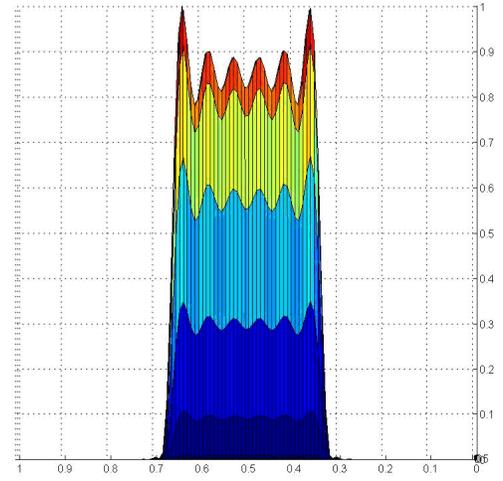
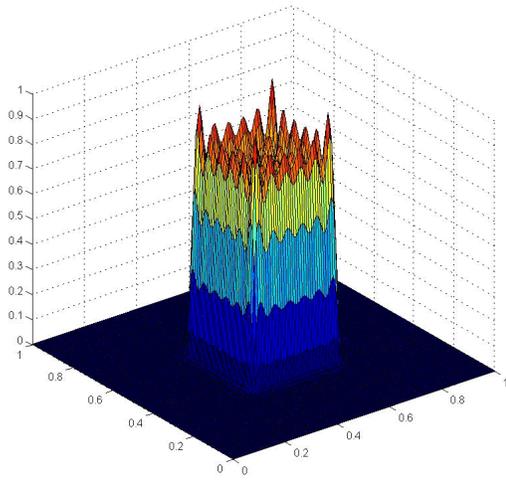


Рис. 23: Восстановление с неоптимизированной голограммы  $T_0$ . Исходная опология - квадрат  $10 \times 10$  мкм на поле  $30 \times 30$  мкм.

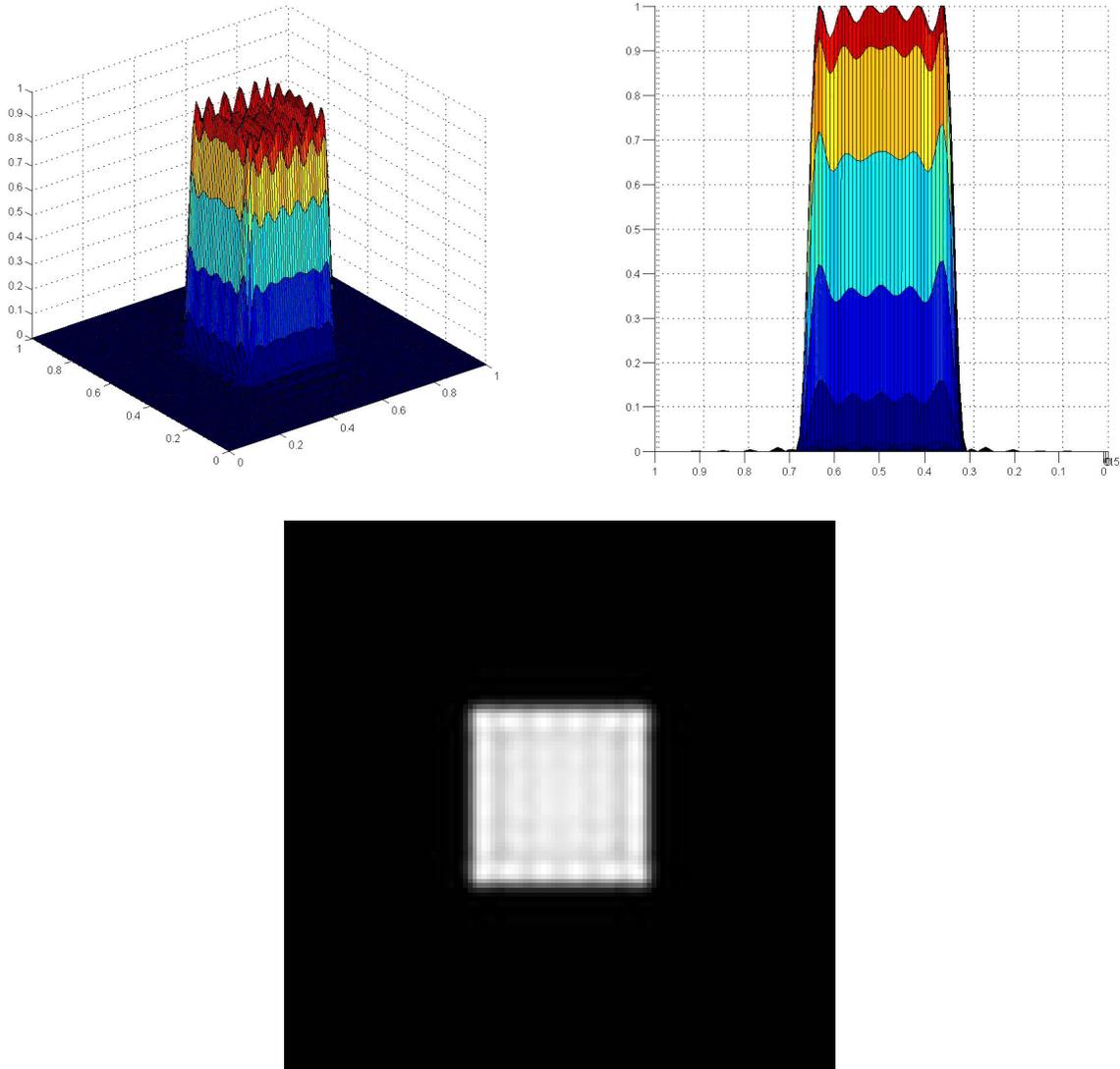


Рис. 24: Восстановление с голограммы, оптимизированной методом локальных вариаций. Исходная топология - квадрат 10x10 мкм на поле 30x30 мкм.

Другой результат относится к сравнению оптимизации качества изображения простого объекта (прямоугольник, уголок) методом локальных вариаций (12) и методом градиентного спуска (13). Метод градиентного спуска показал большие возможности и ряд преимуществ по сравнению с методом локальных вариаций.

Идеальное изображение представляет собой уголок на фоне квадрата 20x20 микрон. На рис. 25 приведены функции интенсивности  $h(\xi, \eta)$  в плоскости фоторезиста, полученные после оптимизации начального приближения Габора

методами локальных вариаций и градиентного спуска, соответственно (метод локальных вариаций отмечен на рисунках буквами ЛВ, а метод градиентного спуска - буквами ГС).

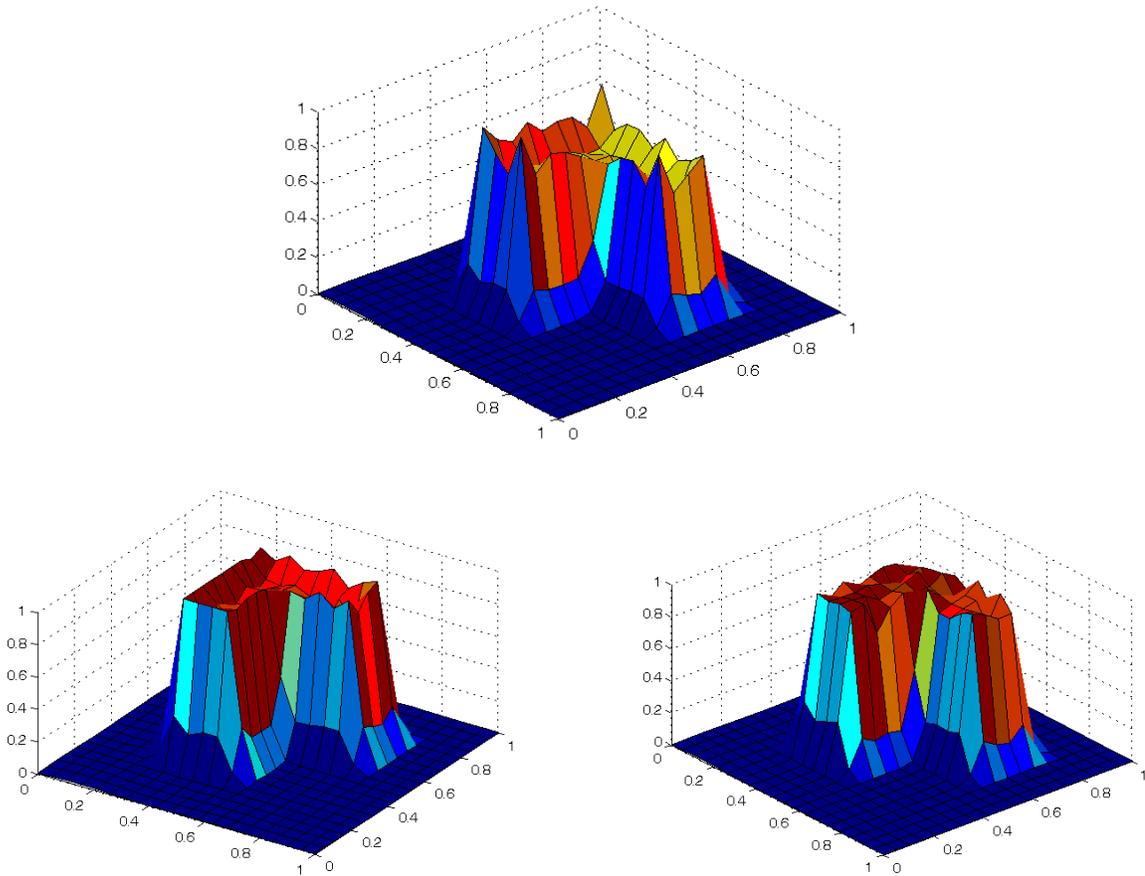


Рис. 25: Изображение, восстановленное с голограммы созданной по методу Габора (вверху), с голограммы оптимизированной локальными вариациями (внизу слева) и с голограммы оптимизированной методом градиентного спуска (внизу справа).

Далее (рис. 26) в некоторых условных единицах показаны последовательности норм погрешностей, полученные по ходу применения итераций для методов локальных вариаций и градиентного спуска, соответственно. Видно, что метод градиентного спуска приводит к существенно меньшему значению погрешности. Как видно также из рисунков, уменьшение погрешности при применении метода градиентного спуска связано с подъемом при оптимизации этим методом "дна" графика яркости внутри "уголка". Метод же локальных вариаций "застревает" в локальном минимуме функционала и "дно" функции яркости остается несколько "опущенным".

Метод градиентного спуска применялся для двух различных норм отличия идеального изображения от корректируемого- для нормы пространства  $L_1$  (график отмечен буквами ГС1) и нормы пространства  $L_2$  (график отмечен буквами ГС2). Видно, что использование второй нормы приводит к заметно лучшим результатам - невязка улучшается примерно на 25%.

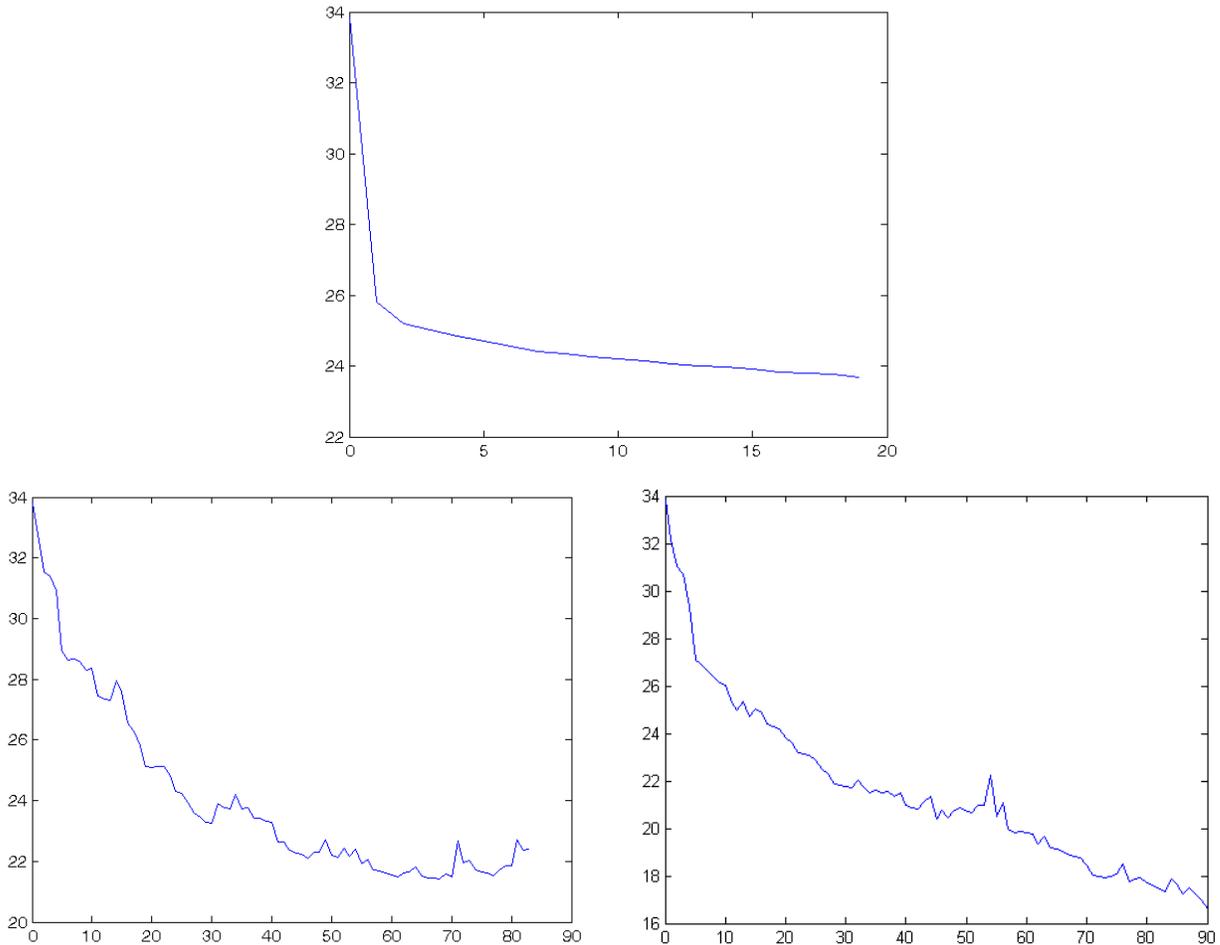


Рис. 26: Уменьшение нормы отклонения изображения от идеального при оптимизации по методу локальных вариаций (сверху) и градиентного спуска (снизу).

Метод локальных вариаций применялся ранее для создания обычных масок. Так в работе [14] минимизируется функционал

$$J(\mathbf{m}) = \|\sqrt{w}(h(\mathbf{m}) - h_0)\|_2^2 + \alpha_1 \|L_1 \mathbf{m}\|_2^2 + \alpha_2 \|L_2 \mathbf{m}\|_2^2 + \alpha_3 \|\mathbf{m} - \mathbf{m}_0\|_2^2 \rightarrow \min \quad (38)$$

где  $L_1, L_2$  - операторы первых и вторых производных,  $w$  - некоторая весовая функция, отражающая важность отдельных элементов топологии, а

$\alpha_i$  - весовые коэффициенты. Результат работы оптимизационного алгоритма представлен на рис. 27.

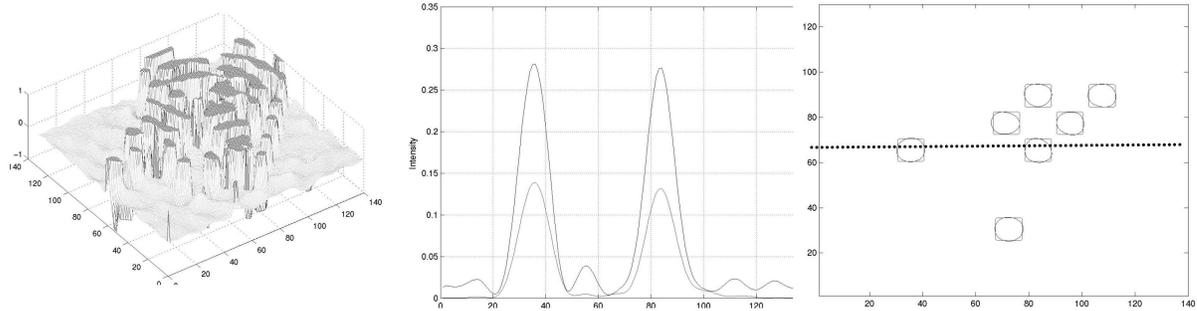


Рис. 27: Оптимизированная PSM-маска из работы Граника [14] (слева) и соответствующий ей профиль в слое металла (справа). На рисунке в центре показана интенсивность излучения в плоскости чипа вдоль линии, отмеченной пунктиром на правом рисунке. График интенсивности, лежащий выше, соответствует оптимизированной PSM-маске, ниже - обычной маске-трафарету.

Все приведенные ранее примеры отличались тем, что маски имели небольшие размеры, а количество элементов на них не превышало и одного десятка, однако в реальных микросхемах количество элементов исчисляется десятками миллиардов. Это, конечно, не сильно усложняло расчет для проекционной литографии, когда маска была просто гомотетично подобна топологии, либо когда нужно было внести оптические коррекции по заранее определенным правилам. В современных же методах, когда появляется необходимость, например, решать оптимизационную задачу для функций от миллиардов переменных, остро встает вопрос вычислительных мощностей. Для обычных масок, в силу локальной зависимости профиля дорожек  $q(\xi, \eta)$  от маски, как правило эту проблему легко решить за счет использования обычных кластерных суперкомпьютеров с раздельной памятью: расчеты для отдельных участков маски  $B(x, y)$  практически независимы друг от друга, вследствие чего алгоритм хорошо распараллеливается и дает линейный прирост производительности при увеличении количества используемых вычислительных узлов [2]. Участки, на которые разбивается маска, берутся с небольшими перекрытиями, а после оптимизации они "сшиваются"; при этом в месте стыка дополнительно проводится оптимизация (рис. 28).

Совсем иначе дело обстоит в голографической литографии. Здесь интенсивность  $h(\xi, \eta)$  в точке  $(\xi_0, \eta_0)$  на плоскости фоторезиста зависит от всей голограммы  $B(x, y)$  и поэтому вычислительная трудоемкость задачи

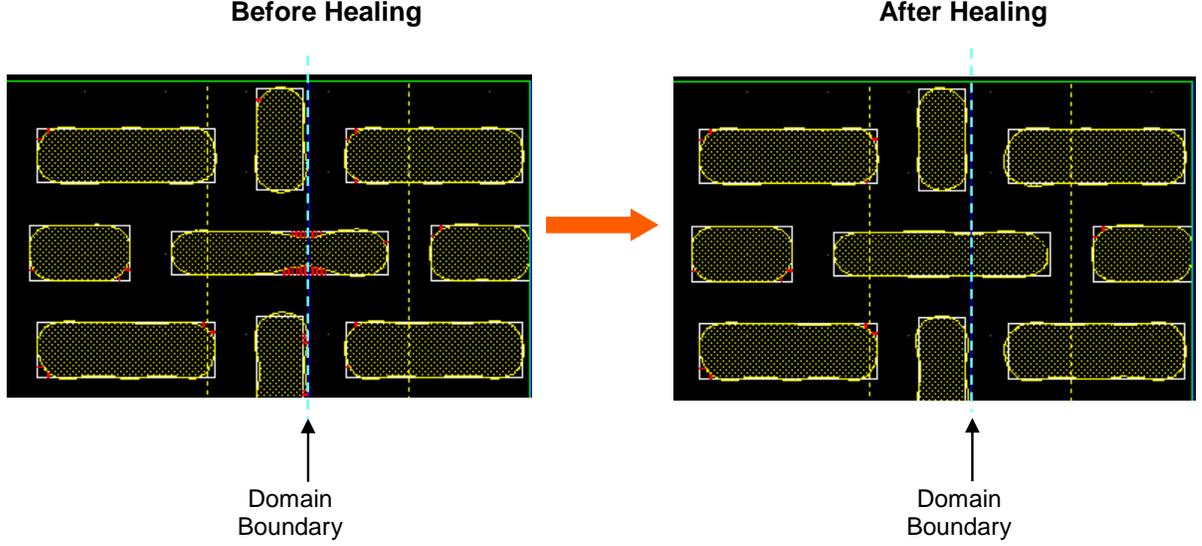


Рис. 28: Оптимизация в месте сшивки двух областей PPM маски [2]

расчета теперь не будет линейно зависеть от площади топологии. При этом существенную сложность представляет уже расчет первого приближения серой голограммы  $T_0(x, y)$  по формулам Габора (31), (33). Необходимо также уметь восстанавливать изображение с серой голограммы по формуле (35) и с бинарной по формуле (36) с шагом в несколько раз меньшим, чем минимальный характерный размер топологии. Это необходимо как для моделирования эксперимента, так и для расчета градиента при оптимизации по методу (13). Самым трудоемким в этих задачах является расчет интегралов сверточного типа:

$$K * f \equiv \iint K(\xi - x, \eta - y) f(\xi, \eta) d\xi d\eta. \quad (39)$$

Для произведения расчетов изображение разбивается на на сетку  $N \times N$  с шагом  $\sigma$ , а голограмма на сетку  $M \times M$  с шагом  $\gamma$  (рис. 29 ). Пусть минимальный размер элемента топологии равен  $\sigma_0$ , а размер ячейки голограммы -  $\gamma_0$ . Так как функция  $K$  сильно осцилирует, квадрат с длиной стороны  $a$ , на котором задается исходное изображение  $q_0(\xi, \eta)$  разбивается на сетку  $N_\lambda \times N_\lambda$  с шагом  $\sigma_\lambda$  в несколько раз меньшим длины волны лазера  $\lambda$ :  $\sigma_\lambda = \lambda/5$ . Голограмма разбивается на сетку  $M_\lambda \times M_\lambda$  с шагом равным размеру ячейки голограммы  $\gamma_\lambda = \gamma_0$ . Тогда  $N_\lambda = 5a/\lambda$ ,  $M_\lambda = b/\gamma_0$  и при прямом расчете результата свертки  $W(x, y) \equiv (K * f)(x, y)$  методом простого суммирования

необходимо вычислять суммы

$$W(k, l) = \sum_{ij} K_{ij}(k, l) q_0(i, j), \quad \text{при этом } \sigma = \lambda/5, \gamma = \gamma_0 \quad (40)$$

где принято  $x_k = \gamma * (k + 1/2)$ ,  $y_l = \gamma * (l + 1/2)$ ,  $\xi_i = \sigma * (i + 1/2)$ ,  $\eta_j = \sigma * (j + 1/2)$ , для любой функции  $v$ :  $v(k, l) = v(x_k, y_l)$ , а  $v(i, j) = v(\xi_i, \eta_j)$ .

Количество операций составит

$$\Upsilon_\lambda = O(a^2 b^2) \quad (41)$$

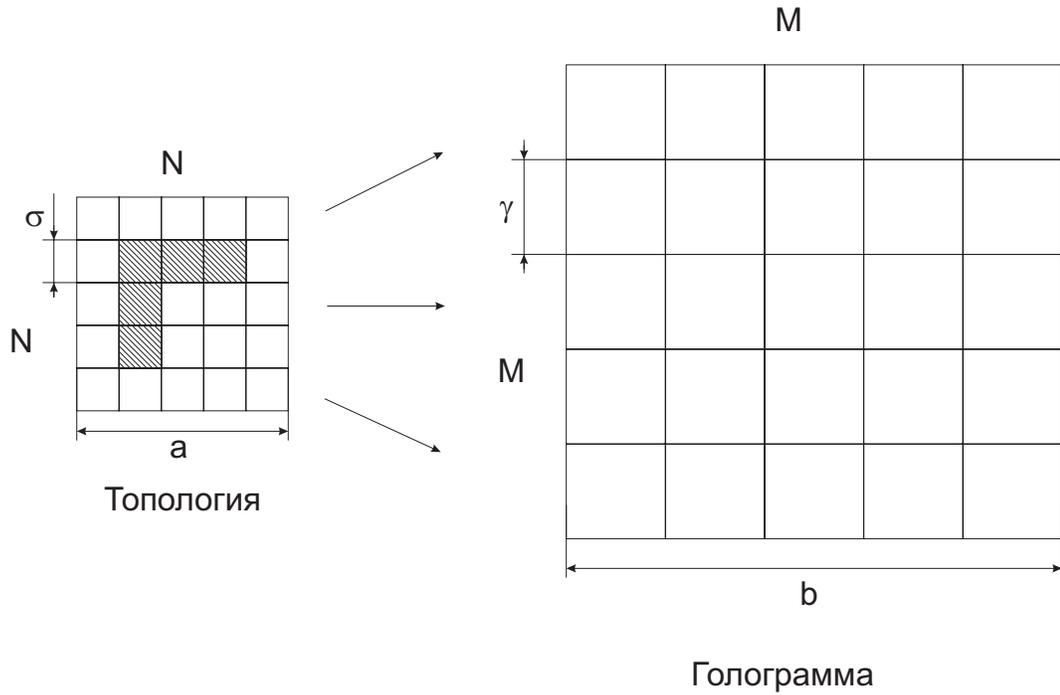


Рис. 29: Сетки разбиения на топологии и голограмме для расчета свертки (39).

Если теперь использовать то обстоятельство, что ядро  $K_{ij}$  имеет вид  $K_{ij}(k, l) = K(k - i, l - j)$ , то

$$W(k, l) = \sum_{ij} K(k - i, l - j) q_0(i, j), \quad \text{при этом } \sigma = \lambda/5, \gamma = \gamma_0 \quad (42)$$

Для вычисления свертки (42) применяется быстрое дискретное преобразование Фурье:

$$\mathbf{W} = F^{-1}[F(\mathbf{K}) \otimes F(\mathbf{Q})] \quad (43)$$

где матрицы  $\mathbf{W}, \mathbf{K}, \mathbf{Q}$  соответствуют функциям  $W(k,l), K(p,m), q_0(i,j)$ ,  $\otimes$  обозначает покоординатное умножение матриц,  $F$  - оператор преобразования Фурье:

$$F(f_L(m,n)) = g_L(p,q) = \sum_{m=s}^{L+s-1} \sum_{n=t}^{L+t-1} f_L(m,n) e^{\frac{2\pi i(mp+nq)}{N}} \quad (44)$$

а  $F^{-1}$  - оператор обратного преобразования Фурье:

$$F^{-1}(g_L(p,q)) = f_L(m,n) = \frac{1}{N^2} \sum_{p=s}^{L+s-1} \sum_{q=t}^{L+t-1} g_L(p,q) e^{-\frac{2\pi i(mp+nq)}{N}} \quad (45)$$

где  $f_L(m,n), g_L(p,q)$  - последовательности, периодичные с периодом  $L$  по обеим координатам.

При использовании БПФ шаг на голограмме придется взять такой же, как и на изображении:  $\gamma_{\text{БПФ}} = \sigma_{\text{БПФ}} = \lambda/5$ . Количество операций:

$$\Upsilon_{\text{БПФ}} = O(ab \ln(ab)) \quad (46)$$

Интеграл (39) можно представить как сумму интегралов по квадратикам с размером стороны, равным минимальному размеру на топологии  $\sigma_0$ :

$$W(x,y) = \sum_{ij} \iint_{\square_{ij}} K(\xi - x, \eta - y) q_0(\xi, \eta) d\xi d\eta \quad (47)$$

Можно аналитически, пользуясь приближением "дальней зоны", вычислить интеграл на квадратике  $(i,j)$ . Учитывая, что  $q_0$  принимает только значения 0 и 1, получим

$$\iint_{\square_{ij}} K(\xi - x, \eta - y) q_0(\xi, \eta) d\xi d\eta = \begin{cases} \iint_{\square_{ij}} K(\xi - x, \eta - y) d\xi d\eta, & \text{при } q_0(\xi_i, \eta_j) = 1 \\ 0, & \text{при } q_0(\xi_i, \eta_j) = 0 \end{cases} \quad (48)$$

Обозначим

$$S_{ij}(x, y) = \iint_{\square_{ij}} K(x - \xi, y - \eta) q_0(\xi, \eta) d\xi d\eta \quad (49)$$

Тогда интеграл (39) равен:

$$W(k, l) = \sum_{ij} S_{ij}(k, l) q_0(i, j), \quad \text{при этом } \sigma = \sigma_0, \gamma = \gamma_0 \quad (50)$$

То, что шаг по сетке теперь в несколько раз больше позволяет для метода аналитического суммирования (50) уменьшить количество операций примерно в 25 раз по сравнению с методом простых сумм (41):

$$\Upsilon_{\square} = \Upsilon_{\lambda}/25 \quad (51)$$

Для маленьких топологий при принятых в данном исследовании параметрах получается  $\Upsilon_{\square} \approx \Upsilon_{\text{БПФ}}$ , что позволяет производить расчеты с использованием (49) на обычном компьютере. При использовании суперкомпьютера методом аналитического приближения (50) удастся рассчитывать голограммы и для средних топологий (рис. 22), однако, с увеличением размера исходного изображения время работы всех описанных выше алгоритмов становится непоправимо большим.

Чтобы избежать таких трудоемких расчетов, можно воспользоваться тем, что топология состоит лишь из нескольких элементарных элементов: прямоугольников со сторонами, параллельными осям координат, квадратов и прямоугольников, наклоненных относительно координатных осей на  $45^\circ$  и голограмму сложной топологии составлять простым суммированием комплексных голограмм  $E_i(x, y)$  элементарных элементов, сдвинутых в нужное положение относительно начала координат. Для этого необходимо один раз провести расчеты голограмм элементарных объектов  $E_i(x, y)$  на поле в два раза превышающем поле топологии и сохранить результаты этих расчетов, а затем использовать их для создания голограмм сложных топологий. Результат склейки для двух полосок приведен на рис. 30. Здесь отдельно были получены голограммы по формулам (31), (33) для горизонтальной и вертикальной полоски, затем эти голограммы были оптимизированы по методу локальных вариаций, сложены и с результирующей голограммы было проведено восстановление.

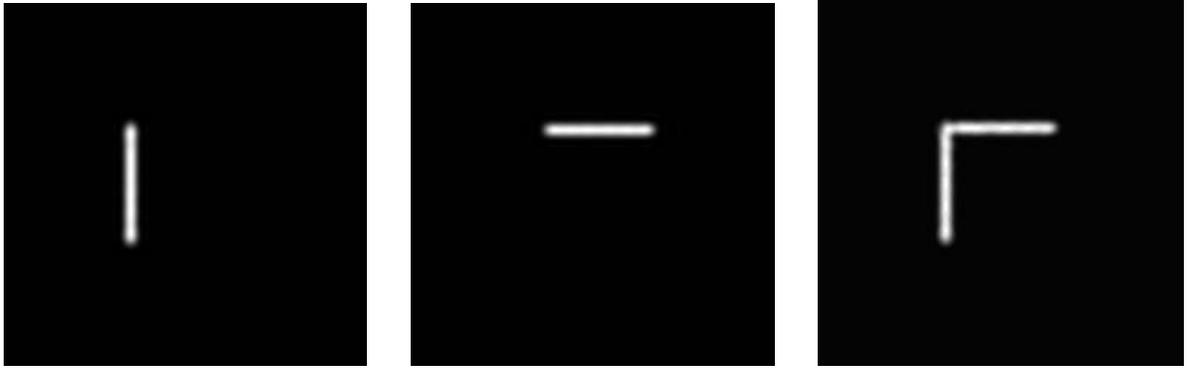


Рис. 30: Составление голограммы уголка из голограмм вертикальной и горизонтальной полосок. Слева результаты восстановления с голограмм вертикальной и горизонтальной полосок, справа - восстановление с результирующей голограммы.

Этот подход требует повышенного качества при проведении оптимизации а также больших объемов памяти для расчета и хранения таких голограмм, представляемых числами повышенной точности. Можно хранить еще голограммы наиболее часто встречающихся конфигураций базовых элементов (например, логические блоки "и", "или" и т.п.). Наиболее эффективен такой способ создания голограмм будет для топологий, состоящих из набора абсолютно одинаковых элементов (например, ячеек памяти - рис. 31). Для таких топологий повторяющийся блок складывается из базовых элементов, а потом быстро "размножается" на всю площадь.

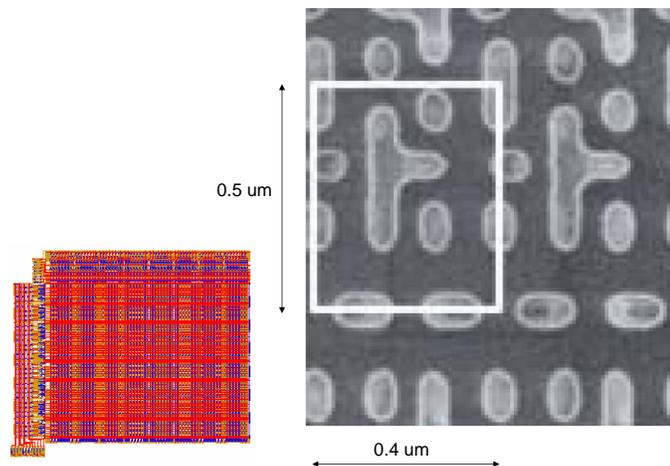


Рис. 31: Участок с периодической структурой (справа) микросхемы SRAM памяти (слева внизу).

Другой способ создания большой голограммы - "метод Большого Пиксела"; предлагаемый в настоящей работе предполагает для расчета по формуле

(39) использовать и аналитическое приближение (49) на большом пикселе и быстрое дискретное преобразование Фурье (44),(45). В методе Большого Пиксела плоскость топологии разбивается на  $M \times M$  "больших пикселей-квадратиков со стороной равной минимальному размеру на топологии  $\sigma_0$ , обычно в несколько раз большей  $\lambda$ . Излучение  $S_{ij}(x, y)$  от большого пиксела с координатами  $(i, j)$  рассчитывается аналитически в приближении дальней зоны (49). Тогда свертку (39) можно, учитывая что  $S_{ij}(k, l) = S(k - i, l - j)$ , рассчитать по формуле

$$W(k, l) = \sum_{ij} S(k - i, l - j)q_0(i, j), \quad \text{при этом } \sigma = \sigma_0, \gamma = \sigma_0 \quad (52)$$

Для вычисления этой суммы типа свертки можно применить БПФ:

$$\mathbf{W} = F^{-1}[F(\mathbf{S}) \otimes F(\mathbf{Q})] \quad (53)$$

где  $\mathbf{Q}$  - матрица, соответствующая функции  $q_0(\xi, \eta)$ , а  $F$  - оператор преобразования Фурье. Асимптотическая формула для количества операций при применении метода Большого Пиксела (53):

$$\Upsilon_{\text{БП}} = O(ab \ln(ab)) \quad (54)$$

такая же как и для метода обычного БПФ (46), однако из-за укрупнения в десять раз сетки, количество операций уменьшается в 100 раз, что позволяет рассчитать серую голограмму по формулам метода Габора (31), (33) для больших топологий. Так расчет голограммы для топологии, заданной на сетке  $40000 \times 40000$ , на 640 ядрах суперкомпьютера МВС100К МСЦ РАН занял 100 минут. Поскольку при восстановлении (35) необходимо сделать шаг хотя бы в 5 раз меньше, чем шаг на заданной топологии, полностью произвести восстановление методом Большого Пиксела с большой голограммы за разумное время невозможно, однако для контроля правильности работы алгоритма, методом аналитического приближения восстанавливается небольшой участок изображения (рис. 32).

Как уже упоминалось ранее, изображения, восстановленные с голограмм Габора имеют неудовлетворительное качество и для их улучшения необходимо проводить оптимизацию. Поскольку метод локальных вариаций даже при использовании всех возможностей восстановления на супер-ЭВМ МВС100-К приводит к слишком большому объему вычислений, для больших топологий

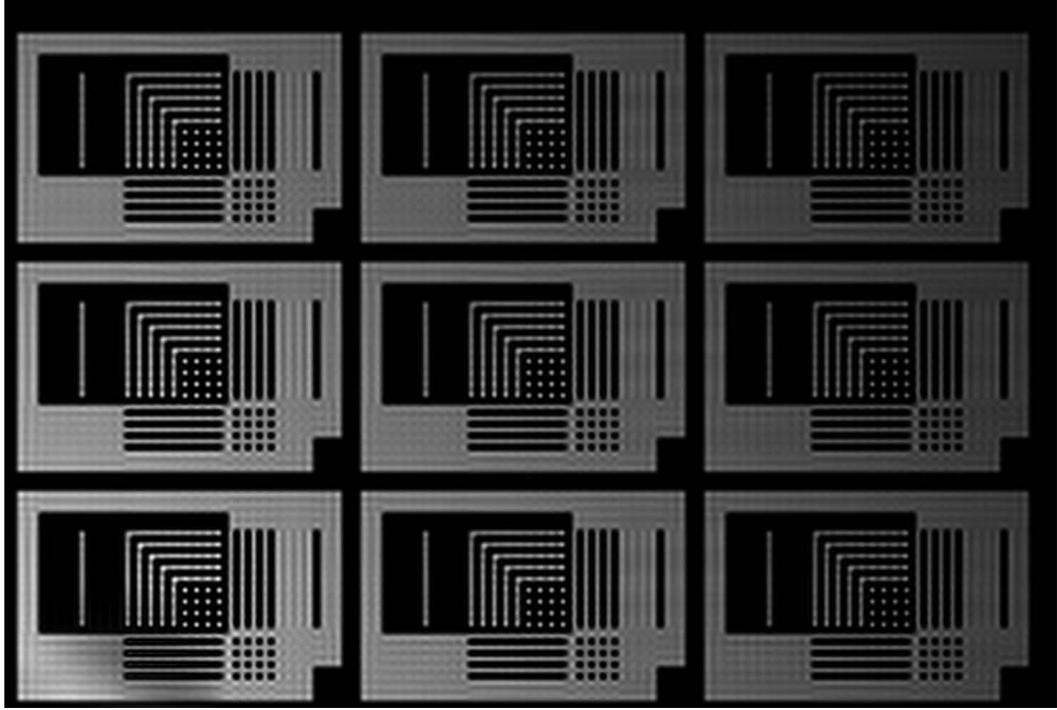


Рис. 32: Восстановление с голограммы, созданной по методу Большого Пиксела.

возможно использовать лишь градиентный метод (13). Вариация функционала  $J_\rho(T)$  из (27) для нормы  $\rho = L_2$  может быть получена аналитически:

$$\delta J_{L_2}(T) = 4\text{Re}(V(x, y), \overline{\delta T(x, y)})_{(x, y)} \quad (55)$$

где

$$V(x, y) = \iint_{\mathbb{H}} K(\xi - x, \eta - y) [ |K * T|^2 \cdot K * T - K * T \cdot q_0^2 ] (\xi, \eta) d\xi d\eta \quad (56)$$

Для вычислений по формуле (55) необходимо посчитать четыре свертки типа (39). При этом шаг по изображению должен быть как при восстановлении -  $\sigma_0/5$ , а по голограмме -  $\gamma_0$ . С учетом того, что необходимо совершить несколько десятков шагов итерационного процесса, метод градиентного спуска с использованием метода Большого Пиксела возможно реализовать лишь для топологий средних размеров. Для больших топологий существующих на данный момент вычислительных мощностей будет недостаточно, однако разработан, но пока не реализован, алгоритм, развивающий идею метода Большого Пиксела и позволяющий провести оптимизацию градиентным

методом для голограмм больших топологий содержащих более миллиарда элементов на суперкомпьютерах уровня МВС100К.

## Список литературы

1. Borodovsky Y., Cheng W.-H., Schenker R., Singh V. Pixelated phase mask as novel lithography RET // Proc of SPIE Vol. 6924
2. Singh V., Hu B., Bollepalli S., Wagner S., Borodovsky Y. Making a trillion pixels dance // Proc of SPIE Vol. 6924
3. Davids S.P., Bollepalli S.B. Generalized inverse problem for partially coherent projection lithography // Proc of SPIE Vol. 6924
4. Born M., Wolf E. Principles of Optics. Pergamon Press, Oxford, 1980
5. Levenson M.D., Johnson K.M., Hanchett V.C., Chiang K. Projection photolithography by wave-front conjugation // J. Opt. Soc. Am., V. 71, No. 6, June 1981
6. Beesley M.J., Foster H., Hambleton K.G. Holographic projection of microcircuit patterns // Electronic letters, Vol. 4, No. 3, 9th February 1968
7. Craighead H.G., White J.C., Howard R.E., Jackel L.D., Behringer R.E., Sweeney J.E., Epworth R.W. Contact lithography at 157 nm with an F2 excimer laser J.Vac. Sci. Technol., B 1(4), 1983
8. Philip S. Considine Effects of coherence on imaging systems // J. Opt. Soc. Am., V. 56, No. 8, August 1966
9. Efremow N.N., Economou N.P, Bezjian K., Dana S.S., Henry I. Smith A simple technique for modifying the profile of resist exposed by holographic lithography // J.Vac. Sci. Technol., Vol. 19, No. 4, 1981
10. Voshchenkov A.M., Hanson R.C. Submicron resolution photolithography by spectral shaping // IEEE Electron Device Letters, Vol. EDL-3, No. 7, July 1982
11. A. Poonawala and P. Milanfar Proc. of SPIE 6154 (2006).
12. A.Poonawala and P. Milanfar IEEE Trans. on Image Proc. 16 (2007).
13. Y. Liu and A. Zakhor Proc. of SPIE 1463 (1991).
14. Y. Granik Proc. of SPIE 5754 (2005).
15. Gabor D. A new microscopic principle // Nature, V. 161, PP. 777-778, 1948

16. Gabor D. Microscopy by reconstructed wavefronts // Proc. Roy. Soc, 1949 (A), 197, pp. 454-486
17. Chris A. Mark Positive resist // J.Electrochem.Soc.: Solid-State Science and Technology, Vol. 134, No 1, pp. 148-152, January 1987
18. N. Cobb and A. Zakhor Fast sparse aerial image calculation for OPC // BACUS Symposium on Photomask Technology, Proc. SPIE 2621, pp. 534-545, 1995.
19. Huang W.C. Two threshold resist models for Lithography simulation
20. Randall J. Variable Threshold Resist Models for Lithography Simulation
21. Черноусько Ф.Л., Баничук Н.В. Вариационные задачи механики и управления. М. "Наука 240 с., 1973.
22. Н. Н. Моисеев, Ю. П. Иванилов, Е. М. Столярова. Методы оптимизации. - М.: <Наука>, 1978. - 351 с.
23. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы