

Эффективные методы расчета электромагнитных полей

Д.Ю. Князьков

Институт проблем механики им. А.Ю. Ишлинского РАН

Рассматривается задача расчета электромагнитных полей от большого количества (10^8 и более) элементарных объектов для нужд голографической литографии. Анализ производительности некоторых методов расчета электромагнитных полей на кластерах МВС100-К МСЦ РАН и МИИТ Т4700 показал невозможность создания этими методами голограммы Габора для микрочипа с реальными размерами на современных суперкомпьютерах. Специально разработанный метод Большого пиксела, позволяющий рассчитывать с его помощью голограммы Габора для современных микрочипов на современных параллельных вычислительных комплексах, был реализован в составе параллельного ПК. Приведен пример синтеза голограммы Габора для топологии из $1.6 \cdot 10^9$ расчетных элементов.

1. Введение

В рамках научно-исследовательских работ по голографической литографии [1–3] необходимо уметь решать задачи создания, оптимизации и восстановления для топологий состоящих из одного или нескольких элементарных объектов (квадратиков, полосок, шевронов, тестовых мир). Такие задачи могут быть решены с использованием кластерного суперкомпьютера небольшой мощности [4].

Расчет голограммы Габора [5] является здесь необходимым технологическим этапом. В обычной проекционной литографии маска фактически является гомотетически растянутой исходной топологией, поэтому ее расчет достаточно прост. Расчет же голографической маски обычными методами для топологий реальных чипов, содержащих миллиарды элементов, будет требовать вычислительных мощностей, на порядки превышающих производительность существующих на сегодня супер-ЭВМ.

Восстановление в проекционной литографии также облегчается тем, что зависимость изображения от маски - локальна [6], тогда как в голографической литографии "информация" о каждом участке изображения распределена по всей маске и, даже при восстановлении небольшого фрагмента изображения, необходимо учитывать дифракцию от всех элементов маски. Таким образом, расчет электромагнитных полей является существенной проблемой в голографической литографии и будет рассмотрен в настоящей работе.

2. Постановка задачи

Схема голографической литографической установки [1–3] изображена на рис. 1. Излучение W падает на голографическую пластинку, которая занимает область Ω_H в плоскости O_Hxy , а ее пропускание описывается вещественной функцией $T(x, y): \Omega_H \rightarrow [0, 1]$. При этом области $\{(x, y) \mid (x, y) \in \Omega_H, T(x, y) = 1\}$ полностью прозрачны для излучения, а области $\{(x, y) \mid (x, y) \in \Omega_H, T(x, y) = 0\}$ излучение W не пропускают. В результате, в плоскости $O_I\xi\eta$ необходимо создать требуемое световое изображение.

Пусть $g_0(\xi, \eta): \Omega_I \rightarrow \{0, 1\}$ - некоторая функция, задающая желаемую конфигурацию дорожек (топологию) на микросхеме. Возьмем в области Ω_I на плоскости $O_I\xi\eta$ соответствующее объектное поле - комплекснозначную функцию $g(\xi, \eta): \Omega_I \rightarrow \mathbb{C}$. Тогда голограммой Габора [5] для объекта $g(\xi, \eta)$ называется функция $T_g(x, y): \Omega_H \rightarrow [0, 1]$, рассчитанная по следующей формуле:

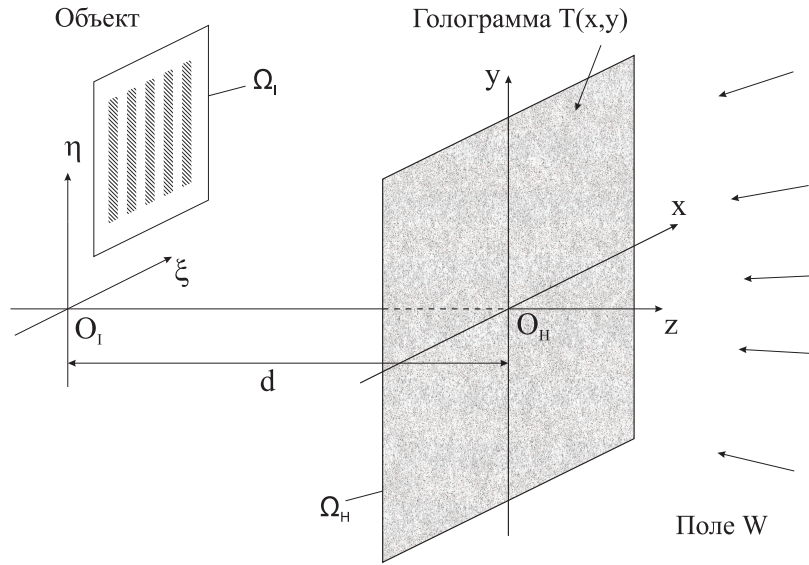


Рис. 1. Оптическая схема.

$$T_g(x, y) = \left| \iint_{\Omega_I} K(x, y, \xi, \eta) g(\xi, \eta) d\xi d\eta + \overline{W}(x, y) \right|^2, \quad (1)$$

где $K(x, y, \xi, \eta) = \frac{e^{i\tilde{k}R}}{R}$, $R = \sqrt{(x - \xi)^2 + (y - \eta)^2 + d^2}$, $\tilde{k} = \frac{2\pi}{\lambda}$ - волновое число, λ - длина волны, а $\overline{W}(x, y)$ - поле, комплексно-сопряженное к восстанавливающему полю $W(x, y)$ (рис. 1).

На расстоянии d от голограммы, в объектной плоскости $O_I\xi\eta$ находится область наблюдения Ω_I , в которой регистрируется интенсивность излучения $I(\xi, \eta)$. Если поле в плоскости O_Hxy до попадания на голограмму с пропусканием $T(x, y)$ описывается функцией $W(x, y)$, то интенсивность излучения в объектной плоскости может быть в скалярном приближении вычислена с использованием интеграла Кирхгофа [1]:

$$I_T(\xi, \eta) = \left| \iint_{\Omega_H} K(x, y, \xi, \eta) W(x, y) T(x, y) dx dy \right|^2. \quad (2)$$

Полученное с голограммы Габора $T_g(x, y)$ изображение $I_{T_g}(\xi, \eta)$ будет достаточно близко к заданной топологии $g_0(\xi, \eta)$.

Заметим, что при больших апертурах необходимо использовать векторную модель излучения. Формула (1) для голограммы при этом не изменится, а расчет восстановленного изображения может быть сведен к расчету нескольких интегралов типа (2), например, с использованием формул Стреттона-Чу [7]. Таким образом, основная проблема как в задаче создания голограммы (1), так и в задаче моделирования эксперимента по ее засветке (2) - это расчет интеграла

$$I(x, y) = \iint_{\Omega} K(x, y, \xi, \eta) \varphi(x, y) d\xi d\eta \quad (3)$$

с сильно осцилирующим ядром

$$K(x, y, \xi, \eta) = \frac{e^{i\frac{2\pi}{\lambda}\sqrt{(x-\xi)^2+(y-\eta)^2+d^2}}}{\sqrt{(x-\xi)^2+(y-\eta)^2+d^2}}. \quad (4)$$

3. Применение прямых методов расчета и методов с БПФ

Без ограничения общности, возьмем в качестве области Ω квадрат со стороной a . Рассмотрим самую трудоемкую операцию в задачах синтеза голограммы (1) и восстановления изображения (2) - расчет интеграла

$$f(x, y) = \int_{-\frac{a}{2}+\eta_0}^{\frac{a}{2}+\eta_0} \int_{-\frac{a}{2}+\xi_0}^{\frac{a}{2}+\xi_0} K(x, y, \xi, \eta)\varphi(\xi, \eta)d\xi d\eta \quad (5)$$

где (ξ_0, η_0) - центр топологии, заданной в квадрате размера $a \times a$. Конкретный вид функции $\varphi(\xi, \eta)$ зависит от решаемой задачи. Например, при создании голограммы $\varphi(\xi, \eta) = R(\xi, \eta)q_0(\xi, \eta)$, где $R(\xi, \eta)$ и $q_0(\xi, \eta)$ - функции, задающие волну подсветки и топологию соответственно. Пусть голограмма размера $b \times b$ состоит из $M \times M$ ячеек размера $\gamma \times \gamma$ (рис. 5): $M = \frac{b}{\gamma}$. Тогда для расчета серой голограммы необходимо найти значения функции f в $M \times M$ точках: $\{(-\frac{b}{2} + \gamma(i + \frac{1}{2}), -\frac{b}{2} + \gamma(j + \frac{1}{2})) | i, j = 0, \dots, M - 1\}$. При этом область интегрирования будет разбиваться различным образом (см. рис. 2) в зависимости от метода, используемого при расчете интеграла (5).

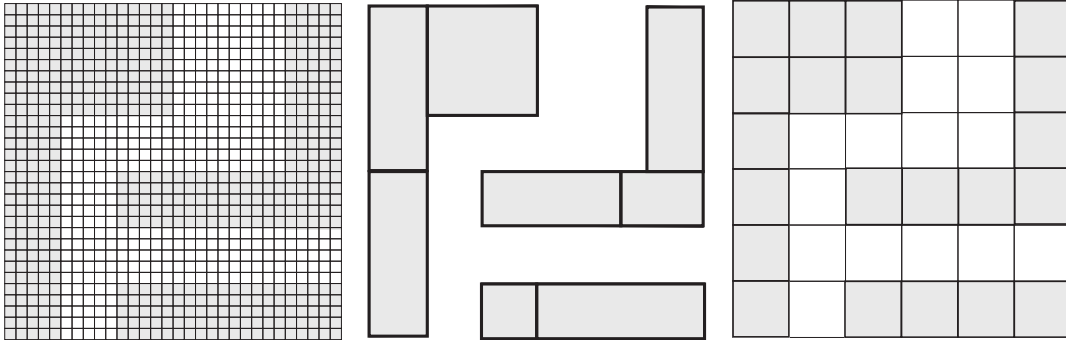


Рис. 2. Разбиение области интегрирования (расчетные сетки) для методов простого суммирования и метода с применением БПФ (слева), метода разбиения на прямоугольники (в центре) и метода Большого Пиксела (справа).

3.1. Метод простого суммирования

Выпишем интегральные суммы для (5). Поскольку функция ядра колеблется с периодом λ , для расчета интеграла нужно брать шаг интегрирования δ равным хотя бы $\frac{\lambda}{10}$ (см. рис. 2 слева). Тогда количество точек разбиения $N_D = \lceil \frac{10a}{\lambda} \rceil$ и интегральные суммы:

$$f(i, j) = \sum_{k, l=0}^{N_D-1} \varphi(k, l)K(i, j, k, l), \quad i, j = 0, \dots, M - 1 \quad (6)$$

где под $f(i, j)$, $\varphi(k, l)$, $K(i, j, k, l)$, $i, j = 0, \dots, M - 1$, $k, l = 0, \dots, N_D - 1$ подразумеваются следующие обозначения:

$$\begin{aligned}
f(i, j) &= f(-\frac{b}{2} + \gamma(i + \frac{1}{2}), -\frac{b}{2} + \gamma(j + \frac{1}{2})) \\
\varphi(k, l) &= \varphi(\xi_0 - \frac{a}{2} + \delta(k + \frac{1}{2}), \eta_0 - \frac{a}{2} + \delta(l + \frac{1}{2})) \\
K(i, j, k, l) &= \\
&= K(-\frac{b}{2} + \gamma(i + \frac{1}{2}), -\frac{b}{2} + \gamma(j + \frac{1}{2}), \xi_0 - \frac{a}{2} + \delta(k + \frac{1}{2}), \eta_0 - \frac{a}{2} + \delta(l + \frac{1}{2}))
\end{aligned}$$

где (ξ_0, η_0) - центр изображения (топологии).

Для расчета по формуле (6) понадобится $\Upsilon_D^* = (C_\varphi^* + C_K^* + 1)M^2N_D^2$ умножений и $\Upsilon_D^+ = (C_\varphi^+ + C_K^+ + 1)M^2N_D^2 - M^2$ сложений, где C_φ^* и C_φ^+ - количество умножений и сложений для вычисления значения функции φ , а C_K^* и C_K^+ - количество умножений и сложений для вычисления значения функции K . Упрощенно можно записать

$$\Upsilon_D = C_D M^2 N_D^2. \quad (7)$$

3.2. Разбиение топологии на прямоугольники

Представим теперь топологию g_0 в виде суммы N_R прямоугольников P_m , $m = 1, \dots, N_R$:

$$\{(\xi, \eta) \mid g_0(\xi, \eta) = 1\} = \bigsqcup_{m=1}^{N_R} P_m.$$

Пример такого разбиения показан на рис. 2 в центре.

При вычислении интеграла (5) можно воспользоваться тем, что подсветка объекта $R(\xi, \eta)$ мало меняется на этих прямоугольниках и можно считать $R(\xi, \eta)|_{P_m} = R_m$, $\forall m$. Тогда, поскольку $\varphi(\xi, \eta) = R(\xi, \eta)g_0(\xi, \eta)$, получаем $\varphi(\xi, \eta)|_{P_m} = \varphi_m$, $\forall m$ и

$$f(x, y) = \int_{-\frac{a}{2} + \eta_0}^{\frac{a}{2} + \eta_0} \int_{-\frac{a}{2} + \xi_0}^{\frac{a}{2} + \xi_0} K(x, y, \xi, \eta) \varphi(\xi, \eta) d\xi d\eta = \sum_{m=1}^{N_R} \varphi_m \iint_{P_m} K(x, y, \xi, \eta) d\xi d\eta.$$

Соотношение геометрических параметров оптической схемы голографической установки таково, что интегралы, участвующие в последней сумме могут быть вычислены аналитически в приближении дальней зоны [8, 9]: $\iint_{P_m} K(x, y, \xi, \eta) d\xi d\eta = I_{P_m}(x, y)$. Тогда

$$f(x, y) = \sum_{m=1}^{N_R} \varphi_m I_{P_m}(x, y),$$

откуда получаем расчетную формулу для (5):

$$f(i, j) = \sum_{m=1}^{N_R} \varphi_m I_{P_m}(i, j), \quad i, j = 0, \dots, M - 1, \quad (8)$$

где под $f(i, j)$, $I_{P_m}(i, j)$, $i, j = 0, \dots, M - 1$ подразумеваются следующие обозначения:

$$\begin{aligned}
f(i, j) &= f(-\frac{b}{2} + \gamma(i + \frac{1}{2}), -\frac{b}{2} + \gamma(j + \frac{1}{2})), \\
I_{P_m}(i, j) &= I_{P_m}(-\frac{b}{2} + \gamma(i + \frac{1}{2}), -\frac{b}{2} + \gamma(j + \frac{1}{2})).
\end{aligned}$$

Для осуществления вычислений по формуле (8) понадобится совершить $\Upsilon_{\Pi}^* = (C_{\varphi}^* + C_I^* + 1)M^2N_R$ умножений и $\Upsilon_{\Pi}^+ = (C_{\varphi}^+ + C_I^+ + 1)M^2N_R - M^2$ сложений, где C_I^* и C_I^+ - количество операций для вычисления значения функции $I_{P_m}(x, y)$. Если в качестве топологии взять структуру, представляющую собой квадрат размером $a \times a$, заполненный маленькими квадратами размера $\sigma \times \sigma$, получим $N_R = \frac{a^2}{4\sigma^2}$. Здесь и далее σ - это минимальный характеристический размер элемента на топологии, $N = \frac{a}{\sigma}$. Тогда получим упрощенную формулу для количества операций:

$$\Upsilon_R = \frac{1}{4}C_R N^4. \quad (9)$$

Такой метод расчета был реализован на языке программирования C++ в составе параллельного программного комплекса BinNet [4]. Этот ПК успешно использовался на кластерах МВС100-К МСЦ РАН и МИИТ Т4700 при проведении исследовательских работ. С его помощью рассчитывались файлы голограммных масок для их дальнейшего изготовления, моделировались процессы создания голограмм, восстановления изображений и проводились различные численные эксперименты для нужд исследовательского проекта по голографической литографии (некоторые результаты этих работ приведены в [1]).

Из формулы (8) видно, что вычисление результатов дифракции от разных прямоугольников в разных точках голограммы совершенно независимы, поэтому можно ожидать, что соответствующий алгоритм будет обладать хорошей масштабируемостью на кластерных ВС. Действительно, результаты численных экспериментов (рис. 3) показывают, что эффективность алгоритма близка к 1 (замерялось суммарное время необходимое для расчета по (8) и межпроцессорного взаимодействия). Размер задачи n - это количество сумм в (8), остальные параметры близки к параметрам экспериментальной установки.

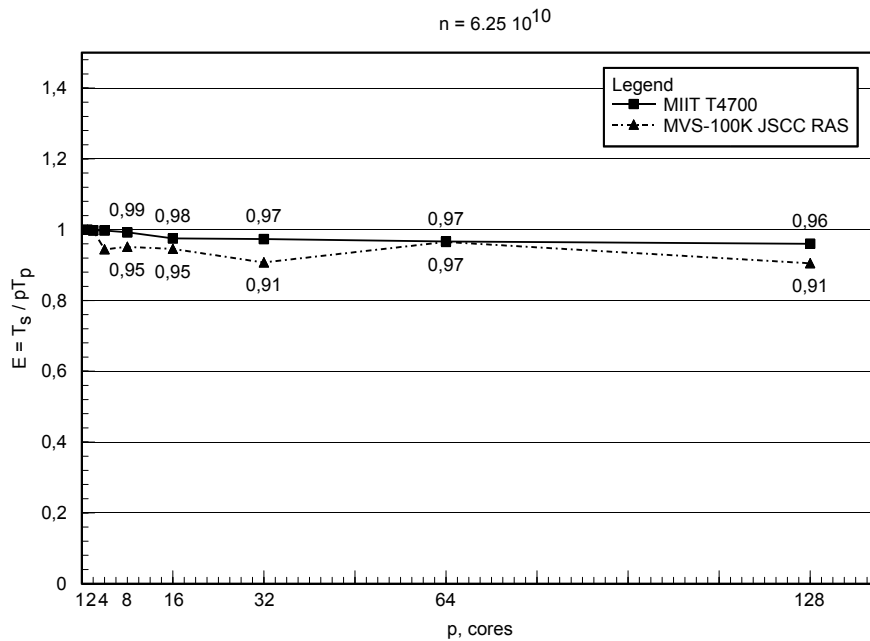


Рис. 3. Эффективность работы алгоритма суммирования излучения от прямоугольников. Количество вычислительных ядер изменяется от 1 до 128, размер задачи постоянен.

3.3. Использование быстрого преобразования Фурье

В описанных ранее алгоритмах расчета по формулам (6) и (8) никак не используется то, что исходный интеграл является, как легко увидеть из (4), сверткой с ядром $K_1(x_1, y_1) =$

$K(x_1, y_1, 0, 0)$. Перепишем теперь (6) с учетом этого свойства:

$$f(i, j) = \sum_{k, l=0}^{N_{FFT}-1} \varphi(k, l) K_1(i - k, j - l), \quad i, j = 0, \dots, M_{FFT} - 1, \quad (10)$$

где $N_{FFT} = \frac{a}{\delta}$, $M_{FFT} = \frac{b}{\delta}$, δ - шаг интегрирования, а под $f(i, j)$, $\varphi(k, l)$, $K_1(s, t)$, $i, j = 0, \dots, M_{FFT} - 1$, $k, l = 0, \dots, N_{FFT} - 1$ подразумеваются следующие обозначения:

$$\begin{aligned} f(i, j) &= f\left(-\frac{b}{2} + \delta\left(i + \frac{1}{2}\right), -\frac{b}{2} + \delta\left(j + \frac{1}{2}\right)\right) \\ \varphi(k, l) &= \varphi\left(-\frac{a}{2} + \delta\left(k + \frac{1}{2}\right), -\frac{a}{2} + \delta\left(l + \frac{1}{2}\right)\right) \\ K_1(s, t) &= K_1\left(-\frac{b}{2} + \delta\left(s + \frac{1}{2}\right), -\frac{b}{2} + \delta\left(t + \frac{1}{2}\right)\right) \end{aligned}$$

Заметим, что в (10) шаг расчета интеграла на голограмме совпадает с шагом интегрирования по плоскости топологии (рис. 2 слева): голограмма рассчитывается не как ранее в узлах сетки со стороной, равной ячейке голограммы γ , а в узлах сетки со стороной, равной δ . Таким образом теперь рассчитывается очень много "лишних" точек на голограмме, однако появляется возможность рассчитывать свертку (10) с помощью преобразования Фурье:

$$\mathbf{f} = F^{-1}[F(\boldsymbol{\varphi}) \otimes F(\mathbf{K}_1)], \quad (11)$$

где F и F^{-1} - прямое и обратное двумерное дискретное преобразование Фурье, \otimes - поэлементное умножение, а \mathbf{f} , $\boldsymbol{\varphi}$, \mathbf{K}_1 - соответствующие матрицы. Использование быстрого преобразования Фурье [10] позволяет получить следующее количество операций: $\mathcal{Y}_{FFT} = 6C_1((k+1)N_{FFT} - 1)^2 \log_2((k+1)N_{FFT} - 1)$, где $k = \frac{b}{a}$. Если отбросить несущественные члены, принять $\delta = \frac{\lambda}{10}$, $N_{FFT} = 10N$, а в качестве константы C_1 выбрать $C_1 = 5$ [11], то получим

$$\mathcal{Y}_{FFT} = 30(k+1)^2 N^2 10^2 \log_2((k+1)^2 N), \quad (12)$$

то есть $N^2 \log_2 N$ в качестве асимптотики количества операций, тогда как простое суммирование (как по формуле (6), так и по (10)) приводит к асимптотике N^4 (здесь принято $N = \frac{a}{\sigma}$, где σ - размер минимального элемента на микросхеме и учтено, что $\sigma \approx \lambda$, $N_D \sim N$, $M \sim N$, $N_{FFT} \sim N$, $M_{FFT} \sim N$).

К сожалению, реальная производительность алгоритма БПФ на кластерной ВС при распределенном хранении соответствующей матрицы будет невысока. На рис. 4 приведен график зависимости эффективности работы такого алгоритма от количества используемых вычислительных ядер (от 1 до 512) при постоянном размере задачи (матрица 15999×15999 комплексных чисел двойной точности). Видно, что эффективность падает в 4 раза при увеличении количества процессов от 1 до 8 на одном узле (вычислительный узел МВС100-К имеет два 4-х ядерных процессора), далее держится примерно на уровне 25% при увеличении количества процессов до 64 (8 узлов) и затем продолжает падать, достигая 10% на 512 вычислительных ядрах.

3.4. Трудоемкость алгоритмов и асимптотика количества операций

Проанализируем возможности использования описанных выше методов для расчета топологий, состоящих из большого количества элементов. Количество операций при использовании метода простого суммирования ($\mathcal{Y}_D^* = (C_\varphi^* + C_K^* + 1)M^2 N_D^2$, $\mathcal{Y}_D^+ = (C_\varphi^+ + C_K^+ + 1)M^2 N_D^2 - M^2$) пропорционально $M^2 N_D^2$. Это слишком много даже если для вычислений используется современный суперкомпьютер. Действительно, пусть рассчитывается топология с минимальным размером элемента $\sigma = 100$ нм на чипе со стороной $a = 1$ см. Тогда

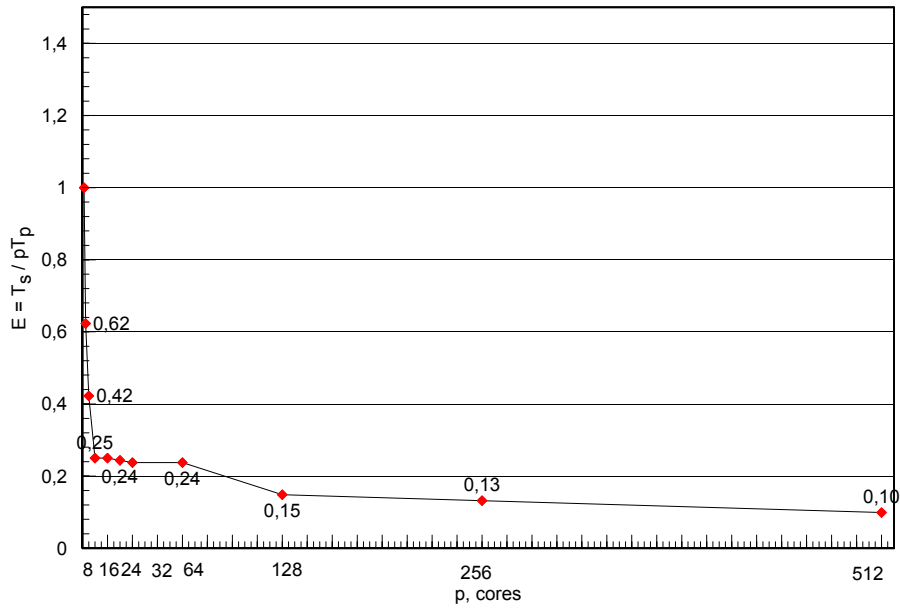


Рис. 4. Эффективность двумерного параллельного БПФ на МВС100-К МСЦ РАН.

$N = 10^5$ и для реальной оптической схемы $M = 10^5$. При длине волны $\lambda = 193$ нм, получаем $N_D \approx 5 \cdot 10^5$, а количество операций при $C_D = 10$: $\mathcal{Y}_D = 2,5 \cdot 10^{22}$. Пусть такой алгоритм реализуется на суперкомпьютере с производительностью 100 TFlops. Тогда, даже если бы при вычислениях достигалась пиковая производительность (а в случае расчета по формуле (6) этого, на самом деле, не сложно добиться), на расчет одной голограммы ушло бы около 8 лет.

Количество операций для формулы излучения от прямоугольников (8) $\mathcal{Y}_R \sim M^2 N_R$. Поскольку количество прямоугольников, из которых состоит топология N_R в общем случае пропорционально N^2 , и для оптической схемы, которая используется в эксперименте $M \sim N$, получаем такую же неприемлимую асимптотику, как и в предыдущем случае: $\mathcal{Y}_R \sim N^4$. Для топологии с $\sigma = 100$ нм и $a = 1$ см из оценки количества операций (9) для синтеза голограммы Габора получается, что расчет займет более 100 суток на суперкомпьютере с производительностью 100 TFlops.

Аналогичные оценки для расчета по методу БПФ (11) дают время около 30 мин, но, учитывая низкую эффективность этого алгоритма на кластерной ВС (см. рис. 4), реальное время работы составит около 10 часов. Этот алгоритм имеет также повышенные требования к общему объему оперативной памяти.

Для расчета голограмм Габора можно предложить ряд методов, использующих предварительно рассчитанные поля от некоторых повторяющихся структур. Такими структурами могут быть как просто прямоугольники (со сторонами, параллельными осям координат или имеющими с ними угол 45 градусов - именно из таких элементов состоят топологии слоев реальных микросхем), так и повторяющиеся структурные элементы (например одинаковые элементарные ячейки в схемах памяти). Количество таких структур N_Ω , на которые разбивается топология пропорционально в общем случае N^2 . Получается, что для всех способов с предварительно рассчитанными полями (голограммами) верна та же асимптотика количества операций - $M^2 N^2$, а с учетом того, что на практике $M \sim N$, имеем асимптотику N^4 . Более того при использовании этих методов для расчета голограмм топологий, близких по сложности к топологиям реальных слоев микрочипов, возникает проблема хранения и загрузки в память огромных массивов данных.

Метод расчета (11) с помощью БПФ, если потребовать размещения всех необходимых

матриц в ОЗУ, успешно может быть применен на персональном компьютере только для расчета голограмм и восстановления с них для маленьких топологий (полосок, квадратиков, уголков, шевронов). Для работы с подобными и более сложными - средними топологиями (например тестовая мира рис. 6) использовался упомянутый выше ПК BinNet. Такие расчеты обычно занимали от 5-10 минут на 32 вычислительных ядрах до суток на всем кластере МИИТ Т4700, состоящем из 64 узлов с двумя 4-х ядерными процессорами и 16 Гб ОЗУ каждый. Однако, из оценок количества необходимых арифметических операций получается, что ни один из описанных выше методов не может быть применен для расчета голограмм реальных чипов на существующих на данный момент суперкомпьютерах.

4. Метод большого пиксела

Рассмотрим два из описанных выше методов создания серой голограммы: метод расчета свертки (10) с помощью БПФ (11) и метод суммирования излучения от прямоугольников, на которые разбивается топология (8). Можно заметить, что в первом случае получается большой выигрыш в асимптотике ($C_{FFT}N^2 \log_2 N$ по сравнению с $C_D N^4$ для метода простого суммирования (6)), но при этом необходимо брать очень маленький шаг по сеткам на топологии и голограмме. Во втором случае, на голограмме не считаются "лишние" точки и излучение от целых прямоугольников рассчитывается по аналитической формуле, но асимптотика остается неприемлимой - $C_R N^4$. Метод большого пиксела (БП), предложенный А.С. Шамаевым, использует преимущества этих двух подходов: с одной стороны излучение от минимального характерного квадрата топологии - большого пиксела размера $\sigma \times \sigma$ - считается аналитически, а с другой - расчет сводится к вычислению свертки, что позволяет воспользоваться БПФ и добиться асимптотики $C_{BP} N^2 \log_2 N$.

4.1. Разбиение топологии на большие пикселы

Разобьем всю область объекта (квадрат $a \times a$ с центром в точке (ξ_0, η_0)) на большие пикселы - квадратики \square_{kl} размера $\sigma \times \sigma$ с центрами в $(\xi_0 - \frac{a}{2} + \sigma(k + \frac{1}{2}), \eta_0 - \frac{a}{2} + \sigma(l + \frac{1}{2}))$, $k, l = 0, \dots, N - 1$, где $N = \lfloor \frac{a}{\sigma} \rfloor$. Таким образом, на каждом \square_{kl} функция топологии q имеет постоянное значение q_{kl} : 1, если квадратик \square_{kl} принадлежит топологии и 0 в противном случае (см. рис. 2 справа). При задании топологии некоторым ее элементам может быть приписана отличная от нуля фаза ψ (например, когда при реализации техники phaseshift соседним элементам приписываются фазы 0 и π). Эта фаза также будет постоянной на \square_{kl} . Подсветка объекта R практически не меняется на площади большого пиксела поэтому можно считать, что функция $\varphi = q\psi R$ на каждом из \square_{kl} постоянна и равна φ_{kl} . Тогда исходный интеграл (5) можно преобразовать следующим образом:

$$f(x, y) = \iint_{\square_{kl}} \varphi(\xi, \eta) K(x, y, \xi, \eta) d\xi d\eta = \sum_{k, l=0}^{N-1} \varphi(k, l) \iint_{\square_{kl}} K(x - \xi, y - \eta) d\xi d\eta. \quad (13)$$

Обозначим $S_{kl}(x, y) = \iint_{\square_{kl}} K(x, y, \xi, \eta) d\xi d\eta$ - результат дифракции на квадратном отверстии - большом пикселе \square_{kl} . В приближении дальней зоны этот интеграл может быть посчитан аналитически. Возьмем теперь шаг расчетной сетки на голограмме равным шагу по топологии σ . Тогда на голограмме будет $H \times H$ расчетных узлов, где $H = \lfloor \frac{b}{\sigma} \rfloor$. Заметим, что интеграл $S_{kl}(x, y)$ инвариантен относительно одновременного сдвига по объекту и по голограмме: $\forall \delta_k, \delta_l \in \mathbb{Z}$ имеем $S_{k+\delta_k, l+\delta_l}(x + \delta_k \sigma, y + \delta_l \sigma) = S_{kl}(x, y)$. Используя это свойство и обозначив $S(i, j) = S_{00}(i\sigma, j\sigma)$ сумму (13) можно преобразовать следующим образом:

$$f(i, j) = \sum_{k, l=0}^{N-1} \varphi(k, l) S(i - k, j - l), \quad i, j = 0, \dots, H - 1, \quad (14)$$

где $f(i, j) = f(-\frac{b}{2} + \sigma(i + \frac{1}{2}), -\frac{b}{2} + \sigma(j + \frac{1}{2}))$.

Для прямого расчета свертки (14) понадобится $N^2 H^2 = [\frac{a}{\sigma}]^2 [\frac{b}{\sigma}]^2$ операций. Таким образом, так как $a \sim b$ при увеличении размера топологии a или уменьшении характерного размера σ количество операций будет расти как 4-ая степень, что при работе с большими топологиями приводит к невозможности произвести расчет за приемлемое время даже на суперкомпьютере. Однако, для вычисления свертки (14) можно применить описанный выше способ, использующий 3 двумерных БПФ (11). Тогда количество операций составит $\gamma_{БП} = 3 \cdot 2C_1((k+1)N - 1)^2 \log_2((k+1)N - 1)$. Если отбросить несущественные члены и принять $C_1 = 5$, будем иметь

$$\gamma_{БП} = 30(k+1)^2 N^2 \log_2((k+1)N). \quad (15)$$

Таким образом, алгоритм большого пиксела дает выигрыш в два порядка по времени счета и по требуемому объему оперативной памяти по сравнению с обычным методом расчета свертки с БПФ (11), (12).

4.2. Реализация метода Большого Пиксела на кластерной ВС

При реализации метода большого пиксела на кластерной ВС, поле на голограмме рассчитывалось по зонам, линейный размер каждой из которых совпадал с размером топологии (рис. 5). Всего тогда надо рассчитать k^2 таких зон. Тогда количество операций составит $\gamma_{БПЗ} = (2k^2 + 1)2C_1(2N - 1)^2 \log_2(2N - 1)$. При $C_1 = 5$ получим

$$\gamma_{БП} = (80k^2 + 40)N^2 \log_2(2N). \quad (16)$$

Это несколько больше, чем для метода большого пиксела (15), но в 37.5 раз меньше, чем для обычного метода с БПФ (12). Более того, в требование к общему объему оперативной памяти для алгоритма БП в таком виде снижается еще на два порядка (точнее, в $\frac{(k+1)^2}{4}$ раз).

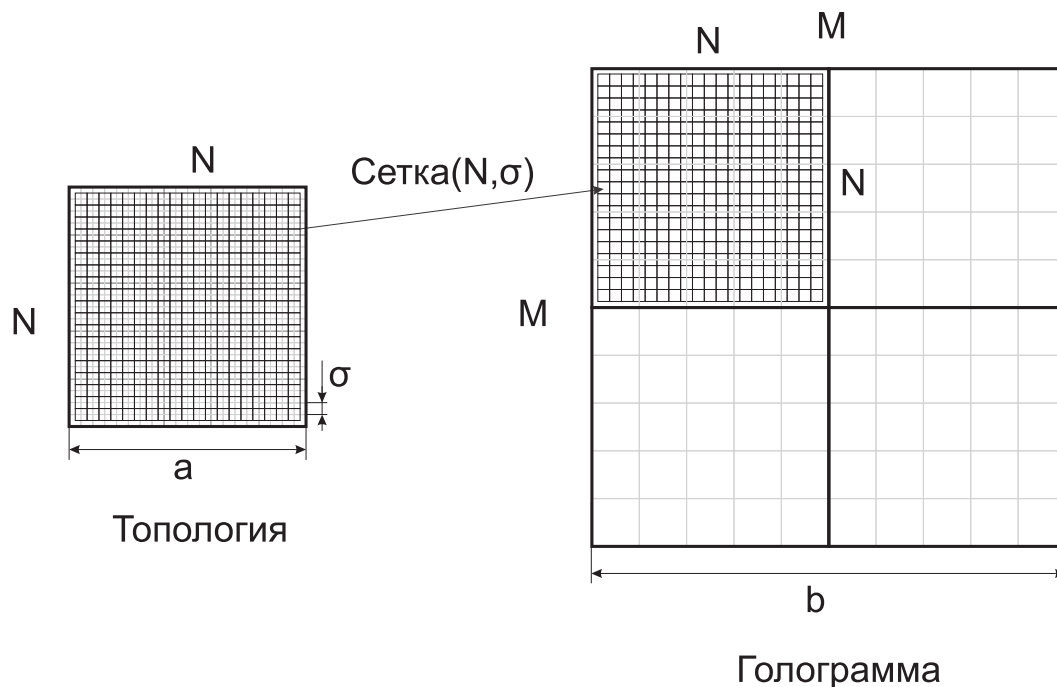


Рис. 5. Расчетные сетки при синтезе голограммы методом большого пиксела с разбиением на зоны.

Метод большого пиксела был реализован на языке C++ в составе ПК BigPixel. Для осуществления 2-мерного распределенного параллельного БПФ в программе использовались пакеты FFTW и Intel MKL.

Поведение эффективности алгоритма практически идентично поведению эффективности алгоритма БПФ (см. рис. 4), так как на каждом из k^2 шагов основное время занимает расчет двух двумерных БПФ, при этом время поточечного умножения матриц пренебрежимо мало, а БПФ матрицы φ осуществляется лишь один раз и результат его запоминается.

4.2.1. Пример численного расчета

В качестве тестовой топологии был выбран объект, состоящий из 40000×40000 расчетных пикселей и заполненный тестовыми мирами размером 57×37 пикселей каждая. Рассчитываемая голограмма Габора была в 10 раз больше объекта, то есть $k = 10$. Этот расчет проводился на 640 вычислительных ядрах суперкомпьютера МВС100-К МСЦ РАН и занял 100 минут. При этом время осуществления каждого БПФ матрицы размера 79999×79999 из комплексных чисел двойной точности составляло 29 секунд. Для контроля правильности расчета голограммы проводилось восстановление (уже по формулам типа (8)) небольшого участка изображения, содержащего 9 мир. Результат этого восстановления показан на рис. 6.

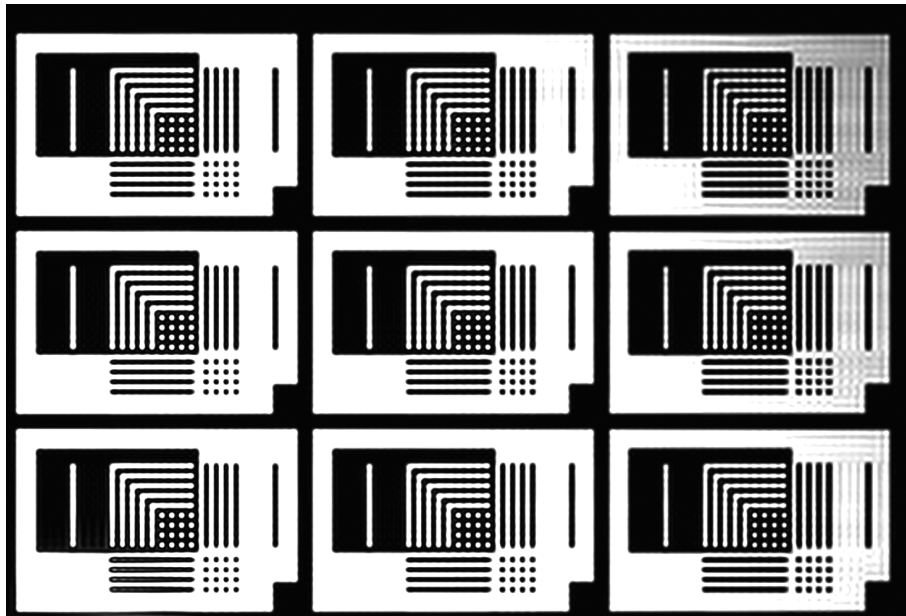


Рис. 6. Результат восстановления участка изображения для топологии, составленной из тестовых мир.

5. Заключение

В работе была исследована эффективность работы алгоритма расчета свертки суммированием излучения от прямоугольников и алгоритма двумерного БПФ на параллельных вычислительных системах кластерного типа. Показана хорошая масштабируемость первого алгоритма и низкая эффективность второго (БПФ) при увеличении кол-ва вычислительных ядер как внутри вычислительного узла, так и при увеличении кол-ва ВУ. Показана невозможность расчета обычными методами голограммы Габора для микрочипа с реальными размерами.

В статье описан метод Большого пиксела, позволяющий рассчитывать с его помощью

голограммы Габора для современных микрочипов с произвольной топологией на существующих параллельных вычислительных комплексах. Модифицированный, позволяющий еще на два порядка снизить объем требуемой оперативной памяти, вариант метода большого пиксела был реализован в составе параллельного программного комплекса BigPixel.

В случае, когда топология микросхемы имеет периодическую (регулярную) структуру, можно добиться повышения производительности алгоритма большого пиксела, если в качестве элементарных пикселей брать не маленькие квадратики, а эти повторяющиеся структуры.

В данной работе не затрагивались вопросы, связанные с улучшением получающегося изображения [12], тогда как из рис. 6 видно, что качество изображения оставляет желать лучшего. Можно предложить градиентный метод оптимизации, где для вычисления вариации соответствующего функционала необходимо будет рассчитывать несколько сверток типа (3). Следовательно, метод большого пиксела может быть применен в решении таких оптимизационных задач.

Литература

1. Борисов М.В., Боровиков В.А., Гавриков А.А., Князьков Д.Ю., Раховский В.И., Челубеев Д.А., Шамаев А.С. Методы создания и коррекции качества голографических изображений геометрических объектов с элементами субволновых размеров // Доклады Академии Наук, 2010, т. 434, № 3, стр. 332-336.
2. Борисов М.В., Гавриков А.А., Князьков Д.Ю., Раховский В.И., Челубеев Д.А., Шамаев А.С. Способ изготовления голографических изображений рисунка // Патент РФ 2396584, опубликован 10.08.2010г.
3. Rakhovskiy V.I., Borisov M.V., Shamaev A.S., Chelubeev D.A., Gavrikov A.A., Knyazkov D.U. Method of producing holographic images of IC topologies // US Patent App. №12836670, 2011.
4. Knyazkov D.Yu. Simulation of Holography Using Multiprocessor Systems // Springer LNCS, Mathematical Modeling and Computational Science, 2011, Vol. 7125 (in print).
5. Gabor D. A new microscopic principle // Nature, 1948, Vol. 161, pp. 777-778.
6. Singh V., Hu B., Bollepalli S., Wagner S., Borodovsky Y. Making a trillion pixels dance // Proc. of SPIE Vol. 6924, 0S1-0S12, 2008.
7. Stratton, J.A., Chu L.J. Diffraction theory of electromagnetic waves // Phys. Rev., 1939, Vol. 56, pp. 99-107.
8. Born M., Wolf E. Principles of Optics. Oxford, 1964
9. Басс Ф.Г., Фукс И.М. Рассеяние волн на статистически неровной поверхности. 1972. М.: Наука. 424 с.
10. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. М.: Бином, 2003, 632 с.
11. Нуссбаумер Г. Быстрое преобразование Фурье и алгоритмы вычисления сверток. М.: Радио и связь, 1985, 248 с.
12. Князьков Д.Ю. Оптимизация электромагнитных полей в голографической литографии с помощью метода локальных вариаций // Известия РАН. Теория и системы управления, 2011, № 6, стр. 99-109.