

ТЕОРИЯ МАКСИМИЗАЦИИ ДИФРАКЦИОННОЙ ЭФФЕКТИВНОСТИ ТОНКОСЛОЙНЫХ АМПЛИТУДНЫХ ГОЛОГРАММ

Э. И. Крупицкий и И. С. Барбанель

Статья посвящена решению задачи оптимизации дифракционной эффективности тонкослойных амплитудных голограмм. Решение производится как в плане нахождения оптимального по дифракционной эффективности фотоносителя (полная задача оптимизации), так и в плане нахождения оптимальных режимов регистрации голограмм на конкретных фотоносителях (частная задача оптимизации). В последнем случае задача оптимизации сводится к задаче нелинейного программирования, которая решается нами графоаналитически.

Введение

Дифракционная эффективность определяет яркость восстановленного изображения при известной мощности излучения, освещающего голограмму источника. Ввиду малой мощности современных газовых лазеров (до десятков мвт), используемых в качестве источников излучения, проблема увеличения дифракционной эффективности становится особенно острой. К сожалению, для тонкослойных амплитудных голограмм достижимая дифракционная эффективность ограничена сравнительно небольшой величиной около 10%. Для получения величин, близких к предельной, необходима оптимизация как регистрирующего фотоматериала (синтез материала с характеристикой специальной формы), так и процесса записи голограммы. Строго говоря, оптимизация материала и режима записи представляют собой задачу полной оптимизации. Целесообразность ее решения следует хотя бы из того факта, что получаемая обычно дифракционная эффективность составляет 0.5–2%, т. е. значительно уступает предельно достижимой.

В настоящей статье даны постановка и анализ задачи полной оптимизации для общего случая. Для частного случая, когда известно семейство характеристик амплитудного пропускания фотоматериала, предлагается метод решения полной задачи оптимизации, сравнительно просто реализуемый практически. Приведен пример расчета по предлагаемому методу для случая использования фотоматериала типа МИКРАТ ВР.

Постановка задачи

Дифракционная эффективность определяется, как известно, соотношением

$$\eta = \frac{\Phi_1}{\Phi_2}, \quad (1)$$

где Φ_1 — световой поток, соответствующий первому дифракционному пучку при восстановлении изображения; Φ_2 — поток света, падающий на голограмму.

Амплитуда интерференционного поля в плоскости регистрации голограммы определяется выражением

$$A^2(x; y) = A_r^2(x; y) + 2A_r(x; y)A_s(x; y)\cos[\Phi_r(x; y) - \Phi_s(x; y)] + A_s^2(x; y), \quad (2)$$

где учтено, что комплексные амплитуды опорной и предметной волн описываются соответственно функциями

$$A_r(x; y)\exp\{j\Phi_r(x; y)\} \text{ и } A_s(x; y)\exp\{j\Phi_s(x; y)\}.$$

Введем для удобства обозначения

$$\left. \begin{aligned} \mathcal{J}_r(x; y) &= \mathcal{J}_{rm}f_r(x; y), \\ \mathcal{J}_s(x; y) &= \mathcal{J}_{sm}f_s(x; y), \end{aligned} \right\} \quad (3)$$

где \mathcal{J}_{rm} и \mathcal{J}_{sm} — максимальные интенсивности опорного и предметного пучков в пределах области регистрации; $f_r(x; y)$ и $f_s(x; y)$ — нормированные к единице функции распределения интенсивности пучков. Учитывая (2) и (3), получим для экспозиции в плоскости регистрации выражение

$$E(x; y) = E_{rm}[f_r(x; y) + 2(q)^{1/2}f_r^{1/2}(x; y)f_s^{1/2}(x; y)\cos\Psi(x; y) + qf_s(x; y)], \quad (4)$$

где

$$\Psi(x; y) = \Phi_r(x; y) - \Phi_s(x; y),$$

$q = \mathcal{J}_{sm}/\mathcal{J}_{rm}$ — отношение максимальных интенсивностей интерферирующих пучков света, $E_{rm} = \mathcal{J}_{rm}t_s$ — максимальная экспозиция при воздействии только опорного пучка (предметный перекрыт), t_s — время экспозиции.

Амплитудное пропускание проэкспонированного и обработанного фотоматериала будет изменяться по закону

$$T(x; y) = F[E(x; y)], \quad (5)$$

где $F(E)$ представляет собой зависимость коэффициента амплитудного пропускания от экспозиции. Эта функция определяется свойствами фотоматериала и режимом обработки. Согласно (4), величина $T(x; y)$ может быть представлена рядом Фурье

$$T(x; y) = C_0(x; y) + 2 \sum_{n=1}^{\infty} C_n(x; y)\cos n\Psi(x; y), \quad (6)$$

где

$$C_n = \frac{1}{\pi} \int_0^{\pi} F[E(x; y)]\cos n\Psi d\Psi. \quad (7)$$

При восстановлении полезному изображению соответствует пучок света, которому соответствует амплитудное распределение на «теневой стороне» голограммы вида

$$A'_s(x; y) = A_r(x; y)C_1(x; y)\exp\{j\Phi_s(x; y)\}. \quad (8)$$

Соответствующий световой поток запишется в виде

$$\Phi_1 = \mathcal{J}_{rm} \iint_{(S_r)} C_1^2(x; y)f_r(x; y) dx dy, \quad (9)$$

где S_r — рабочая область голограммы.

Падающий на голограмму поток восстанавливающего пучка¹

$$\Phi_{\Sigma} = \mathcal{J}_{rm} \iint_{(S_r)} f_r(x; y) dx dy. \quad (10)$$

¹ При получении формул (9) и (10) для простоты принято, что восстановленный полезный пучок света и восстанавливающий образуют сравнительно небольшие углы с нормалью к поверхности голограммы.

Следовательно, с учетом (1), (9) и (10) имеем

$$\eta = \frac{\iint_{(S_r)} C_1^2(x; y) f_r(x; y) dx dy}{\iint_{(S_r)} f_r(x; y) dx dy}, \quad (11)$$

где

$$C_1(x; y) = \frac{1}{\pi} \int_0^\pi F \{ E_{rm} [f_r(x; y) + 2(q)^{1/2} f_r^{1/2}(x; y) f_s^{1/2}(x; y) \cos \Psi + q f_s(x; y)] \} \cos \Psi d\Psi. \quad (12)$$

Из (11) и (12) следует, что дифракционная эффективность η есть сложный нелинейный функционал, зависящий от параметров режима записи голограммы E_{rm} и q , характеристики фотоматериала $F(E)$ и функций распределения $f_r(x; y)$ и $f_s(x; y)$.

На практике функции f_r и f_s обычно заданы и определяются свойствами объекта и выбранной схемой голографирования. Таким образом, варьироваться могут только параметры E_{rm} и q , а также функция $F(E)$. При этом параметры E_{rm} и q должны подчиняться некоторым нелинейным ограничениям, вытекающим из условия

$$E_{\min.} \leq E(x; y) \leq E_{\max.} \quad (13)$$

Учитывая (4), нетрудно показать, что для выполнения условия (13) достаточно, чтобы параметры E_{rm} и q удовлетворяли ограничениям

$$E_{rm} (1 + \sqrt{q})^2 \leq E_{\max.}; \quad E_{rm} (1 - \sqrt{q})^2 \geq E_{\min.} \quad (14)$$

Все величины, входящие в (14), не могут быть отрицательными. Граничные значения экспозиции $E_{\min.}$ и $E_{\max.}$ определяются условиями технической реализации голограммы, а также условиями задания функции $F(E)$.² Функция $F(E)$ при решении полной задачи оптимизации может выбираться из некоторого класса G_i ($i=1, 2, \dots, N$). При оптимизации необходимым (но не достаточным) условием физической реализуемости является

$$0 \leq F(E) \leq 1. \quad (15)$$

Задача полной оптимизации сформулируется теперь следующим образом: требуется определить значения параметров режима E_{rm} и q , удовлетворяющие нелинейным ограничениям (14), а также функцию $F(H) \in G_i$, которые обеспечивают максимум функционалу η .

Данная задача относится к разряду труднейших задач нелинейного программирования [1]. Однако, как будет показано ниже, в некоторых важных частных случаях она допускает сравнительно простое решение.

Предельно достижимая эффективность голограммы

В данном разделе мы дадим строгий вывод условий достижения предельной эффективности, решив попутно частную задачу полной оптимизации.

Будем предполагать, что в (11) являются варьируемыми не только E_{rm} , q и $F(E)$, но и функции $f_r(x; y)$ и $f_s(x; y)$. При этом по физическому смыслу

$$0 \leq f_r(x; y) \leq 1; \quad 0 \leq f_s(x; y) \leq 1. \quad (16)$$

² Кроме того, величины $E_{\max.}$ и $E_{\min.}$ могут выбираться из соображений исключения сильно нелинейных режимов записи, приводящих к большим искажениям полезного изображения и появлению ярких изображений высших порядков.

Оптимальную функцию $F(E)$ будем отыскивать в классе ограниченных функций (G_1), удовлетворяющих единственному условию (15). Для числителя (11) можно записать оценку

$$\iint_{(S_T)} C_1^2(x; y) f_r(x; y) dx dy \leq \max_{(x; y)} C_1^2(x; y) \iint_{(S_T)} f_r(x; y) dx dy. \quad (17)$$

При этом знак равенства достигается в единственном случае, когда $C_1^2(x; y) = \text{const}$. С учетом (17) имеем

$$\eta \leq \max_{(x; y)} C_1^2(x; y). \quad (18)$$

Следовательно, максимум η достигается при условии, что $C_1^2(x; y) = \text{const}$. Из (12) следует, что это условие выполняется только при $f_r(x; y) = \text{const}$ и $f_s(x; y) = \text{const}$, т. е.

$$f_r(x; y) = f_s(x; y) = 1.$$

Таким образом, мы пришли к соотношению

$$\max \eta = \max_{\substack{(E_{rm}, q) \\ F \in G_1}} C_1^2(E_{rm}; q; F), \quad (19)$$

где

$$C_1^2(E_{rm}; q; F) = \frac{1}{\pi^2} \left\{ \int_0^\pi F [E_{rm}(1 + 2\sqrt{q} \cos \Psi + q)] \cos \Psi d\Psi \right\}^2. \quad (20)$$

Оптимальная функция $F \in G_1$ определяется следующим образом: для положительного фотоматериала F должна быть максимальна при $\cos \Psi > 0$ и минимальна при $\cos \Psi < 0$ (для отрицательного фотоматериала условия меняются на противоположные), т. е. с учетом условий (15)

$$\left. \begin{aligned} F [E_{rm}(1 + 2\sqrt{q} \cos \Psi + q)] &= \begin{cases} 1 & \cos \Psi > 0 \\ 0 & \cos \Psi < 0 \end{cases} \text{ для положительного} \\ & \text{фотоматериала (а),} \\ F [E_{rm}(1 + 2\sqrt{q} \cos \Psi + q)] &= \begin{cases} 0 & \cos \Psi > 0 \\ 1 & \cos \Psi < 0 \end{cases} \text{ для отрицательного} \\ & \text{фотоматериала (б).} \end{aligned} \right\} \quad (21)$$

Из (21) нетрудно видеть, что

$$F [E_{rm}(1 + 2\sqrt{q} \cos \Psi + q)] = \frac{1}{2} [1 \pm \text{sign}(\cos \Psi)] \quad (22)$$

или же, после замены переменной

$$F(E) = \frac{1}{2} [1 \pm \text{sign}(E - E_{cp})]; \quad E_{cp} - \Delta E \leq E \leq E_{cp} + \Delta E, \quad (23)$$

где

$$E_{cp} = E_{rm}(1 + q); \quad \Delta E = 2E_{rm}\sqrt{q}.$$

Знак плюс в выражении (23), определяющем вид функции, которая обеспечивает максимум функционалу (20), соответствует случаю положительного фотоматериала, минус соответствует случаю отрицательного. На рис. 1, а дан график этой функции для случая отрицательного фотоматериала. Соответствующая ему характеристическая кривая показана на рис. 1, б. Нахождение оптимальной функции определило в данном случае и оптимальный режим записи, который не требует дополнительного пояснения. Подставляя (23) в (20), получим

$$\max \eta = \max C_1^2 = \frac{1}{\pi^2} \left| \int_0^{\pi/2} \cos \Psi d\Psi \right|^2 = 10.15\%. \quad (24)$$

Приведенное доказательство является строгим и позволяет утверждать, что ни при каких условиях эффективность тонкослойной амплитудной голограммы не может превышать величину 10.15%. Заметим, что этой величине соответствует так называемая бинарная голограмма — для нее характерен высокий уровень яркости изображений высоких порядков.³

Решение полной задачи оптимизации в случае известного фотоматериала

Решение, найденное в предыдущем разделе, имеет тот недостаток, что оно соответствует идеальному фотоматериалу. Характеристика амплитудного пропускания реального материала не может иметь вид (23) из-за разрывности функции $\text{sign}(E - E_{cp.})$. Другим недостатком этого решения является то, что оно соответствует случаю равномерного распределения яркостей пучков при голографировании, что не всегда выполняется на практике. В связи с этим возникают две проблемы: проблема синтеза практически реализуемого фотоматериала с оптимальной характеристикой $F(E)$ и проблема полной оптимизации в случае готового материала, для которого известно семейство функций $F_j(E)$, а именно: $\{F_j(E)\}_{j=1}^n$. Отдельные функции $F_j(E)$ могут, в частности, соответствовать различным режимам обработки фотоматериала.

В настоящей работе мы ограничимся решением лишь второй задачи. Обозначим соответствующий узкоспециальный класс функций $F(E)$ через G_2 , учитывая, что под G_2 понимается семейство $\{F_j(E)\}_{j=1}^n$. Полная задача оптимизации может быть в данном случае решена следующим образом. Введем в рассмотрение величины $\eta_j = \eta(E_{rm}; q; F_j)$; $j=1, 2 \dots n$, где функционал η определяется выражениями (11) и (12). Определим максимум каждой из величин η_j по переменным E_{rm} и q с учетом ограничений (14). Тогда получим набор величин η_{jmax} . После этого остается определить номер j , которому соответствует наибольшее из чисел совокупности $\{\eta_{jmax}\}_{j=1}^n$. Тем самым определяется как вид оптимальной характеристики $F_j(E)$, так и соответствующие ей параметры оптимального режима записи $E_{rm}^{(j)}$ и $q^{(j)}$.

Реализация решения может быть осуществлена методом нелинейного программирования [1], причем благодаря наличию только двух переменных E_{rm} и q решение легко получить графо-аналитическим способом. Для простоты пояснения метода будем полагать, что семейство $\{F_j\}$ состоит только из одной функции. Во-первых, на плоскости $(E_{rm}; q)$ строятся кривые, соответствующие равенствам

$$\left. \begin{aligned} E_{rm}(1 - \sqrt{q})^2 &= E_{мин.}, \\ E_{rm}(1 + \sqrt{q})^2 &= E_{макс.} \end{aligned} \right\} \quad (25)$$

которые определяют область допустимых значений E_{rm} и q в соответствии с неравенствами (14). Далее, полагая последовательно $\eta(E_{rm}; q; F) = k_l$, где $l=1, 2 \dots M$, причем $k_1 < k_2 < \dots < k_M$, строим соответствующие кривые $E_{rm} = \Phi(q; k_l)$.

Из множества кривых, пересекающих область допустимых значений E_{rm} и q , выбирается та, которая соответствует наибольшему значению

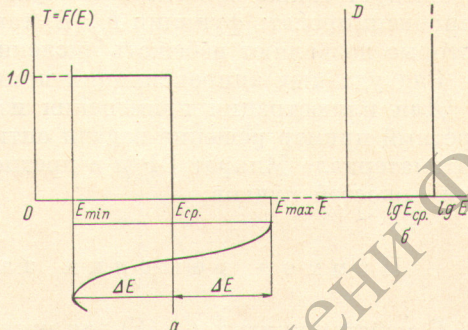


Рис. 1. Оптимальные характеристики фотоматериала.

а — амплитудное пропускание, б — характеристическая кривая.

³ Максимально достижимая эффективность амплитудной голограммы при отсутствии нелинейных искажений составляет 6.25%.

параметра k_i . Кривая, как правило, касается границы указанной области в точке, координаты которой и определяют оптимальные значения E_{rm} и q . Максимум параметра k_i соответствует максимуму эффективности η . Для выбора оптимальной функции $F_j(E)$ такого рода построение должно быть проделано для всех $j=1, 2 \dots n$. Ввиду однотипности необходимых вычислений они легко могут быть проделаны на ЦВМ. Решение задачи по предлагаемому методу существенно упрощается в случае, когда распределения интенсивности пучков равномерные.⁴ При этом вместо (11) получаем

$$\eta = C_1^2(E_{rm}; q; F), \quad (26)$$

где C_1^2 определяется формулой (20). Ограничения же на параметры E_{rm} и q по-прежнему будут иметь вид (14).

Следует также отметить, что для упрощения решения целесообразно аппроксимировать функции $F_j(E)$ такими аналитическими выражениями, которые позволяют избежать численного интегрирования в (12) и (20).

Ряд удобных аппроксимаций приведен в работе [2].

Для иллюстрации применения изложенного метода рассмотрим конкретный пример решения задачи оптимизации для случая использования фотоматериала Микрат ВР и в предположении равномерного распределения яркостей пучков.

Пример решения задачи оптимизации

Семейство кривых $F_j(E)$ для материала Микрат ВР, соответствующих различному времени проявления $t_{пр}$, приведено на рис. 2 [3]. Они хорошо аппроксимируются функциями

$$F_j(E) = \beta_j e^{-\alpha_j (E - \delta_j)} \quad (27)$$

путем выбора коэффициентов β_j ; α_j ; и δ_j ⁵ (соответствующие аппроксимирующие кривые показаны на рис. 2 штриховыми линиями, а значения параметров аппроксимации — в таблице).

Учитывая соотношение (21) и формулу

$$e^{-z \cos \Phi} = I_0(z) + + 2 \sum_{n=1}^{\infty} (-1)^n I_n(z) \cos n\Phi, \quad (28)$$

где $I_n(z)$ — функции Бесселя от чисто мнимого аргумента, легко получим

$$\eta_j(X; Y) = P_j e^{-2\alpha_j X} I_1^2(Y). \quad (29)$$

Здесь обозначено

$$\left. \begin{aligned} P_j &= (\beta_j e^{\alpha_j \delta_j})^2, \\ X &= E_{rm} (1 + q) = E_{ср}, \\ Y &= 2E_{rm} \sqrt{q} = \Delta E. \end{aligned} \right\} \quad (30)$$

⁴ Это, в частности, имеет место при голографировании фазовых и диффузно рассеивающих объектов.

⁵ Граничные экспозиции $E_{j\text{мин}}$ и $E_{j\text{макс}}$ выбраны в примере так, чтобы исключить сильно нелинейные режимы записи. Для $j=1$ более удобна аппроксимация $T = -\beta_1 \ln \alpha_1 H$; дифракционная эффективность в этом случае определяется как $\eta = \beta^2 q$.

Значения параметров аппроксимации

| j | Время проявления $t_{пр. j}$, мин. | α_j | β_j | δ_j | $E_{j \text{ мин.}}$ | $E_{j \text{ макс.}}$ |
|-----|-------------------------------------|------------|-----------|------------|----------------------|-----------------------|
| 1 | 1 | 0.023 | 0.248 | — | 1 | 36 |
| 2 | 2 | 0.18 | 0.15 | 11 | 0.5 | 17 |
| 3 | 4 | 0.32 | 1.065 | 0 | 0.56 | 9 |
| 4 | 8 | 0.76 | 1.1 | 0 | 0.56 | 4 |
| 5 | 16 | 1.33 | 1.08 | 0 | 0.32 | 3 |

Все построения удобнее в данном случае проводить на плоскости (X, Y) вместо плоскости (E_{rm}, q) . Для определения величин E_{rm} и q следует воспользоваться соотношениями

$$E_{rm} = \frac{\sqrt{X+Y} \pm \sqrt{X-Y}}{4} \quad (31)$$

$$q = \left(\frac{\sqrt{X+Y} \mp \sqrt{X-Y}}{\sqrt{X+Y} \pm \sqrt{X-Y}} \right)^2 \quad (32)$$

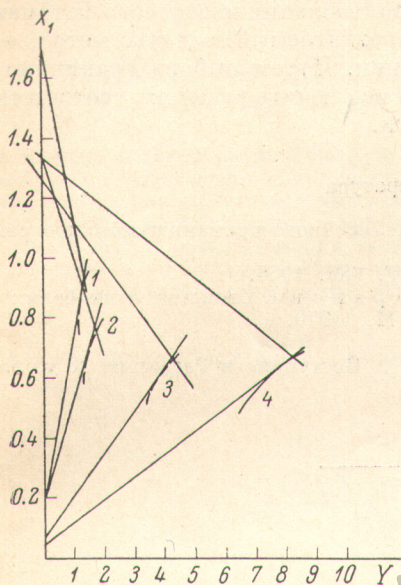


Рис. 3. Графическое построение.

1 — $\eta_5 = 2.7\%$; 2 — $\eta_4 = 2.74\%$; 3 — $\eta_3 = 3.75\%$; 4 — $\eta_2 = 4.84\%$.

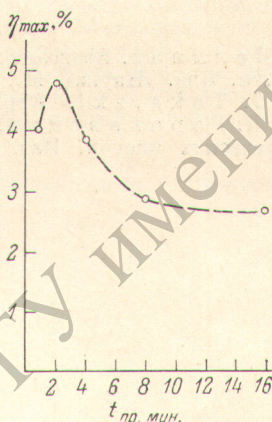


Рис. 4. Зависимость максимальной дифракционной эффективности от времени проявления.

Двузначность определяет то положение, что при равномерном распределении интенсивностей пучков, случай, когда интенсивность опорного пучка больше интенсивности предметного (верхние знаки в (31) и (32)), эквивалентен обратному.

Картина графического построения приведена на рис. 3. Области допустимых значений для каждого j представляют собой треугольники.⁶ Штриховыми линиями показаны кривые $X_1 = f(Y; k_l)$, соответствующие $\eta_j = \eta_{j \text{ макс}}$ для $j = 3, 4, 5$. Сплошной линией показана кривая, соответствующая $\max(\eta_{j \text{ макс}}) = 4.84\%$ (ей соответствует номер $j = 2$), т. е. время проявления $t_{пр.} = 2$ мин. Оптимальные параметры режима записи имеют при этом значения

$$E_{rm} = 1.42 \text{ (или } 0.05), q = 0.026 \text{ (или } 40). \quad (33)$$

На рис. 4 приведена зависимость η_{max} от времени проявления, из которой следует, что величина η_{max} имеет резко выраженный максимум при $t_{пр} = 2$ мин. При отклонении от этой величины возможен существенный проигрыш по дифракционной эффективности. Это говорит о важности оптимального выбора кривой $F(E)$ из возможного семейства.

⁶ Для упрощения графо-аналитического построения, которое подробно в данной статье не рассматривается, во всех случаях $j = 2, \dots, 5$ в уравнениях (29) и ограничениях (30) произведена замена переменной вида $X_1 = \alpha_j X / 2.303$.

Значительное отличие полученной величины $\max \eta$ от предельной, равной 10.15%, объясняется неблагоприятной формой всех кривых заданного семейства $\{F_j(E)\}$ и исключением сильно нелинейных режимов записи.

З а к л ю ч е н и е

В работе строго доказано, что предельно достижимая дифракционная эффективность тонкослойных амплитудных голограмм равна 10.15%, и реализуется только при равномерном распределении интенсивностей предметного и опорного пучков по области регистрации. Получение максимальной эффективности требует совместной оптимизации как режима записи голограммы, так и синтеза характеристики пропускания фотоматериала, т. е. решения полной задачи оптимизации.

Для решения задачи в случае использования известного фотоматериала может быть эффективно применен предлагаемый в статье метод, основанный на нелинейном программировании. Путем выбора граничных экспозиций $E_{j\text{мин}}$ и $E_{j\text{макс}}$ метод позволяет исключить решения, соответствующие сильно нелинейным режимам записи.

Литература

- [1] Р. Беллман. Введение в теорию нелинейного и динамического программирования. Изд. «Наука», М., 1968.
- [2] М. Ж. Токарский. Appl. Optics, 7, 989, 1968.
- [3] Ю. Н. Гороховский, В. П. Баранова. Свойства черно-белых фотографических пленок. Изд. «Наука», М., 1970.

Поступило в Редакцию 20 июля 1971 г.
