Исследование возможности построения блочного алгоритма BPM-метода

Д.Л. Головашкин[1,2], Д.С. Калачян[2], Л.В. Яблокова[2]

[1]Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001
[2]Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. Работа посвящена применению методики построения разностных схем на основе комбинации явного и неявного дифференциальных шаблонов при разработке блочного алгоритма BPM-метода. Особое внимание уделяется экспериментальному исследованию эффективности построенного блочного алгоритма и сходимости разностного решения. Подтверждено трехкратное ускорение вычислений по сравнению с классической явной разностной схемой, выявлены особенности поведения разностного решения при сгущении сеточной области.

1. Введение

Развитие компьютерной оптики [1], как предметной области, объединяющей вычислительные и оптические технологии, принято связывать с одновременным совершенствованием методов производства оптических элементов и методов их расчета. Видное место среди последних занимает метод распространяющегося пучка (Beam propagation method, BPM); в частности его вариант, основанный на конечно-разностном решении уравнения Гельмгольца (Finite-difference Beam propagation method, FD-BPM) в “приближении медленно меняющейся огибающей” (Slowly varying envelop approximation, SVEA) [2]. Программная реализация такого решения – BeamPROP, интегрированная в среду RSoft CAD Environment компании Synopsys, широко используется при расчете распространения света в оптических волноводах [4,5].

Перечисляя проблемы разностного решения уравнений в частных производных, среди прочего принято указывать на высокую длительность вычислений. Особенно остро эта проблема проявляется в случае протяженных вычислительных областей, которые приходится задавать при моделировании распространения излучения в упомянутых волноводах. Ее решение традиционно видят в организации параллельных вычислений по BPM-методу на многоядерных процессорных архитектурах, в том числе графических процессорах [6,7].

Вместе с тем, результаты недавних исследований в области разработки блочных алгоритмов для разностных схем свидетельствуют о возможности в несколько раз снизить длительность вычислений за счет совершенствования собственно алгоритмов расчета. Так, при разностном решении уравнений Максвелла (FDTD-метод, Finite-Difference Time-Domain) по блочному алгоритму, вычисления ускоряются на той же аппаратной базе и без изменения разностной схемы [8]; то есть модифицируется алгоритм метода (задается новый порядок обхода узлов сеточной области), но не сам численный метод. Все математические свойства последнего (аппроксимация и устойчивость) остаются прежними. Использование блочности не исключает переход к параллельным вычислениям, позволяя синтезировать блочно-параллельные алгоритмы FDTD-метода, характеризующиеся большей эффективностью, чем просто параллельные [9]. Уместным применение блочности оказалось и при организации расчетов по FDTD-методу на графических процессорах [10]. Оправданным также является переход к блочным алгоритмам при разностном решении волнового уравнения на центральном процессоре [11] и видеокарте [12].

В связи с изложенным представляется актуальным исследовать возможность синтеза блочного алгоритма BPM-метода и его эффективность, ведь применение блочных алгоритмов других численных методов компьютерной оптики (FDTD и разностного решения волнового уравнения) оказалось удачным [8-12].

2. Выбор уравнения и разностных схем

В фундаментальной монографии [2] представлен вывод из уравнения Гельмгольца обобщенной формы уравнения Фока-Леонтовича, являющегося в указанной работе отправной точкой при описании FD-BPM метода:

$$\frac{\partial \Psi}{\partial z} = \frac{-i}{2k_0 n_*}\left(\frac{\partial}{\partial x^2} + \frac{\partial}{\partial y^2} + k_0^2 n^2 - k_0^2 n_*^2\right)\Psi, \tag{1}$$

где Ψ – комплексная амплитуда (описывающая огибающую поля), n_* – базовый показатель преломления (считается, что в целом распространение излучения сходно с распространением в однородной среде с таким показателем), k_0 – волновое число, n – показатель преломления.

Для целей настоящей работы удобно максимально упростить (1), сведя его к уравнению Шредингера из [13], где детально обсуждаются разностные схемы для такого уравнения:

$$i\frac{\partial \Psi}{\partial z} = \frac{\partial \Psi}{\partial x^2}. \tag{2}$$

При рассмотрении распространения волны выбранной длины (для которой $2k_0 n_* = 1$) в однородной ($n_*=n$) двумерной ($\Psi=\Psi(x,z)$) среде (1) действительно переходит в (2). Завершая постановку задачи примем в качестве граничных условий $\Psi(0,z)=\Psi(1,z)=0$ и начального условия (направление распространения вдоль z можно уподобить оси времени) $\Psi(x,0)=\sqrt{2}\sin(\pi x)$. Очевидное аналитическое решение задачи: $\Psi(x,z)=\sqrt{2}\sin(\pi x)\exp(i\pi^2 z)$ на вычислительной области $\omega=\{(x,z)$: $0\le x\le 1$, $0\le z\le L\}$ будем использовать для оценки погрешности разностного решения.

Выбирая из многообразия представленных в [13] разностных схем для уравнения (2), ограничимся далее рассмотрением простейших явной:

$$i\frac{\Psi_j^{k+1} - \Psi_j^k}{h_z} = \frac{\Psi_{j-1}^k - 2\Psi_j^k + \Psi_{j+1}^k}{h_x^2}; \tag{3}$$

и неявной двухслойных схем:

$$i\frac{\Psi_j^{k+1} - \Psi_j^k}{h_z} = \frac{\Psi_{j-1}^{k+1} - 2\Psi_j^{k+1} + \Psi_{j+1}^{k+1}}{h_x^2}, \tag{4}$$

где функция Ψ_j^k определена на сеточной области $\omega_h=\{(x_j,z_k)$: $x_j=j\cdot h_x$, $0\le j\le J$, $h_x=1/J$; $z_k=k\cdot h_z$, $0\le k\le K$, $h_z=L/K\}$.

Авторам известно, что в вычислительной практике при решении (1) по BPM-методу принято использовать схемы типа Кранка-Николсон (CN) [2], однако составление блочных алгоритмов уместней начать с более простых примеров.

3. Смешанная разностная схема и блочный алгоритм

К сожалению, блочные алгоритмы решения сеточных уравнений двухслойных разностных схем, разработанные без модификации самих схем, в известной авторам литературе не встречаются. Упомянутые алгоритмы решения уравнений Максвелла и волнового основаны на трехслойных схемах. Вместе с тем, в работе [14] приводится успешный прием составления смешанной разностной двухслойной схемы для уравнения теплопроводности, применение которого позволяет сформулировать блочный алгоритм решения сеточных уравнений смешанной схемы. Суть приема состоит в комбинации явного и неявного дифференциальных шаблонов на одном слое по времени, при которой сеточная функция в одних узлах вычисляется по явному шаблону, в других по неявному. На следующем слое шаблоны меняются местами. Так, на нечетных слоях в нечетных узлах применяется явный шаблон (3), затем в четных – неявный (4); на четных слоях наоборот: в четных узлах применяется явный шаблон (3), в нечетных – неявный (4).

В рассматриваемом случае BPM-метода не блочный алгоритм вычислений по смешанной схеме, записанный на языке фортран, выглядит следующим образом.

```
c1=1+2*i*hz/hx/hx; c2=i*hz/hx/hx; ! коэффициенты явной части разностной схемы
c3=1/(1-2*i*hz/hx/hx); c4=i*hz/hx/hx/(1-2*i*hz/hx/hx); ! коэффициенты неявной части
do k=1,K,2 ! перебор слоев сеточной области с шагом 2
    Psi(3:J-1:2)=c1*Psi(3:J-1:2)-c2*(Psi(2:J-2:2)+Psi(4:J:2)); ! явный шаблон (3), нечетные узлы
    Psi(2:J-1:2)=c3*Psi(2:J-1:2)-c4*(Psi(1:J-2:2)+Psi(2:J:2)); ! неявный шаблон (4), четные узлы
    Psi(2:J-1:2)=c1*Psi(2:J-1:2)-c2*(Psi(1:J-2:2)+Psi(2:J:2)); ! явный шаблон (3), нечетные узлы
    Psi(3:J-1:2)=c3*Psi(3:J-1:2)-c4*(Psi(2:J-2:2)+Psi(4:J:2));!неявный шаблон(4), нечетные узлы
enddo
```

Теперь становится возможной организация обхода сеточной области, при котором перебираются узлы внутри блоков, имеющих форму параллелепипедов (рисунок 1 из [14]). Ни для явной, ни для неявной схем, такой обход невозможен [14]. Аналогичные форма блоков и правило обхода характеризуют блочные алгоритмы разностного решения уравнений Максвелла и волнового уравнения.

Не решаясь затруднить читателя представлением всего блочного алгоритма, ограничимся демонстрацией его фрагмента для центральных блоков одного блочного слоя, где m – ширина блока, delta – его высота, а P – число блоков в одном блочном слое.

```
do p=2,P-1 ! проход по блокам
    jl=(p-1)*m+2; jr=jl+m-1; ! границы блока
    do k=2,delta,2 ! цикл по парам слоев
        Psi(jl+1:jr:2)=c1*Psi(jl+1:jr:2)-c2*(Psi(jl:jr-1:2)+Psi(jl+2:jr+1:2)); ! вычисления по ф. (3)
        Psi(jl:jr:2)=c3*Psi(jl:jr:2)-c4*(Psi(jl-1:jr-1:2)+Psi(jl+1:jr+1:2)); ! вычисления по ф. (4)
        Psi(jl:jr-1:2)=c1*Psi(jl:jr-1:2)-c2*(Psi(jl-1:jr-2:2)+Psi(jl+1:jr:2)); ! вычисления по ф. (3)
        Psi(jl-1:jr-1:2)=c3*Psi(jl-1:jr-1:2)-c4*(Psi(jl-2:jr-2:2)+Psi(jl:jr:2)); ! вычисления по ф. (4)
        jl=jl-2; jr=jr-2; ! новые границы блока
    enddo
enddo
```

Весьма схожие алгоритмы (не блочный и блочный), правда без представления программной реализации, приводились в [14] для уравнения теплопроводности и характеризовались пятикратным ускорением вычислений по сравнению с классической явной схемой. Переход к BPM-методу теоретически должен сопровождаться снижением ускорения: работа с комплексными числами обуславливает увеличение объема коммуникаций между оперативной памятью и кэш-памятью процессора вдвое, а количество арифметических операций возрастает на большую величину. Следовательно, доля последних в общей длительности расчетов увеличится, а эффект от блочности, связанный с сокращением коммуникаций, уменьшится. Проверим это предположение на практике.

4. Экспериментальное исследование блочного алгоритма

Эксперименты по исследованию ускорения блочного алгоритма относительно программной реализации явной разностной схемы (3) проводились на процессоре Intel Core i7-4500U (2,4 GHz, кэш-память L3 4Мб), работающем под управлением операционной системы Ubuntu 16.04 и с использованием компилятора gfortran 5.3. В таблице 1 представлены результаты, подтверждающие теоретические предположения; для блочного алгоритма задавались параметры delta=20, m= $J\times10^5$.

Таблица 1. Ускорение вычислений (S) по блочному алгоритму (время расчетов Тбл.) относительно явной схемы (Тявн.) при K=120.

$J\times10^6$	10	20	30	40	50
Тявн.,с.	9,76	19,2	28,55	38,19	48,02
Тбл.,с.	3,5	6,5	9,79	13,22	17,17
S	2,79	2,95	2,92	2,89	2,8

Ускорение снизилось до трех раз (в [14] при решении задачи теплопроводности оно достигало пяти), оставаясь при этом достаточно существенным: ведь для его получения не были задействованы новые вычислительные потоки. Сокращение длительности вычислений по блочному алгоритму связано с уменьшением интенсивности коммуникаций между оперативной памятью и кэш-памятью процессора.

Длительность вычислений по неявной схеме (4), характеризующейся большей вычислительной сложностью по сравнению с явной, будет еще выше. Вместе с тем, согласно [13] такая схема абсолютно устойчива, в то время, как явная (3) характеризуется сверхжестким критерием устойчивости ($h_z / h_x^4 < const.$). Что при этом можно сказать об устойчивости новой смешанной схемы? В [14] при исследовании смешанной схемы для уравнения теплопроводности этот вопрос не обсуждался.

Экспериментально изучая поведение погрешности разностного решения для предложенной здесь смешанной схемы авторы настоящей работы пришли к выводу, что новая схема характеризуется тем же критерием устойчивости, что и явная. Однако, рассматривая поведение погрешности модуля разностного решения (а именно квадрат модуля определяет интенсивность излучения, интересующую специалистов-оптиков при решении (1)) авторы получили достаточно интересные результаты, представленные в таблице 2.

Таблица 2. Погрешность модуля решения для явной (ε явн.), неявной (ε н/явн.) и смешанной (ε смеш.) разностных схем при hz/hx<const.

J	10	20	30	40	50
ε явн.	4,49E+01	1,77E+19	6,45E+43	1,18E+69	9,78E+95
ε н/явн.	1,35E+00	1,26E+00	1,12E+00	9,91E-01	8,80E-01
ε смеш.	7,79E-01	2,62E-01	1,43E-01	9,60E-02	7,21E-02

Как видно из таблицы, со сгущением сеточной области при соблюдении критерия Куранта, налагаемого на соотношение шагов, явная разностная схема становится неустойчивой (как и ожидалось), неявная сохраняет устойчивость, а смешанная ведет себя как неявная. Более того, при выбранных параметрах сеточной области погрешность модуля решения по смешанной схеме значительно ниже, чем по неявной. Этот занятный экспериментальный результат безусловно нуждается в теоретическом истолковании, которым может заняться любознательный читатель.

5. Заключение

В работе предложены смешанная разностная схема для решения уравнения Фока-Леонтовича, связанного с BPM-методом, и блочный алгоритм организации расчетов по ней. Продемонстрированы эффективность упомянутого алгоритма и некоторые свойства разностного решения. Авторы полагают полученные результаты началом исследования и считают необходимым далее перенести использованную здесь методику конструирования блочных алгоритмов на более сложные схемы типа Кранка-Николсон, широко применяемые в BPM-методе.

6. Благодарности

Работа выполнена при поддержке гранта РФФИ 19-07-00423 А.

7. Литература

[1] Сойфер, В.А. Дифракционная компьютерная оптика – М.: Физматлит, 2007. – 736 с.
[2] Сойфер, В.А. Дифракционная нанофотоника – М.: Физматлит, 2011. – 680 с.
[3] BeamPROP Product Overview [Электронный ресурс]. – Режим доступа: https://www.synopsys.com/optical-solutions/rsoft/passive-device-beamprop.html.
[4] Shvets, G. Polarization properties of chiral fiber gratings // Journal of Optics. – 2009. – Vol. 11(7).
[5] Qian, J. Coupled-Mode Analysis for Chiral Fiber Long-Period Gratings Using Local Mode Approach // IEEE J. of Quantum Electronics. – 2011. – Vol. 47(11).
[6] Shaaban, A. Fast parallel beam propagation method based on multi-core and many-core architectures / A. Shaaban, M. Sayed, M. Farhat // Optik. – 2019. – Vol. 180. – P. 484-491.
[7] Golovashkin, D.L. Solving finite-difference equations for diffractive optics problems using graphics processing units / D.L. Golovashkin, D.G. Vorotnikova, A.V. Kochurov, S.A. Malysheva //Optical Engineering. – 2013. – Vol. 52(9). – P. 091719-1-091719-5.
[8] Takeshi, M. Automatic Parameter Tuning of Three-Dimensional Tiled FDTD Kernel / M. Takeshi // High Performance Computing for Computational Science – VECPAR. – 2014. – Vol. 8969. – P. 284-297.
[9] Orozco, D.A. Mapping the FDTD Application to Many-Core Chip Architectures / D.A. Orozco, G.R. Gao // Parallel Processing, 2009. – P. 309-316.
[10] Перепелкина, А.Ю. Алгоритм DiamondTorre и высокопроизводительная реализация FDTD метода для суперкомпьютеров с графическими ускорителями / А.Ю.Перепелкина, А.В. Закиров, В.Д. Левченко – М.: Ин-т прикладной математики им. М.В. Келдыша РАН, 2015. – 22 с.
[11] Яблокова, Л.В. Блочные алгоритмы совместного разностного решения уравнений Даламбера и Максвелла / Л.В. Яблокова, Д.Л. Головашкин // Компьютерная оптика. – 2018. – Т. 42, № 2. – С. 320-327.
[12] Perepelkina, A.Yu. Diamond Torre Algorithm for High-Performance Wave Modeling / A.Yu. Perepelkina, V.D. Levchenko / Keldysh Institute Preprints. – 2015. – Vol. 18. – P. 20.
[13] Самарский, А.А. Теория разностных схем: учеб. пособие – М.: Наука, 1977. – 656 с.
[14] Головашкин, Д.Л. Ускорение вычислений по блочным алгоритмам разностного решения уравнения теплопроводности / Д.Л. Головашкин, Л.В. Яблокова, И.Д. Резник // Сборник трудов V Международной конференции и молодёжной школы «Информационные технологии и нанотехнологии» (ИТНТ) – Самара: Новая техника, 2019. – Т. 4. – С. 601-606.

Investigation of the possibility of building a block algorithm of the BPM method

D.L. Golovashkin[1,2], D.S. Kalachian[2], L.V. Yablokova[2]

[1]Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS, Molodogvardejskaya street 151, Samara, Russia, 443001
[2]Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. The article deals with building methodology application of the difference schemes. The basis of the proposed approach is a combination of explicit and implicit differential patterns during developing a block algorithm using the BPM method. Special attention is given to practical investigation of the block algorithm building efficiency and the convergence of the difference solution. Threefold acceleration of calculations is confirmed in comparison with the classical explicit difference scheme. The features of the difference solution behavior when the grid region thickening are revealed.