

УДК 519.63

## УСТОЙЧИВОСТЬ ЯВНЫХ СХЕМ РЕШЕНИЯ УРАВНЕНИЙ МАКСВЕЛЛА МЕТОДОМ КОНТРОЛЬНЫХ ОБЪЕМОВ ВЫСОКОГО ПОРЯДКА ТОЧНОСТИ

Д. К. Фирсов<sup>1</sup>

Предложен новый критерий устойчивости явных схем решения уравнений Максвелла методом контрольных объемов высокого порядка точности. Доказательство опирается на обобщение критерия устойчивости для схемы, основанной на аппроксимации первого порядка точности по пространству, на схемы более высокого порядка точности. Приводится оценка влияния разрывов в решении на устойчивость схем высокого порядка точности. Обсуждается принцип максимума для метода контрольных объемов, аппроксимирующих векторные законы сохранения.

**Ключевые слова:** уравнения Максвелла, метод контрольных объемов, устойчивость явных схем, высокий порядок точности, уравнения в частных производных.

Решение динамических уравнений в частных производных первого порядка зачастую проводится с использованием явных схем. Одним из стандартных требований к явным схемам является необходимость использования максимально возможного шага по времени, при котором схема сохраняет абсолютную устойчивость. Более того, для ускорения решения при помощи явных схем обычно применяется метод на основе локальных шагов по времени [1], в котором требуется надежная оценка шага по времени для каждого из элементов. Так, для получения максимально возможного шага по времени на структурированных сетках обычно применяется анализ фон Неймана [2]. Однако этот анализ не применим на неструктурированных сетках, для которых шаг по времени меняется от элемента к элементу и зависит от их геометрии. По этой причине для анализа устойчивости явных схем на неструктурированной сетке используется подходящая оценка энергетической нормы, например дискретная норма  $L_2$  оператора пространственной дискретизации [3, 4]. Необходимо также отметить, что для сходного набора задач существуют оценки и на основе нормы  $L_\infty$  [5, 6].

В настоящей статье обсуждается новый критерий устойчивости для явных схем на основе метода контрольных объемов на неструктурированной сетке для уравнений Максвелла [7, 8], записанных в виде

$$\begin{cases} \epsilon \partial_t \mathbf{E} - \nabla \times \mathbf{H} + \sigma \mathbf{E} = -\mathbf{J}, \\ \mu \partial_t \mathbf{H} + \nabla \times \mathbf{E} = \mathbf{0}, \end{cases}$$

где  $\epsilon$  — диэлектрическая проницаемость (Ф/м),  $\mu$  — магнитная проницаемость (Н/А<sup>2</sup>),  $\mathbf{E}$  — напряженность электрического поля (В/м) и  $\mathbf{H}$  — напряженность магнитного поля (А/м).

Явные схемы аппроксимации для уравнений Максвелла интересны тем, что требуют использования векторных законов сохранения с учетом векторных потоков между элементами на основе сложного взаимодействия между компонентами вектора потока, что влечет за собой необходимость применения матричной записи противопоточной схемы. В этой связи, исследование устойчивости аппроксимации методом контрольных объемов уравнений Максвелла позволят сформировать достаточно универсальные подходы к изучению устойчивости схем на основе векторных законов сохранения.

Критерий устойчивости для явных схем высокого порядка точности интегрирования по времени связан с критерием устойчивости для схемы Эйлера при условии применения пространственных аппроксимаций того же самого порядка. Например, схема предиктор–корректор устойчива при  $\Delta t^{\text{CFL}} = 2\Delta t_E^{\text{CFL}}$ , где  $\Delta t_E^{\text{CFL}}$  — максимально возможный шаг по времени, удовлетворяющий критерию устойчивости CFL (Courant–Friedrichs–Lewy) для схемы Эйлера. При этом для схем Рунге–Кутты третьего и четвертого порядков, представленных в работах [9, 10], критерий устойчивости может быть в 3–4 раза больше, чем критерий устойчивости для схем Эйлера. Таким образом, оценка устойчивости явных схем проводится для схемы Эйлера, а устойчивость для схем более высокого порядка точности по времени получается в виде  $\Delta t^{\text{CFL}} = a\Delta t_E^{\text{CFL}}$ , где  $a$  — коэффициент подобия схемам первого порядка точности.

<sup>1</sup> Geomodeling Technology Corp., 1100-665 8 Street SW, Suite 1100, Calgary, AB T2P 3K7, Canada, Numerical Programmer, e-mail: d.k.firsov@gmail.com; Исследовательский Томский государственный университет, лаборатория молекулярной динамики, просп. Ленина, 36, 634050, г. Томск, ст. науч. сотр., e-mail: d.k.firsov@gmail.com

Оценки шага по времени, полученные в работах [3, 11, 12], для метода контрольного объема при решении уравнений Максвелла хотя и являются достаточными, но не содержат точной оценки и состоятельны только для схем первого порядка точности. Вследствие этого в общем случае применение таких оценок приведет либо к меньшему шагу по времени, чем это возможно, либо к нарушению устойчивости схемы высокого порядка точности.

Кроме того, для исключения ограничений на потоки брался либо небольшой шаг по времени [1], либо применялся принцип максимума для скалярных законов сохранения [13]. В результате численные схемы имели меньшую скорость сходимости и более низкое качество результатов. Ограничение же на потоки между соседними элементами или принцип максимума [13] не был объяснен для векторных законов сохранения. Итак, в работах [3, 11] были представлены достаточные критерии устойчивости для шага по времени на неструктурированной сетке для аппроксимации первого порядка точности. В 1999 г. в работе [11] была получена оценка

$$\Delta t = \min_{1 \leq i \leq N} \frac{V_i}{c_i A_i}, \tag{1}$$

где  $i$  — индекс элемента,  $c_i$  — скорость света,  $V_i$  — объем элемента и  $A_i$  — площадь внешней поверхности элемента. Однако уже в 2000 г. эта оценка была улучшена в два раза [3]:

$$\Delta t = \min_{1 \leq i \leq N} \frac{2V_i}{c_i A_i}. \tag{2}$$

Тем не менее, критерий (2) не является необходимым, поэтому возможно получить оценку для большего шага по времени. Например, на структурированной сетке, состоящей из кубических элементов с одинаковыми диэлектрическими свойствами, оценка (2) дает соотношение  $\Delta t = \frac{h}{3c}$ , что в 1.5 раза меньше, чем получаемая из анализа фон Неймана [11]:  $\Delta t = \frac{h}{2c}$ . Именно по этой причине критерий (2) не является необходимым. Поэтому в 2007 г. для схемы первого порядка точности был получен более универсальный вычислимый критерий устойчивости [15], который детально описан в работе [8]. Последний критерий сходимости совпал с анализом фон Неймана на структурированной сетке [11], состоящей из кубических элементов. Более того, этот критерий позволил получить аналогичный результат для равно-сторонних тетраэдров:

$$\Delta t = \min_{1 \leq i \leq N} \frac{3V_i}{c_i A_i}.$$

Целью настоящей работы является формулирование критерия сходимости для схем, имеющих высокий порядок точности для пространственных производных, и получение обобщения принципа максимума на векторные законы сохранения. Важно отметить следующее: доказательство того факта, что критерий является необходимым на неструктурированной сетке, не является тривиальным, поэтому мы можем говорить только лишь о достаточном критерии сходимости. Так же как и в доказательстве, приведенном в работах [3, 8], вывод базируется на использовании предположения, что расчетная область не имеет источников при переходе с одного шага по времени на другой; следовательно, энергия при интегрировании по времени не нарастает. Тогда энергия электромагнитного поля в области  $\Omega \subset \mathbb{R}^3$ , покрытой сеткой  $\Delta = \left\{ T_i : \bigcup_{i=1}^N T_i = \Omega \right\}$ , состоящей из элементов  $T_i$ , определяет норму решения следующим образом:

$$\|\mathbf{u}\| = \sqrt{\int_{\Omega} \epsilon \mathbf{E} \cdot \mathbf{E} + \mu \mathbf{H} \cdot \mathbf{H} \, dx} = \sqrt{\sum_{i=1}^N \int_{T_i} \epsilon \mathbf{E} \cdot \mathbf{E} + \mu \mathbf{H} \cdot \mathbf{H} \, dx}, \quad \mathbf{E} \cdot \mathbf{E} = \mathbf{E}^T \mathbf{E}. \tag{3}$$

Ключевым моментом в доказательстве является удачное представление потоков на гранях элементов для схемы высокого порядка точности как интерполяции между осредненными значениями на соседних элементах; последнее было предложено в работе [17]. Такой подход позволяет свести задачу к нахождению нормы матрицы потоков для каждого элемента из сетки  $\Omega$  и ввести ограничения на потоки, полученные схемой высокого порядка точности.

Численный критерий устойчивости был получен с использованием оценки нормы  $L_2$  специально сконструированной матрицы  $G_i$  для каждого контрольного объема  $T_i$  из разбиения области  $\Omega = \bigcup_{i=0}^N T_i$ . Каждая такая матрица была создана с использованием скалярного произведения собственных векторов операторов

из противоположного разложения на гранях элемента контрольного объема  $T_i$ . Новый критерий получен численно с использованием геометрии элементов  $T_i$  и дает больший шаг по времени, чем результаты, опубликованные в предыдущих работах [3, 11]. Более того, на структурированной сетке критерий дает тот же самый результат, что и критерий фон Неймана [11]. Другими словами, новая оценка позволяет соединить условия устойчивости на структурированной и не структурированной сетках, что приводит к снижению вычислительных затрат при использовании нового критерия оценки шага по времени. Для обобщения численного метода оценки шага по времени на схемы высокого порядка точности по пространству приводится способ оценки максимального наклона градиентов в схеме MUSCL (Monotonic Upwind Scheme for Conservation Laws). Предложенный способ оценки наклона градиентов позволил сформулировать принцип максимума [13] для метода контрольных объемов применительно к векторным законам сохранения.

Итак, для схем высокого порядка точности запишем критерий устойчивости для векторных законов сохранения следующим образом. Для каждого из элементов  $T_i \in \Delta$  введем систему обозначений: через  $c_i = (\sqrt{\epsilon_i \mu_i})^{-1}$  обозначим локальную скорость света; через  $I_i = \{j_1, j_2, \dots, j_{m_i}\}$  — множество индексов элементов, имеющих общую грань с элементом  $T_i$ ; грани между соседними элементами  $T_i$  и  $T_j$ ,  $j \in I_i$ , обозначим через  $\partial T_{ij}$ , нормали к которым обозначим через  $\hat{\mathbf{n}}_{ij}$ , а площади граней — через  $S_{ij}$ . В результате запишем критерий устойчивости для схем высокого порядка точности по пространству и явной схемы Эйлера, имеющей первый порядок точности по времени, в следующем виде:

$$\Delta t^h \leq \min_i \frac{V_i}{c_i \delta_i \|G_i\|}.$$

Здесь

$$G_i = \left[ g_l^i g_k^i \hat{\mathbf{e}}_l^i \cdot \hat{\mathbf{e}}_k^i \right]_{l,k=1}^{2m_i} \in \mathbb{R}^{2m_i \times 2m_i}, \quad g_{2n-1}^i = g_{2n}^i = \sqrt{S_{ij_n}},$$

$$\hat{\mathbf{e}}_{2n-1}^i = \hat{\mathbf{e}}_{ij_n}^1 = \frac{1}{\sqrt{2}} \left( (\hat{\mathbf{n}}_{ij_n} \times \hat{\mathbf{b}}_{ij_n})^T \hat{\mathbf{b}}_{ij_n}^T \right)^T, \quad \hat{\mathbf{e}}_{2n}^i = \hat{\mathbf{e}}_{ij_n}^2 = \frac{1}{\sqrt{2}} \left( -\hat{\mathbf{b}}_{ij_n}^T (\hat{\mathbf{n}}_{ij_n} \times \hat{\mathbf{b}}_{ij_n})^T \right)^T;$$

$n = 1, \dots, m_i$ ;  $\hat{\mathbf{b}}_{ij_n}$  — нормализованный вектор, полученный из одного из ребер на грани  $\partial T_{ij_n}$ ; коэффициент  $\delta_i \geq 1$  зависит от ограничения на вектора-поток  $\mathbf{u}_{ij_n}$  размерности 6 между соседними элементами, записанными для схемы высокого порядка точности: для значений на элементах  $\mathbf{u}_i$  имеем

$$\frac{1 - \delta_i}{2} \leq \frac{\tilde{\mathbf{e}}_{ij_n}^s \cdot (\mathbf{u}_{ij_n} - \mathbf{u}_i)}{\tilde{\mathbf{e}}_{ij_n}^s \cdot (\mathbf{u}_{j_n} - \mathbf{u}_i)} \leq \frac{1}{2}, \quad s = 1, 2, \quad (4)$$

где  $\mathbf{u}_i$  — значения на элементах и

$$\hat{\mathbf{e}}_{ij_n}^1 = \frac{1}{\sqrt{2}} \left( \sqrt{\epsilon} (\hat{\mathbf{n}}_{ij_n} \times \hat{\mathbf{b}}_{ij_n})^T, \sqrt{\mu} \hat{\mathbf{b}}_{ij_n}^T \right)^T, \quad \tilde{\mathbf{e}}_{ij_n}^2 = \frac{1}{\sqrt{2}} \left( -\sqrt{\epsilon} \hat{\mathbf{b}}_{ij_n}^T, \sqrt{\mu} (\hat{\mathbf{n}}_{ij_n} \times \hat{\mathbf{b}}_{ij_n})^T \right)^T, \quad n = 1, \dots, m_i.$$

**1. Аппроксимация методом контрольных объемов для уравнений Максвелла.** Для того чтобы записать аппроксимацию уравнений Максвелла для метода контрольных объемов в векторной форме, перепишем векторное произведение в матричном виде:

$$\phi \times \mathbf{b} = C(\phi) \mathbf{b} = \begin{bmatrix} 0 & -\phi_3 & \phi_2 \\ \phi_3 & 0 & -\phi_1 \\ -\phi_2 & \phi_1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \phi_2 b_3 - \phi_3 b_2 \\ \phi_3 b_1 - \phi_1 b_3 \\ \phi_1 b_2 - \phi_2 b_1 \end{bmatrix}.$$

Это позволяет записать  $\text{rot } \mathbf{b} = \nabla \times \mathbf{b}$  в виде дивергенции матрицы:

$$\nabla \times \mathbf{b} = C(\nabla) \mathbf{b} = (\text{div } C(\mathbf{b}))^T. \quad (5)$$

Воспользовавшись соотношением  $C(\mathbf{b})^T \hat{\mathbf{n}} = -C(\mathbf{b}) \hat{\mathbf{n}} = -\mathbf{b} \times \hat{\mathbf{n}} = \hat{\mathbf{n}} \times \mathbf{b} = C(\hat{\mathbf{n}}) \mathbf{b}$  и применив формулу Грина к (5), получим

$$\int_{T_i} (\text{div } C(\mathbf{b}))^T dx = \int_{\partial T_i} C(\mathbf{b})^T \hat{\mathbf{n}} ds = \int_{\partial T_i} C(\hat{\mathbf{n}}) \mathbf{b} ds. \quad (6)$$

Тогда с использованием нового оператора (5) уравнения Максвелла могут быть записаны в виде

$$\begin{cases} \epsilon \partial_t \mathbf{E} - (\operatorname{div} C(\mathbf{H}))^\top + \sigma \mathbf{E} = -\mathbf{J}, \\ \mu \partial_t \mathbf{H} + (\operatorname{div} C(\mathbf{E}))^\top = \mathbf{0}. \end{cases} \quad (7)$$

Последнее позволяет записать уравнения (7) в компактной операторной форме:

$$\partial_t \mathbf{u} + \alpha^{-1} R \mathbf{u} = \alpha^{-1} (G + B \mathbf{u}), \quad (8)$$

$$\alpha = \begin{bmatrix} \epsilon I & \mathbf{0} \\ \mathbf{0} & \mu I \end{bmatrix}, \quad R \mathbf{u} = \begin{bmatrix} -\nabla \times \mathbf{H} \\ \nabla \times \mathbf{E} \end{bmatrix}, \quad B \mathbf{u} = \begin{bmatrix} \sigma I \\ \mathbf{0} \end{bmatrix}, \quad G = \begin{bmatrix} \mathbf{J} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{E} \\ \mathbf{H} \end{bmatrix}. \quad (9)$$

Проинтегрировав выражение (7) на элементе  $T_i$ , получим

$$\begin{cases} \int_{T_i} \epsilon \partial_t \mathbf{E} \, dx - \int_{T_i} (\operatorname{div} C(\mathbf{H}))^\top \, dx + \int_{T_i} \sigma \mathbf{E} \, dx = - \int_{T_i} \mathbf{J} \, dx, \\ \int_{T_i} \mu \partial_t \mathbf{H} \, dx + \int_{T_i} (\operatorname{div} C(\mathbf{E}))^\top \, dx = \mathbf{0}. \end{cases} \quad (10)$$

Для произвольной сеточной величины  $\phi(\mathbf{x})$  введем обозначение  $\phi_i = \frac{1}{V_i} \int_{T_i} \phi(\mathbf{x}) \, dV$  и, применив соотношение (6), перепишем систему (10) в виде

$$\begin{cases} \partial_t \mathbf{E}_i - \frac{1}{V_i} \int_{\partial T_i} (\epsilon)^{-1} C(\hat{\mathbf{n}}) \mathbf{H} \, ds + \frac{1}{V_i} \int_{T_i} (\epsilon)^{-1} \sigma \mathbf{E} \, dx = - \frac{1}{V_i} \int_{T_i} (\epsilon)^{-1} \mathbf{J} \, dx, \\ \partial_t \mathbf{H}_i + \frac{1}{V_i} \int_{\partial T_i} (\mu)^{-1} C(\hat{\mathbf{n}}) \mathbf{E} \, ds = \mathbf{0}. \end{cases}$$

Отсюда выражение (8) в интегральной форме примет вид

$$\partial_t \mathbf{u}_i + \frac{1}{V_i} \int_{\partial T_i} \alpha^{-1} R(\hat{\mathbf{n}}) \mathbf{u} \, ds = \frac{1}{V_i} \int_{\partial T_i} \alpha^{-1} (G + B \mathbf{u}) \, dx. \quad (11)$$

Здесь под матрицей  $R(\hat{\mathbf{n}})$  подразумевается симметрическая матрица  $R(\hat{\mathbf{n}}) = R(\hat{\mathbf{n}})^\top = \begin{bmatrix} 0 & -C(\hat{\mathbf{n}}) \\ C(\hat{\mathbf{n}}) & 0 \end{bmatrix}$ .

Введем обозначение  $\tilde{R}(\hat{\mathbf{n}}) = \alpha^{-1} R(\hat{\mathbf{n}})$ . Предположив однородность и изотропность пространства и постоянство  $\epsilon = \operatorname{diag} \{ \epsilon, \epsilon, \epsilon \}$  и  $\mu = \operatorname{diag} \{ \mu, \mu, \mu \}$  на всем элементе  $T_i \in \Delta$ , мы заключаем, что матрица  $\tilde{R}(\hat{\mathbf{n}})$  имеет шесть собственных значений, представленных в виде  $\Lambda = \operatorname{diag} \{ 0, 0, c, c, -c, -c \}$ , где  $c = (\sqrt{\epsilon \mu})^{-1}$  — скорость света [11]. Последнее вытекает из того, что матрица  $\tilde{R}(\hat{\mathbf{n}})$  является самосопряженной в пространстве, заданном скалярным произведением

$$(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \alpha \mathbf{v} = \sum_{i=1}^N u_i^\top \alpha_i v_i, \quad (12)$$

определяющим дискретный аналог энергетической нормы (3):

$$\|\mathbf{u}\| = \sqrt{(\mathbf{u}, \mathbf{u})} = \sqrt{\sum_{i=1}^N u_i^\top \alpha_i u_i},$$

где  $\alpha_i$  — значение  $\alpha$  на элементе  $T_i$ .

Для получения векторного аналога противоточной схемы разложим матрицу  $\tilde{R}(\hat{\mathbf{n}})$  на сумму матриц с положительными и отрицательными собственными значениями:  $\tilde{R}(\hat{\mathbf{n}}) = \tilde{R}(\hat{\mathbf{n}})^+ + \tilde{R}(\hat{\mathbf{n}})^-$ . Итак, в

пространстве, заданном скалярным произведением (12), мы можем записать  $\tilde{R}(\hat{\mathbf{n}}) = \tilde{Q}(\hat{\mathbf{n}})^{-1} \Lambda \tilde{Q}(\hat{\mathbf{n}})$ , где матрица  $\tilde{Q}(\hat{\mathbf{n}})$  состоит из единичных собственных векторов-строк  $(\tilde{\mathbf{e}}^k)^T$ ,  $k = 1, \dots, 6$ , матрицы  $\tilde{R}(\hat{\mathbf{n}})$ :

$$\tilde{Q}(\hat{\mathbf{n}}) = [\tilde{\mathbf{e}}^1(\hat{\mathbf{n}}), \dots, \tilde{\mathbf{e}}^6(\hat{\mathbf{n}})]^T = \begin{bmatrix} \sqrt{\epsilon} \hat{\mathbf{a}} & -\sqrt{\epsilon} \hat{\mathbf{b}} & \sqrt{\epsilon} \hat{\mathbf{a}} & \sqrt{\epsilon} \hat{\mathbf{b}} & \sqrt{\epsilon} \hat{\mathbf{n}} & 0 \\ \sqrt{\mu} \hat{\mathbf{b}} & \sqrt{\mu} \hat{\mathbf{a}} & -\sqrt{\mu} \hat{\mathbf{b}} & \sqrt{\mu} \hat{\mathbf{a}} & 0 & \sqrt{\mu} \hat{\mathbf{n}} \end{bmatrix}^T. \quad (13)$$

Здесь  $\hat{\mathbf{b}} \perp \hat{\mathbf{n}}$ ,  $\|\hat{\mathbf{n}}\| = \|\hat{\mathbf{b}}\| = 1$ ,  $\hat{\mathbf{a}} = \hat{\mathbf{n}} \times \hat{\mathbf{b}}$ . Следовательно, матрица  $\tilde{R}(\hat{\mathbf{n}})$  может быть представлена в виде

$$\tilde{R}(\hat{\mathbf{n}}) = \tilde{Q}(\hat{\mathbf{n}})^{-1} \Lambda \tilde{Q}(\hat{\mathbf{n}}) = \tilde{Q}(\hat{\mathbf{n}})^{-1} \Lambda^+ \tilde{Q}(\hat{\mathbf{n}}) + \tilde{Q}(\hat{\mathbf{n}})^{-1} \Lambda^- \tilde{Q}(\hat{\mathbf{n}}) = \tilde{R}(\hat{\mathbf{n}})^+ + \tilde{R}(\hat{\mathbf{n}})^-, \quad (14)$$

где  $\Lambda = \Lambda^+ + \Lambda^-$ ,  $\Lambda^+ = \text{diag}\{c, c, 0, 0, 0, 0\}$ ,  $\Lambda^- = \text{diag}\{0, 0, -c, -c, 0, 0\}$ ; матрица  $Q(\hat{\mathbf{n}})^{-1} = \boldsymbol{\alpha}^{-1} Q(\hat{\mathbf{n}})^T$  состоит из собственных векторов-столбцов матрицы  $\tilde{R}(\hat{\mathbf{n}})$ . Так же как и в работах [3, 8], элементы разложения (14) можно переписать в виде блочных матриц

$$\tilde{R}(\hat{\mathbf{n}})^+ = \frac{1}{2} \begin{bmatrix} -cC(\hat{\mathbf{n}})^2 & -\epsilon^{-1}C(\hat{\mathbf{n}}) \\ \mu^{-1}C(\hat{\mathbf{n}}) & -cC(\hat{\mathbf{n}})^2 \end{bmatrix}, \quad \tilde{R}(\hat{\mathbf{n}})^- = \frac{1}{2} \begin{bmatrix} cC(\hat{\mathbf{n}})^2 & -\epsilon^{-1}C(\hat{\mathbf{n}}) \\ \mu^{-1}C(\hat{\mathbf{n}}) & cC(\hat{\mathbf{n}})^2 \end{bmatrix},$$

для которых выполняются обычные свойства противоположных схем:

$$|\tilde{R}(\hat{\mathbf{n}})| = \tilde{R}(\hat{\mathbf{n}})^+ + \tilde{R}(-\hat{\mathbf{n}})^+, \quad \tilde{R}(\hat{\mathbf{n}})^- = -\tilde{R}(-\hat{\mathbf{n}})^+, \quad \tilde{R}(\hat{\mathbf{n}})^+ = \frac{1}{2} \left( \tilde{R}(\hat{\mathbf{n}}) + |\tilde{R}(\hat{\mathbf{n}})| \right).$$

Более того, с использованием внутренней структуры матрицы (13) мы можем записать

$$\tilde{R}(\hat{\mathbf{n}})^+ = c \sum_{j=1}^2 \boldsymbol{\alpha}^{-1} \tilde{\mathbf{e}}^j(\hat{\mathbf{n}}) [\tilde{\mathbf{e}}^j(\hat{\mathbf{n}})]^T.$$

В результате вся разница между методом контрольных объемов для скалярных и векторных величин заключается в том, что противоположное разложение записано не для скаляров, а для собственных значений матрицы  $\tilde{R}(\hat{\mathbf{n}})$ . Отметим важную связь между матрицами  $\tilde{R}(\hat{\mathbf{n}})^+$  и  $R(\hat{\mathbf{n}})^+$ , полученную в [11]:

$$\begin{aligned} \tilde{R}(\hat{\mathbf{n}})^- &= \frac{1}{2} \begin{bmatrix} cC^2(\hat{\mathbf{n}}) & -\epsilon^{-1}C(\hat{\mathbf{n}}) \\ \mu^{-1}C(\hat{\mathbf{n}}) & cC^2(\hat{\mathbf{n}}) \end{bmatrix} = c \begin{bmatrix} \mu & 0 \\ 0 & c^{-1} \end{bmatrix} \frac{1}{2} \begin{bmatrix} C^2(\hat{\mathbf{n}}) & -C(\hat{\mathbf{n}}) \\ C(\hat{\mathbf{n}}) & C^2(\hat{\mathbf{n}}) \end{bmatrix} \begin{bmatrix} \mu^{-1} & 0 \\ 0 & c \end{bmatrix} = \\ &= \begin{bmatrix} \epsilon^{-1/2} & 0 \\ 0 & \mu^{-1/2} \end{bmatrix} \frac{1}{2} \begin{bmatrix} C^2(\hat{\mathbf{n}}) & -C(\hat{\mathbf{n}}) \\ C(\hat{\mathbf{n}}) & C^2(\hat{\mathbf{n}}) \end{bmatrix} \begin{bmatrix} \mu^{-1/2} & 0 \\ 0 & \epsilon^{-1/2} \end{bmatrix} = (\sqrt{\alpha})^{-1} R^-(\hat{\mathbf{n}}) (\sqrt{\beta})^{-1}, \quad (\sqrt{\beta})^{-1} = c\sqrt{\alpha}. \end{aligned}$$

При расчете интеграла по грани  $\int_{\partial T_i} \boldsymbol{\alpha}^{-1} R(\hat{\mathbf{n}}) \mathbf{u} ds$  в выражении (11) мы применим разложение по потокам:

$$\int_{\partial T_i} \boldsymbol{\alpha}^{-1} R(\hat{\mathbf{n}}) \mathbf{u} ds = \int_{\partial T_i} \tilde{R}(\hat{\mathbf{n}})^+ \mathbf{u}^* + \tilde{R}(\hat{\mathbf{n}})^- \mathbf{u}^{**} ds, \quad \text{в котором для } \mathbf{x} \in \partial T_i \text{ мы получаем } \mathbf{u}^*(\mathbf{x}) = \lim_{\mathbf{y} \rightarrow \mathbf{x}} \mathbf{u}(\mathbf{y}) \text{ и}$$

$$\mathbf{u}^{**}(\mathbf{x}) = \lim_{\substack{\mathbf{y} \rightarrow \mathbf{x}, \\ \mathbf{y} \in \Omega \setminus T_i}} \mathbf{u}(\mathbf{y}).$$

В общем случае значения слева и справа от грани могут различаться, т.е. функция  $\mathbf{u}(\mathbf{x})$  может терпеть разрыв на грани. Поэтому значения  $\mathbf{u}^*$  и  $\mathbf{u}^{**}$  должны вычисляться путем интерполяции слева и справа. Тогда мы можем обозначить поток через плоскую грань  $\partial T_{ij}$  из элемента  $T_i$  в элемент  $T_j$  следующим образом:

$$\mathbf{u}_{ij} = \frac{1}{S_{ij}} \int_{\partial T_{ij}} \mathbf{u}^*(\mathbf{y}) dS, \quad (15)$$

при этом аппроксимация первого порядка точности получается из предположения, что  $\mathbf{u}_{ij} = \mathbf{u}_i$  и  $\mathbf{u}_{ji} = \mathbf{u}_j$  [11]. В случае же применения схем высокого порядка точности значение  $\mathbf{u}_{ij}$  зависит от  $\mathbf{u}_k$  на соседних элементах. Для случая внешней грани мы предполагаем, что  $\mathbf{u}_{ji} = 0$ ; как и в работе [11], предполагаем,

что на грани помещен идеальный металлический проводник:  $\mathbf{u}_{ji} = (-\mathbf{E}_{ij}^T \mathbf{H}_{ij}^T)^T$ . Для упрощения записи полученных схем введем обозначения  $\tilde{R}_{ij}^+ = \tilde{R}(\hat{\mathbf{n}}_{ij})^+$  и  $\tilde{R}_{ij}^- = \tilde{R}(\hat{\mathbf{n}}_{ij})^-$ , где  $\hat{\mathbf{n}}_{ij}$  — нормаль к грани  $\partial T_{ij}$ .

С использованием введенных обозначений запишем аппроксимацию на основе метода контрольных объемов первого порядка точности для уравнений Максвелла с использованием явной схемы Эйлера для интегрирования по времени в виде [3, 8, 11]

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t}{V_i} \sum_{j \in I_i} S_{ij} \boldsymbol{\alpha}_i^{-1} \left( T_{ij} \boldsymbol{\alpha}_i \tilde{R}_{ij}^+ \mathbf{u}_{ij}^n - T_{ji} \boldsymbol{\alpha}_j \tilde{R}_{ji}^+ \mathbf{u}_{ji}^n \right), \tag{16}$$

где  $I_i$  — множество элементов, имеющих общую грань с элементом  $T_i$ ,  $S_{ij}$  — площадь грани между соседними элементами  $i$  и  $j$ . Оператор согласования на гранях с разными диэлектрическими свойствами запишем в форме

$$T_{ij} = \begin{bmatrix} \frac{2Y_j}{Y_i + Y_j} I & \mathbf{0} \\ \mathbf{0} & \frac{2Y_j^{-1}}{Y_i^{-1} + Y_j^{-1}} I \end{bmatrix}, \quad Y_i = \sqrt{\frac{\epsilon_i}{\mu_i}}, \quad Y_j = \sqrt{\frac{\epsilon_j}{\mu_j}}.$$

Матрица диэлектрической проницаемости из (9) имеет вид

$$\boldsymbol{\alpha}_i = \begin{bmatrix} \epsilon_i I & \mathbf{0} \\ \mathbf{0} & \mu_i I \end{bmatrix}.$$

Далее рассмотрим явную схему Эйлера для уравнений Максвелла, записанных в консервативной форме:

$$\mathbf{u}^{n+1} = \mathbf{u}^n - \Delta t L \mathbf{u}^n, \tag{17}$$

где  $L$  — оператор дискретизации пространственных производных.

**2. Локальная интерполяция MUSCL второго порядка точности.** Воспользовавшись формулой Тейлора второго порядка точности

$$u(\mathbf{x}) = u(\mathbf{x}_i) + (\mathbf{x} - \mathbf{x}_i) \cdot \nabla u(\mathbf{x}_i) + O(|\mathbf{x} - \mathbf{x}_i|^2), \tag{18}$$

мы можем интерполировать решение в любую точку внутри контрольного объема, для чего необходимо построить аппроксимацию градиента с порядком точности не меньше первого.

Для применения формулы (18) рассмотрим аппроксимацию градиента скалярной величины  $u(\mathbf{x})$  на элементах  $T_i$  неструктурированной сетки. Для этого найдем значение оператора градиента на элементе  $T_i$ , применив формулу Грина

$$\nabla u(\mathbf{x}_i) \approx \frac{1}{V_i} \int_{T_i} \nabla u(\mathbf{x}) dV = \frac{1}{V_i} \int_{\partial T_i} \hat{\mathbf{n}}(\mathbf{x}) u(\mathbf{x}) dS \approx \nabla_0 u_i = \frac{1}{V_i} \sum_{j \in I_i} S_{ij} \hat{\mathbf{n}}_{ij} [u_i + \beta_{ij}(u_j - u_i)],$$

где  $\hat{\mathbf{n}}_{ij}$  — нормаль к грани  $\partial T_{ij}$ ,  $\beta_{ij}$  — коэффициент линейной интерполяции в точку на грани с использованием значений  $u_i$  и  $u_j$ . Коэффициент линейной интерполяции  $\beta_{ij}$  на точку грани может быть легко получен по формуле

$$\beta_{ij} = \frac{|(\mathbf{x}_i - \mathbf{x}_{ij}) \cdot \hat{\mathbf{n}}_{ij}|}{|(\mathbf{x}_j - \mathbf{x}_{ij}) \cdot \hat{\mathbf{n}}_{ij}| + |(\mathbf{x}_i - \mathbf{x}_{ij}) \cdot \hat{\mathbf{n}}_{ij}|},$$

где  $\mathbf{x}_{ij}$  — центр грани  $\partial T_{ij}$ . Заметим, что для линейной интерполяции между двумя точками мы получаем центральную аппроксимацию  $\nabla_0 u_i$  со вторым порядком точности. Вместе с тем, исключая из аппроксимации один из соседних элементов  $k$  (т.е. задавая  $\beta_{ik} = 0$ ), мы можем получить направленные аппроксимации первого порядка точности  $\nabla_k u_i$ .

В итоге мы получим несколько градиентов для одного элемента. Для получения аппроксимации (18) второго порядка точности достаточно использовать только лишь центральную аппроксимацию градиента  $\nabla_0 u_i$ , что является достаточным для аппроксимации уравнений Максвелла, решение которых обычно не содержит разрывов. Тем не менее, для расчета практических задач, с целью надежного исключения колебаний решения вокруг областей с большими градиентами решения мы оставим из всех градиентов

$\nabla_k u_i$ ,  $k \in I_i$ , именно тот, который обладает наименьшей нормой, т.е. позволяющий соединять наиболее гладко решение между соседними элементами:

$$\nabla u_i = \nabla_{k_{\min}} u_i, \quad k_{\min} \in I_i, \quad \|\nabla_{k_{\min}} u_i\| \leq \|\nabla_k u_i\|, \quad \forall k \in I_i, \quad (19)$$

что позволяет существенно снизить влияние разрывов на интерполяцию и как итог снизить осцилляции численного решения.

С другой стороны, в работе [16] было предложено использовать взвешенные комбинации аппроксимации градиентов  $\nabla_k u_i$ ,  $k \in I_i$ , тем самым позволяя получить наиболее оптимальную аппроксимацию  $\nabla u(\mathbf{x}_i)$ , и учитывать ограничения на центральный градиент  $\nabla_0 u_i$ . Другими словами, для расчета аппроксимации  $\nabla u(\mathbf{x}_i)$  в работе [16] был предложен наиболее простой и быстрый способ исключения осцилляций решения, основанный на вычислении ограничения оператора градиента  $\nabla_0 u_i$ :

$$\nabla u_i = \min_{k \in I_i} \alpha_k \nabla_0 u_i, \quad \alpha_k = \begin{cases} \frac{\max(u_k - u_i, 0)}{(\mathbf{x}_k - \mathbf{x}_i) \cdot \nabla_0 u_i}, & (\mathbf{x}_k - \mathbf{x}_i) \cdot \nabla_0 u_i > \max(u_k - u_i, 0), \\ \frac{\min(u_k - u_i, 0)}{(\mathbf{x}_k - \mathbf{x}_i) \cdot \nabla_0 u_i}, & (\mathbf{x}_k - \mathbf{x}_i) \cdot \nabla_0 u_i < \min(u_k - u_i, 0), \\ 1, & \text{во всех других случаях.} \end{cases} \quad (20)$$

Смысл этого способа заключается в том, что мы ограничиваем норму градиента так, чтобы интерполяционная формула (18) при интерполяции в центр соседнего элемента не давала значений, превосходящих по модулю осредненные величины на соседних элементах. В результате мы можем записать аппроксимацию компонент вектора потока или скалярных величин через грань (15) в виде

$$u_{ij} = u_i + (\mathbf{x}_{ij} - \mathbf{x}_i) \cdot \nabla u_i, \quad (21)$$

где  $\mathbf{x}_{ij}$  — центр грани  $\partial T_{ij}$ ,  $\nabla u_i$  — подходящая аппроксимация градиента от скалярной величины  $u$  (компоненты векторной величины  $\mathbf{u}$ ) на элементе  $T_i$ .

В случае скалярных законов сохранения в методе контрольных объемов для получения того же самого критерия устойчивости для схем высокого порядка, что и для схем первого порядка точности, обычно применяется принцип максимума [13]

$$\min(u_i, u_j) \leq u_{ij} \leq \max(u_i, u_j), \quad (22)$$

закрывающийся в требовании ограничения на величины потоков  $u_{ij}$  между соседними элементами  $T_i$  и  $T_j$ . Легко заметить, что принцип максимума для аппроксимации (21) выполняется только при использовании аппроксимации градиента (20). Для его выполнения и в случае аппроксимации градиента (19) можно ввести аналогичное ограничение, как и для  $\nabla_0 u_i$  в (20), либо просто ограничить потоки на элементах с помощью формулы

$$u_{ij} = \min\left(\max(u_{ij}, \min(u_i, u_j)), \max(u_i, u_j)\right), \quad (23)$$

что позволяет применять принцип максимума для всех видов аппроксимаций градиента. Поэтому в настоящей работе для векторных законов сохранения нам требуется изучить влияние ограничений на потоки и сформулировать векторный аналог принципа максимума.

**3. Шаг по времени для метода контрольных объемов для уравнений Максвелла.** Задача получения критерия устойчивости может быть сформулирована как получение максимально возможного значения  $\Delta t$ , при котором схема (17) будет устойчива. Для этого сформулируем определение устойчивости схемы в терминах энергетической нормы [3].

**Определение.** Говорят, что численная схема  $L_2$ -устойчива, если при интегрировании по времени в области без внешних источников энергия не нарастает, т.е. выполняется условие

$$\|\mathbf{u}^{n+1}\| \leq \|\mathbf{u}^n\|. \quad (24)$$

Тогда, записав (24) через скалярное произведение (12), получим  $(\mathbf{u}^{n+1}, \mathbf{u}^{n+1}) \leq (\mathbf{u}^n, \mathbf{u}^n)$ ; затем, воспользовавшись выражением (17), запишем

$$\Delta t(L\mathbf{u}^n, L\mathbf{u}^n) \leq 2(L\mathbf{u}^n, \mathbf{u}^n). \quad (25)$$

**3.1. Метод контрольного объема первого порядка точности для уравнений Максвелла.** Рассмотрим доказательство критерия устойчивости для схемы первого порядка точности, обозначив через  $L_1$  аппроксимацию первого порядка точности, для которой  $\mathbf{u}_{ij} = \mathbf{u}_i$ . Для упрощения доказательства

предположим, что электрофизические свойства на всех элементах одни и те же ( $\epsilon_i = \epsilon$  и  $\mu_i = \mu$ ). Для случая элементов с разными электрофизическими свойствами доказательство может быть изменено так же, как и в работе [3]. Итак, для элементов с одинаковыми электрофизическими свойствами ( $T_{ij} = T_{ji} = I$ ) для схемы первого порядка точности запишем

$$(L_1 \mathbf{u}, \mathbf{u}) = \frac{1}{2} \sum_{i=1}^N \mathbf{u}_i^T \boldsymbol{\alpha}_i \left[ \sum_{k \in I_i} S_{ik} \left( \tilde{R}_{ik}^+ \mathbf{u}_i^n + \tilde{R}_{ik}^- \mathbf{u}_k^n \right) \right], \quad (26)$$

$$(L_1 \mathbf{u}, L_1 \mathbf{u}) = \frac{1}{2} \sum_{i=1}^N \frac{1}{V_i} \left\{ \left[ \sum_{j \in I_i} S_{ij} \left( \tilde{R}_{ij}^+ \mathbf{u}_i^n + \tilde{R}_{ij}^- \mathbf{u}_j^n \right) \right]^T \boldsymbol{\alpha}_i \left[ \sum_{k \in I_i} S_{ik} \left( \tilde{R}_{ik}^+ \mathbf{u}_i^n + \tilde{R}_{ik}^- \mathbf{u}_k^n \right) \right] \right\}, \quad (27)$$

где  $L_1$  — аппроксимация пространственных производных первого порядка точности. Воспользовавшись тем фактом, что  $\sum_{j \in I_i} S_{ij} \left( \tilde{R}_{ij}^+ \mathbf{u}_i^n + \tilde{R}_{ij}^- \mathbf{u}_j^n \right) = 0$ , мы можем записать

$$\sum_{j \in I_i} S_{ij} \tilde{R}_{ij}^+ \mathbf{u}_i^n = - \sum_{j \in I_i} S_{ij} \tilde{R}_{ij}^- \mathbf{u}_j^n. \quad (28)$$

Применив выражение (28) к (27) и (26), получаем

$$(L_1 \mathbf{u}, \mathbf{u}) = \frac{1}{2} \sum_{i=1}^N \mathbf{u}_i^T \boldsymbol{\alpha}_i \sum_{k \in I_i} S_{ik} \tilde{R}_{ik}^- (\mathbf{u}_k^n - \mathbf{u}_i^n),$$

$$(L_1 \mathbf{u}, L_1 \mathbf{u}) = \frac{1}{2} \sum_{i=1}^N \frac{1}{V_i} \left\{ \left[ \sum_{j \in I_i} S_{ij} \tilde{R}_{ij}^- (\mathbf{u}_j^n - \mathbf{u}_i^n) \right]^T \boldsymbol{\alpha}_i \left[ \sum_{k \in I_i} S_{ik} \tilde{R}_{ik}^- (\mathbf{u}_k^n - \mathbf{u}_i^n) \right] \right\}.$$

Для упрощения выражения  $(L_1 \mathbf{u}, \mathbf{u})$  воспользуемся тем, что

$$\tilde{R}_{ik}^- = (\tilde{R}_{ik}^-)^T, \quad B \mathbf{a} \cdot \mathbf{b} = \frac{1}{2} B \mathbf{b} \cdot \mathbf{b} + \frac{1}{2} B \mathbf{a} \cdot \mathbf{a} - \frac{1}{2} B (\mathbf{b} - \mathbf{a}) \cdot (\mathbf{b} - \mathbf{a}),$$

и получим

$$\sum_{k \in I_i} S_{ik} \left[ \mathbf{u}_i \cdot \tilde{R}_{ik}^- \mathbf{u}_k - \mathbf{u}_i \cdot \tilde{R}_{ik}^- \mathbf{u}_i \right] = \frac{1}{2} \sum_{k \in I_i} S_{ik} \left[ \mathbf{u}_k \cdot \tilde{R}_{ik}^- \mathbf{u}_k - \mathbf{u}_i \cdot \tilde{R}_{ik}^- \mathbf{u}_i - (\mathbf{u}_k - \mathbf{u}_i) \cdot \tilde{R}_{ik}^- (\mathbf{u}_k - \mathbf{u}_i) \right].$$

**3.2. Выражения для максимальной энергии.** Подставив выражения для  $(L_1 \mathbf{u}, L_1 \mathbf{u})$  и  $(L_1 \mathbf{u}, \mathbf{u})$  в формулу (25), получим

$$\sum_{i=1}^N \frac{\Delta t}{V_i} \left\{ \left[ \sum_{j \in I_i} S_{ij} \tilde{R}_{ij}^- (\mathbf{u}_j^n - \mathbf{u}_i^n) \right]^T \boldsymbol{\alpha}_i \left[ \sum_{k \in I_i} S_{ik} \tilde{R}_{ik}^- (\mathbf{u}_k^n - \mathbf{u}_i^n) \right] \right\} \leq$$

$$\leq \sum_{i=1}^N \sum_{k \in I_i} S_{ik} \left[ \mathbf{u}_k \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- \mathbf{u}_k - \mathbf{u}_i \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- \mathbf{u}_i - (\mathbf{u}_k - \mathbf{u}_i) \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- (\mathbf{u}_k - \mathbf{u}_i) \right]. \quad (29)$$

В последней формуле легко проверить, что

$$\sum_{i=1}^N \sum_{k \in I_i} S_{ik} \left[ \mathbf{u}_k \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- \mathbf{u}_k - \mathbf{u}_i \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- \mathbf{u}_i \right] = - \sum_{i=1}^{N_b} \sum_{k \in I_i^s} S_{ik} \mathbf{u}_i \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- \mathbf{u}_i \geq 0, \quad (30)$$

где  $N_b$  — номер элементов с гранями на границе расчетной области и  $I_i^s \subset I_i$  элементы размерности  $\dim - 1$ , лежащие на границе расчетной области. В том случае, если грань с индексом  $k \in I_i^s$  идеальный проводник, мы можем записать части суммы в левой части формулы (30) в виде

$$\mathbf{u}_j \cdot \boldsymbol{\alpha}_i \tilde{R}_{ij}^- \mathbf{u}_j - \mathbf{u}_i \cdot \boldsymbol{\alpha}_i \tilde{R}_{ij}^- \mathbf{u}_i = \begin{bmatrix} -\mathbf{E}_i \\ \mathbf{H}_i \end{bmatrix}^T \boldsymbol{\alpha}_i \frac{1}{2} \begin{bmatrix} -c_i C(\hat{\mathbf{n}}_{ij})^2 & -\epsilon_i^{-1} C(\hat{\mathbf{n}}_{ij}) \\ \mu_i^{-1} C(\hat{\mathbf{n}}_{ij}) & -c_i C(\hat{\mathbf{n}}_{ij})^2 \end{bmatrix} \begin{bmatrix} -\mathbf{E}_i \\ \mathbf{H}_i \end{bmatrix} \begin{bmatrix} \mathbf{E}_i \\ \mathbf{H}_i \end{bmatrix}^T -$$

$$- \boldsymbol{\alpha}_i \frac{1}{2} \begin{bmatrix} -c_i C(\hat{\mathbf{n}}_{ij})^2 & -\epsilon_i^{-1} C(\hat{\mathbf{n}}_{ij}) \\ \mu_i^{-1} C(\hat{\mathbf{n}}_{ij}) & -c_i C(\hat{\mathbf{n}}_{ij})^2 \end{bmatrix} \begin{bmatrix} \mathbf{E}_i \\ \mathbf{H}_i \end{bmatrix} = 0. \quad (31)$$

В выражении (30) грани на совершенных проводниках не оказывают никакого влияния; следовательно, используя выражения (30) и (31), мы можем переписать (29) в виде

$$\begin{aligned} \sum_{i=1}^N \frac{\Delta t}{V_i} \left\{ \left[ \sum_{j \in I_i} S_{ij} \tilde{R}_{ij}^- (\mathbf{u}_j^n - \mathbf{u}_i^n) \right]^T \boldsymbol{\alpha}_i \left[ \sum_{k \in I_i} S_{ik} \tilde{R}_{ik}^- (\mathbf{u}_k^n - \mathbf{u}_i^n) \right] \right\} &\leq \\ &\leq \sum_{i=1}^N \sum_{k \in I_i} S_{ik} \left[ -(\mathbf{u}_k - \mathbf{u}_i) \cdot \boldsymbol{\alpha}_i \tilde{R}_{ik}^- (\mathbf{u}_k - \mathbf{u}_i) \right] \leq 2(L_1 \mathbf{u}, \mathbf{u}). \end{aligned} \quad (32)$$

Пусть  $\sqrt{\boldsymbol{\beta}} = c^{-1} \sqrt{\boldsymbol{\alpha}}$ ; тогда

$$\begin{aligned} \left( \tilde{R}_{ij}^- \right)^T \boldsymbol{\alpha}_i \tilde{R}_{ik}^- &= \sqrt{\boldsymbol{\beta}^{-1}} R_{ij}^- \sqrt{\boldsymbol{\alpha}^{-1}} \boldsymbol{\alpha}_i \sqrt{\boldsymbol{\alpha}^{-1}} R_{ik}^- \sqrt{\boldsymbol{\beta}^{-1}} = \sqrt{\boldsymbol{\beta}^{-1}} R_{ij}^- R_{ik}^- \sqrt{\boldsymbol{\beta}^{-1}}, \\ \boldsymbol{\alpha}_i \tilde{R}_{ik}^- &= \boldsymbol{\alpha}_i \sqrt{\boldsymbol{\alpha}^{-1}} R_{ik}^- \sqrt{\boldsymbol{\beta}^{-1}} = \sqrt{\boldsymbol{\alpha}_i} R_{ik}^- \sqrt{\boldsymbol{\beta}^{-1}} = c_i \sqrt{\boldsymbol{\beta}^{-1}} R_{ik}^- \sqrt{\boldsymbol{\beta}^{-1}}. \end{aligned}$$

Введя обозначение  $\mathbf{w}_{ik} = \sqrt{\boldsymbol{\beta}^{-1}} (\mathbf{u}_k - \mathbf{u}_i)$ , перепишем выражение (32) в виде

$$\sum_{i=1}^N \frac{\Delta t}{V_i} \left\{ \left[ \sum_{j \in I_i} S_{ij} R_{ij}^- \mathbf{w}_{ij} \right] \cdot \left[ \sum_{k \in I_i} S_{ik} R_{ik}^- \mathbf{w}_{ik} \right] \right\} \leq \sum_{i=1}^N \sum_{k \in I_i} S_{ik} \left[ -\mathbf{w}_{ik}^T R_{ik}^- \mathbf{w}_{ik} \right] \leq 2(L_1 \mathbf{u}, \mathbf{u}). \quad (33)$$

Для упрощения оценки (33) потребуем выполнения следующего локального неравенства для каждого элемента  $i$ :

$$\frac{\Delta t}{V_i} \left[ \sum_{j \in I_i} S_{ij} R_{ij}^- \mathbf{w}_{ij} \right] \cdot \left[ \sum_{k \in I_i} S_{ik} R_{ik}^- \mathbf{w}_{ik} \right] \leq \sum_{k \in I_i} S_{ik} \left[ -W_{ik}^T R_{ik}^- \mathbf{w}_{ik} \right]. \quad (34)$$

Обозначим через  $W_i = \{I, \dots, I\}$  матрицу размером  $6 \times 6m_i$ , где  $m_i$  — число граней элемента  $T_i$ , а  $I$  — единичная матрица шестого порядка. Введя блочно-диагональную матрицу  $Z_i = -\text{diag} \left\{ S_{ik} R_{ik}^- (\hat{\mathbf{n}}_{ik}) \right\}_{k \in I_i}$  и вектор-столбец  $\mathbf{x}_i = \text{vector} \{ \mathbf{w}_{ik} \}_{k \in I_i}$ , мы перепишем выражение (34) в виде

$$c \frac{\Delta t}{V_i} (W_i Z_i \mathbf{x}_i, W_i Z_i \mathbf{x}_i) \leq (Z_i \mathbf{x}_i, \mathbf{x}_i) \quad \forall i. \quad (35)$$

**3.3. Оптимальный способ вычисления критерия устойчивости.** Рассмотрим эффективный способ вычисления шага по времени в формуле (35). Воспользуемся свойством  $R_{ij}^- = (R_{ij}^-)^T$ . Тогда получим, что матрица  $Z_i$  является симметрической и положительно определенной. Это позволяет записать  $Z_i = Q_i \Lambda_i Q_i^T = Q_i \sqrt{\Lambda_i} Q_i^T Q_i \sqrt{\Lambda_i} Q_i^T = \sqrt{Z_i} \sqrt{Z_i}$ , где  $\Lambda_i = \text{diag} \{ \lambda_j \}_{j=1}^{6m_i}$  — диагональная матрица собственных значений  $Z_i$  и  $\sqrt{\Lambda_i} = \text{diag} \{ \sqrt{\lambda_j} \}_{j=1}^{6m_i}$ .

Введя новое обозначение  $\mathbf{y} = \sqrt{Z_i} \mathbf{x}_i$  и переписав неравенство (35) в форме  $\frac{V_i}{c \Delta t} \geq \frac{(W_i \sqrt{Z_i} \mathbf{y}, W_i \sqrt{Z_i} \mathbf{y})}{(\mathbf{y}, \mathbf{y})}$  для всех  $i$ , получим критерий устойчивости в виде сумм по граням:

$$\frac{V_i}{c \Delta t} \geq \frac{\left[ \sum_{k \in I_i} \sqrt{S_{ik}} \sqrt{-R_{ik}^-} \mathbf{y}_{ik} \right] \cdot \left[ \sum_{j \in I_i} \sqrt{S_{ij}} \sqrt{-R_{ij}^-} \mathbf{y}_{ij} \right]}{\sum_{k \in I_i} \mathbf{y}_{ik} \cdot \mathbf{y}_{ik}} \quad \forall i. \quad (36)$$

Из предыдущего уравнения, используя численные методы, мы можем вычислить максимальное  $\Delta t$ , при котором схема является устойчивой. Однако эти вычисления могут быть упрощены, если мы воспользуемся свойством, что  $C^3(\mathbf{n}_{ij}) = -C(\mathbf{n}_{ij})$ , которое влечет за собой  $[-R_{ij}^-]^2 = -R_{ij}^-$ ; следовательно,  $\sqrt{-R_{ij}^-} = -R_{ij}^-$ , а это означает, что все собственные значения матрицы  $-R_{ij}^-$  есть единица с кратностью два и ноль с кратностью четыре. Поэтому, по аналогии с (13), выразим собственные векторы для собственного значения, равного единице, в виде

$$\hat{\mathbf{e}}_{ij}^1 = \frac{1}{\sqrt{2}} \left( (\hat{\mathbf{n}}_{ij} \times \hat{\mathbf{b}}_{ij})^T \hat{\mathbf{b}}_{ij} \right)^T, \quad \hat{\mathbf{e}}_{ij}^2 = \frac{1}{\sqrt{2}} \left( -\hat{\mathbf{b}}_{ij}^T (\hat{\mathbf{n}}_{ij} \times \hat{\mathbf{b}}_{ij})^T \right)^T.$$

Четыре оставшихся собственных вектора матрицы  $-R_{ij}^-$ , соответствующие нулевому собственному значению (или взятые из ядра), могут быть записаны в виде  $\hat{e}_{ij}^3 = \frac{1}{\sqrt{2}} \left( -(\hat{n}_{ij} \times \hat{b}_{ij})^T \hat{b}_{ij} \right)^T$ ,  $\hat{e}_{ik}^4 = \frac{1}{\sqrt{2}} \left( \hat{b}_{ij}^T (\hat{n}_{ij} \times \hat{b}_{ij})^T \right)^T$ ,  $\hat{e}_{ij}^5 = \frac{1}{\sqrt{2}} \left( \hat{n}_{ij}^T \mathbf{0}^T \right)^T$ ,  $\hat{e}_{ik}^5 = \frac{1}{\sqrt{2}} \left( \mathbf{0}^T \hat{n}_{ij}^T \right)^T$ . Следовательно, взяв один вектор  $\hat{b}_{ij}$  на поверхности грани между элементами  $i$  и  $j$ , мы можем задать все собственные векторы оператора  $\sqrt{-R_{ij}^-} = -R_{ij}^-$ . С целью наиболее экономичного вычисления (36) мы можем разложить каждый вектор  $\mathbf{y}_{ij}$  на сумму собственных векторов  $\hat{e}_{ij}^1$  и  $\hat{e}_{ij}^2$ , соответствующих единичному собственному значению матрицы  $\sqrt{-R_{ij}^-} = -R_{ij}^-$ :

$$\mathbf{y}_{ij} = \sum_{s=1}^2 w_{ij}^s \hat{e}_{ij}^s.$$

Подставив это выражение в формулу (36), получим следующее выражение для шага по времени:

$$\frac{V_i}{c\Delta t} \geq \frac{\sum_{k \in I_i} \sum_{j \in I_i} \sqrt{S_{ik} S_{ij}} \sum_{s_1=1}^2 \sum_{s_2=1}^2 w_{ik}^{s_1} w_{ij}^{s_2} \hat{e}_{ik}^{s_1} \cdot \hat{e}_{ij}^{s_2}}{\sum_{k \in I_i} \sum_{s=1}^2 (w_{ik}^s)^2},$$

которое перепишем в виде  $\frac{V_i}{c\Delta t} \geq \frac{\mathbf{w} \cdot G_i \mathbf{w}}{\mathbf{w} \cdot \mathbf{w}}$ , где элементы матрицы  $G_i \in \mathbb{R}^{2m_i \times 2m_i}$  являются скалярными произведениями:  $G_i = \left[ g_l^i g_k^i \hat{e}_l^i \cdot \hat{e}_k^i \right]_{l,k=1}^{2m_i}$ ,  $g_{2n-1}^i = g_{2n}^i = \sqrt{S_{ij_n}}$ ,  $\hat{e}_{2n-1}^i = \hat{e}_{ij_n}^1$ ,  $\hat{e}_{2n}^i = \hat{e}_{ij_n}^2$ ,  $n = 1, \dots, m_i$ . Тогда критерий устойчивости для схем первого порядка точности записывается в виде

$$\Delta t^1 \leq \min_i \frac{V_i}{c \|G_i\|},$$

или так же, как и в работе [3], для элементов с индивидуальными  $\epsilon_i$  и  $\mu_i$  в виде

$$\Delta t^1 \leq \min_i \frac{V_i}{c_i \|G_i\|},$$

где  $c_i = \sqrt{(\epsilon_i \mu_i)^{-1}}$ . Норма матрицы  $G_i$  вычисляется достаточно быстро, поскольку  $4 \leq m_i = |I_i| \leq 6$  для сетки, состоящей из тетраэдров, призм, пирамид и шестигранников.

**3.4. Устойчивость пространственной аппроксимации уравнений Максвелла высокого порядка точности.** Для обобщения критерия устойчивости со схем первого порядка точности на схемы высокого порядка точности, такие как схемы MUSCL и полиномиальные интерполяции высокого порядка точности [14], рассмотрим по аналогии с (16) следующую схему высокого порядка точности по пространству:

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t^h}{V_i} \sum_{j \in I_i} S_{ij} \alpha_i^{-1} \left( \mathbf{T}_{ij} \alpha_i \tilde{R}_{ij}^+ \mathbf{u}_{ij}^n - \mathbf{T}_{ji} \alpha_j \tilde{R}_{ji}^+ \mathbf{u}_{ji}^n \right). \tag{37}$$

Здесь  $\mathbf{u}_{ij}^n$  — поток через грань из элемента  $i$  в элемент  $j$ , вычисленный с использованием схем высокого порядка точности, примененных покомпонентно к каждой компоненте  $[u_i^n]_k$ ,  $k = 1, \dots, 6$ , вектора  $\mathbf{u}_i$ ,  $i = 1, \dots, N$ .

Для упрощения анализа рассмотрим схему MUSCL, в которой выполняется требование минимальности градиента (19). Тогда, по аналогии с работой [17], потоки между соседними элементами линейно выразимы из осредненных величин на соседних элементах:

$$[u_{ij}^n]_k = [u_i^n]_k \left( 1 - [d_{ij}^n]_k \right) + [u_j^n]_k [d_{ij}^n]_k = [u_i^n]_k + [d_{ij}^n]_k \left( [u_j^n]_k - [u_i^n]_k \right), \quad k = 1, \dots, 6.$$

Последнее выражение в векторной форме имеет вид

$$\mathbf{u}_{ij}^n = (1 - D_{ij}^n) \mathbf{u}_i^n + D_{ij}^n \mathbf{u}_j^n = \mathbf{u}_i^n + D_{ij}^n (\mathbf{u}_j^n - \mathbf{u}_i^n), \tag{38}$$

где  $D_{ij}^n = \text{diag}(\mathbf{d}_{ij}^n)$  — диагональная матрица, полученная из вектора  $\mathbf{d}_{ij}^n$ , состоящего из компонент  $[d_{ij}^n]_k$ ,  $k = 1, \dots, 6$ : в случае  $[u_i^n]_k = [u_j^n]_k$  мы предполагаем, что  $[d_{ij}^n]_k = 0$ .

Поскольку поток  $\mathbf{u}_{ij}^n$  умножается на матрицу  $\tilde{R}_{ij}^+$ , мы можем заменить этот поток через грань и интерполяцию (38) на  $\tilde{R}_{ij}^+$  заменить на эквивалентную  $\tilde{R}_{ij}^+ \tilde{\mathbf{u}}_{ij}^n = \tilde{R}_{ij}^+ \mathbf{u}_{ij}^n$ :

$$\tilde{\mathbf{u}}_{ij}^n = \mathbf{u}_i^n + \tilde{\gamma}_{ij}^n (\mathbf{u}_j^n - \mathbf{u}_i^n). \quad (39)$$

Здесь  $\tilde{\gamma}_{ij}^n$  — интерполяционная матрица, записанная в виде

$$\tilde{\gamma}_{ij}^n = \tilde{Q}_{ij}^{-1} \text{diag} \left( [d_{ij}^n]_k : k = 1 : m_{ij}^+ \right) \tilde{Q}_{ij} = \sum_{k=1}^{m_{ij}^+} \alpha^{-1} \tilde{\mathbf{e}}_{ij}^k [d_{ij}^n]_k \left( \tilde{\mathbf{e}}_{ij}^k \right)^T, \quad (40)$$

где  $m_{ij}^+$  — количество положительных собственных значений матрицы  $\tilde{R}_{ij}^+$ , а

$$[d_{ij}^n]_k = \frac{\tilde{\mathbf{e}}_{ij}^k \cdot D_{ij}^n (\mathbf{u}_j^n - \mathbf{u}_i^n)}{\tilde{\mathbf{e}}_{ij}^k \cdot (\mathbf{u}_j^n - \mathbf{u}_i^n)}, \quad k = 1, m_{ij}^+, \quad (41)$$

являются коэффициентами интерполяции между проекциями  $\tilde{\mathbf{e}}_{ij}^k \cdot \mathbf{u}_i^n$  и  $\tilde{\mathbf{e}}_{ij}^k \cdot \mathbf{u}_j^n$  на векторы, соответствующие положительным собственным значениям  $\tilde{R}_{ij}^+$ . При этом проекция на собственные векторы из ядра  $\tilde{R}_{ij}^+$  предполагается равной нулю. Другими словами, мы рассматриваем потоки  $\tilde{\mathbf{u}}_{ij}^n$  как комбинацию интерполяций проекций на столбцы матрицы  $\tilde{Q}_{ij}$ , соответствующие ненулевым собственным значениям матрицы  $\tilde{R}_{ij}^+$ . В этом случае матрица  $\tilde{\gamma}_{ij}^n$  коммутует с  $\tilde{R}_{ij}^+$  ( $\tilde{R}_{ij}^+ \tilde{\gamma}_{ij}^n = \tilde{\gamma}_{ij}^n \tilde{R}_{ij}^+$ ). Легко заметить, что в случае, когда для всех  $s$  выполнены неравенства  $[d_{ij}^n]_s \geq 0$ , для коэффициентов формулы (41) выполняются неравенства

$$[d_{ij}^n]_k \leq \max_s [d_{ij}^n]_s, \quad k = 1, m_{ij}^+. \quad (42)$$

Применяя покомпонентное ограничение на потоки (22)  $\min \left( [u_i^n]_k, [u_j^n]_k \right) \leq [u_{ij}^n]_k \leq \max \left( [u_i^n]_k, [u_j^n]_k \right)$ , из неравенства (42) получим

$$0 \leq [d_{ij}^n]_k \leq \max_s [d_{ij}^n]_s \leq 1. \quad (43)$$

Использование выражения (39) позволяет переписать схему высокого порядка (37) в виде симметрического оператора

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t^h}{V_i} \sum_j S_{ij} \left( \tilde{R}_{ij}^+ \mathbf{u}_i^n - \tilde{R}_{ji}^+ \mathbf{u}_j^n + [\tilde{R}_{ij}^+ \tilde{\gamma}_{ij}^n + \tilde{R}_{ji}^+ \tilde{\gamma}_{ji}^n] [\mathbf{u}_j^n - \mathbf{u}_i^n] \right),$$

что совместно с (17) дает

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t^h}{V_i} \sum_j S_{ij} \left( \tilde{R}_{ji}^+ - \tilde{R}_{ij}^+ \tilde{\gamma}_{ij}^n - \tilde{R}_{ji}^+ \tilde{\gamma}_{ji}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n). \quad (44)$$

Предполагая, что матрица  $\tilde{\gamma}_{ij}^n$  имеет вид (40) и выполняются условия  $\tilde{R}_{ij}^+ \cdot \tilde{R}_{ji}^+ = 0$  и  $\tilde{R}_{ij}^+ + \tilde{R}_{ji}^+ = |\tilde{R}_{ij}|$ , мы можем записать

$$\begin{aligned} \sum_j S_{ij} \left( \tilde{R}_{ij}^+ \mathbf{u}_i^n - \tilde{R}_{ji}^+ \mathbf{u}_j^n \right) &= \sum_j S_{ij} \left( \tilde{R}_{ji}^+ - \tilde{R}_{ij}^+ \tilde{\gamma}_{ij}^n - \tilde{R}_{ji}^+ \tilde{\gamma}_{ji}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n) = \\ &= \sum_j S_{ij} \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n), \end{aligned} \quad (45)$$

где  $\gamma_{ij}^n = \tilde{\gamma}_{ij}^n + \tilde{\gamma}_{ji}^n = \gamma_{ji}^n$ . В результате выражение (44) с использованием (45) дает

$$\mathbf{u}_i^{n+1} = \mathbf{u}_i^n - \frac{\Delta t^h}{V_i} \sum_j S_{ij} \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n), \quad (46)$$

что позволяет обозначить оператор аппроксимации высокого порядка точности через  $L_h^n$ , который мы можем переписать в компактном виде:

$$L_h^n = L_1 - \gamma^n.$$

Другими словами, симметризация (46) позволяет записать оператор аппроксимации схемы высокого порядка точности как сумму оператора первого порядка точности с некоторой добавкой.

Выражение (46) позволяет записать условие устойчивости схемы высокого порядка точности в виде

$$\begin{aligned} \|\mathbf{u}^{n+1}\|^2 &= \sum_i V_i \mathbf{u}_i^{n+1} \cdot \alpha_i \mathbf{u}_i^{n+1} = \sum_i V_i \left( \mathbf{u}_i^n - \frac{\Delta t^h}{V_i} \sum_j S_{ij} \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n) \right) \times \\ &\quad \times \alpha_i \left( \mathbf{u}_i^n - \frac{\Delta t^h}{V_i} \sum_k S_{ik} \left( \tilde{R}_{ki}^+ - |\tilde{R}_{ik}| \gamma_{ik}^n \right) (\mathbf{u}_i^n - \mathbf{u}_k^n) \right) = \\ &= \|\mathbf{u}^n\|^2 - 2\Delta t^h \sum_i \mathbf{u}_i^n \cdot \alpha_i \sum_j S_{ij} \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n) + \\ &\quad + (\Delta t^h)^2 \sum_i \frac{1}{V_i} \sum_j \sum_k S_{ij} S_{ik} \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \alpha_i \left( \tilde{R}_{ki}^+ - |\tilde{R}_{ik}| \gamma_{ik}^n \right) (\mathbf{u}_i^n - \mathbf{u}_k^n). \end{aligned}$$

Упростив выражения для граней  $ij$  в скалярном произведении  $(\mathbf{u}^{n+1}, \mathbf{u}^n) = \sum_i V_i \mathbf{u}_i^{n+1} \cdot \alpha_i \mathbf{u}_i^n$ , получим

$$\begin{aligned} \mathbf{u}_i^n \cdot \alpha_i S_{ij} \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n) + \mathbf{u}_j^n \cdot \alpha_j S_{ji} \left( \tilde{R}_{ij}^+ - |\tilde{R}_{ji}| \gamma_{ji}^n \right) (\mathbf{u}_j^n - \mathbf{u}_i^n) = \\ = \mathbf{u}_i^n \cdot \alpha_i S_{ij} \tilde{R}_{ji}^+ (\mathbf{u}_i^n - \mathbf{u}_j^n) + \mathbf{u}_j^n \cdot \alpha_j S_{ji} \tilde{R}_{ij}^+ (\mathbf{u}_j^n - \mathbf{u}_i^n) - (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \alpha_i S_{ij} |\tilde{R}_{ij}| \gamma_{ij}^n (\mathbf{u}_i^n - \mathbf{u}_j^n). \end{aligned}$$

Последнее выражение совместно с (24) дает

$$(L_h^n \mathbf{u}, \mathbf{u}) \geq \sum_i \sum_j S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \alpha_i \tilde{R}_{ji}^+ (\mathbf{u}_i^n - \mathbf{u}_j^n) - \sum_i \sum_j S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \alpha_i |\tilde{R}_{ij}| \gamma_{ij}^n (\mathbf{u}_i^n - \mathbf{u}_j^n),$$

что после упрощения приводит к неравенству

$$\begin{aligned} (L_h^n \mathbf{u}, \mathbf{u}) &\geq \sum_i \sum_j S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \alpha_i \left( \tilde{R}_{ji}^+ - |\tilde{R}_{ij}| \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n) = \\ &= \sum_i \sum_j S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \alpha_i |\tilde{R}_{ij}| \left( \frac{1}{2} I - \gamma_{ij}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n). \end{aligned} \tag{47}$$

При применении произвольной интерполяции неравенство (47) не всегда больше нуля. Следовательно, получение оценки (1) с выбранной нормой для произвольной аппроксимации невозможно. Поэтому для получения критерия сходимости схем высокого порядка точности необходимо выбрать либо подходящую норму, которая позволит получить надежный критерий устойчивости, либо ввести требование того, чтобы линейный оператор  $L_h^n$ , эквивалентный аппроксимации высокого порядка, был положительно определенным. Последнее возможно в случае выполнения условия

$$\left[ \tilde{d}_{ij}^n \right]_k \leq \frac{1}{2}, \quad k = 1, \dots, (m_{ij}^+ + m_{ji}^+). \tag{48}$$

При применении схемы MUSCL с покомпонентным ограничением на векторные потоки между соседними элементами (23) можно показать, что коэффициенты подчиняются условиям

$$0 \leq \left[ \tilde{d}_{ij}^n \right]_k \leq \frac{1}{2}, \tag{49}$$

а оператор  $L_h^n$  является положительным и самосопряженным. Другими словами, неравенство (48) дает возможность показать, что оператор (47) является положительным. Более того, неравенство (48) является векторным аналогом принципа максимума (22).

С целью оценки устойчивости схем, для которых не выполняется ограничение на потоки снизу (например, схема MUSCL, основанная на (19)), получим коэффициент подобия критерию устойчивости схеме первого порядка точности. Другими словами, исключив ограничения снизу, разрешив существование

отрицательных  $\left[ \tilde{d}_{ij}^n \right]_k$  и оставив только условие минимальности нормы градиента в схеме MUSCL (19), которое дает выполнение условия положительности (48) оператора  $L_h^n$ , мы можем получить оценку шага по времени. Из того, что норма положительного самосопряженного оператора равна его максимальному собственному значению, мы можем вычислить коэффициент разности между нормами для схемы высокого порядка точности и для нормы схемы первого порядка точности:

$$\delta = \max_{\mathbf{u}} \frac{\mathbf{u} \cdot L_h^n \mathbf{u}}{\mathbf{u} \cdot L \mathbf{u}},$$

что поэлементно дает

$$\delta_i \geq \max_{\mathbf{u}} \frac{\sum_{j \in I_i} S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot |\tilde{R}_{ij}| \left( \frac{1}{2} I - \gamma_{ji}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n)}{\sum_{j \in I_i} S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \frac{1}{2} |\tilde{R}_{ij}| (\mathbf{u}_i^n - \mathbf{u}_j^n)}. \quad (50)$$

Последнее выражение для каждой из граней поэлементно записывается в виде

$$\delta_i \geq \max_{\mathbf{u}, j \in I_i} \frac{S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot |\tilde{R}_{ij}| \left( \frac{1}{2} I - \gamma_{ji}^n \right) (\mathbf{u}_i^n - \mathbf{u}_j^n)}{S_{ij} \frac{1}{2} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot \frac{1}{2} |\tilde{R}_{ij}| (\mathbf{u}_i^n - \mathbf{u}_j^n)}. \quad (51)$$

Оценка (51) хоть и не является оптимальной и может быть улучшена путем численного расчета (50) для каждой аппроксимации, но является достаточно удобным способом получения критерия устойчивости схем высокого порядка точности.

Итак, в результате использования (51) коэффициент подобия со схемой высокого порядка точности можно получить в виде

$$\delta_i \geq 1 - 2 \min_{\mathbf{u}, j \in I_i} \frac{S_{ij} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot |\tilde{R}_{ij}| \gamma_{ji}^n (\mathbf{u}_i^n - \mathbf{u}_j^n)}{S_{ij} (\mathbf{u}_i^n - \mathbf{u}_j^n) \cdot |\tilde{R}_{ij}| (\mathbf{u}_i^n - \mathbf{u}_j^n)}.$$

Тогда из требования минимальности для каждой из граней мы получаем соотношение

$$\delta_i \geq 1 - 2 \min_{\mathbf{u}, j, k} \left[ \tilde{d}_{ij}^n \right]_k \geq 0, \quad (52)$$

которое позволяет получить ограничение (4), являющееся векторным аналогом принципа максимума. Последнее условие для схем высокого порядка точности позволяет сформулировать оценку шага по времени в виде

$$\Delta t^h \leq \min_i \frac{V_i}{c_i \delta_i \|G_i\|}, \quad 1 \leq \delta_i, \quad (53)$$

где на величину  $\delta_i$  можно либо наложить ограничение из оценки максимально возможных наклонов градиентов, либо предположить  $\delta_i = 1$ . Следует отметить, что для металлических стенок ( $j$  — номер плоского элемента на металлической стенке) для расчета ограничения на потоки мы предполагаем, что  $\mathbf{u}_j^n = -\mathbf{u}_i^n$ , а для внешней границы расчетной области можно и вовсе отказаться от ограничений на потоки. Полученное ограничение на минимальный шаг по времени на неструктурированной сетке, дающей разброс шагов по времени по элементам в (53), позволяет в существенной мере смягчить ограничения на потоки между элементами, имеющими больший шаг по времени, чем элемент, на котором мы получаем минимальный шаг по времени. Последнего можно добиться за счет использования для каждого элемента индивидуального коэффициента

$$\delta_i = \delta \frac{V_i \|G\|}{V \|G_i\|}, \quad (54)$$

где величины без индексов соответствуют элементу, на котором достигается минимальный шаг по времени. Использование так вычисляемых коэффициентов  $\delta_i$  позволяет не только смягчить ограничения на потоки на большей части элементов, но и вовсе отказаться от каких бы то ни было ограничений на элементах, для которых  $\delta_i \geq 2$ , что позволяет записать  $\Delta t^h = \Delta t^1 / \delta$ .

При использовании схемы MUSCL в местах локализации разрывов в случае допущения существования отрицательных  $[\tilde{d}_{ij}^n]_k$  в неравенстве (43) мы имеем отрицательные собственные значения у матрицы  $\gamma_{ij}^n$ , что дает уменьшение шага по времени. При этом минимальная величина  $[\tilde{d}_{ij}^n]_k$  на грани может быть найдена из максимальной величины наклона градиента в схеме MUSCL, которая зависит лишь от позиции центров масс элементов, используемых при аппроксимации градиента.

В схеме MUSCL (19) центральная аппроксимация используется, когда угол наклона односторонних аппроксимаций больше, чем угол наклона центральной аппроксимации, а минимальные значения  $[d_{ij}^n]_k$  достигаются, когда односторонний градиент несколько меньше, чем центральный. Поэтому с целью получения минимального значения  $[d_{ij}^n]_k$  и соответственно переключения способа аппроксимации градиента (19) рассмотрим детально упрощенный двумерный случай, представленный на рис. 1, согласно которому мы можем получить геометрическое условие использования разных аппроксимаций. В случае значительных разрывов мы получаем одностороннюю схему, в которой мы можем вывести соотношение между соседними величинами  $u_1$  и  $u_2$  и значением  $u_0$  на противоположных к разрыву элементах, при этом минимум  $d_{12}$  достигается, когда угол наклона центрального градиента совпадает с углом наклона одностороннего градиента. В результате мы можем заключить, что односторонняя схема применяется при  $|u_2 - u_0| > (b/a) \cdot |u_1 - u_0|$ , что позволяет взять такое  $\beta_1 = \frac{1 + \delta_1}{2} \geq 1 - d_{12} > \frac{1}{2}$ , чтобы  $\beta_1 |u_2 - u_1| \geq |u_{12} - u_2|$ . Здесь необходимо отметить, что  $a$  и  $b$  в трехмерном случае суть расстояния до плоскости, на которой лежат вершины, противоположные разрыву. Тогда последнее неравенство позволяет получить максимальную оценку для произвольного  $\beta_i = 1.5$ . В случае расчета оценки коэффициента  $\delta_i$  мы можем получить значения  $1 < \delta_i \leq 2$ . Другими словами, отказ от ограничения (49) на потоки между соседними элементами приводит к снижению допустимого шага по времени по отношению к схеме первого порядка точности.

Оценка (52) для схем высокого порядка точности позволяет сформулировать принцип максимума (4), который применим для схем высокого порядка точности и заключается в ограничении на собственные значения  $[\tilde{d}_{ij}^n]_k$  матрицы  $\tilde{\gamma}_{ij}^n$ , что требует дополнительных сложных вычислений. Поэтому на практике можно использовать более практичное, но менее строгое ограничение на компоненты векторов потоков. В результате для покомпонентных величин мы можем получить более практичное, но менее надежное ограничение в виде

$$\left| [u_{ij}^n]_k - \frac{[u_j^n]_k + [u_i^n]_k}{2} \right| \leq (1 + \delta_i) \frac{[u_j^n]_k - [u_i^n]_k}{2},$$

для которого в общем случае может и не выполняться условие (42), которое, тем не менее, применимо на практике и не приводит к неустойчивости вычислений, так же как и отмена на ограничение на потоки между соседними элементами с одинаковыми значениями расчетных величин.

**4. Численные эксперименты.** Широко известно, что далеко не все аппроксимационные схемы удовлетворяют принципу максимума, а схемы, к которым применено ограничение на потоки, приводят к снижению точности аппроксимации пространственных производных до первого порядка [16]. В случае применения принципа максимума мы получаем возможность гарантированно решать задачу с большим шагом по времени, но с потерей точности на разрывах. Примерами задач для уравнений Максвелла, в которых наблюдаются разрывы, являются задачи о столкновении разнонаправленных волн, которые возникают при расчете волноводов, эффектов дифракции и отражении радиоволн от различных предметов. Именно по этой причине важно изучение влияния ограничений, накладываемых принципом максимума на достаточно гладкие решения. Источником снижения точности для гладких решений может служить как несогласованность значений  $u_{ij}^n$  с осредненными величинами  $u_i^n$  и  $u_j^n$  на соседних элементах, так и неоднородность сетки. Именно поэтому в качестве примера такого расчета рассмотрим расчет отра-

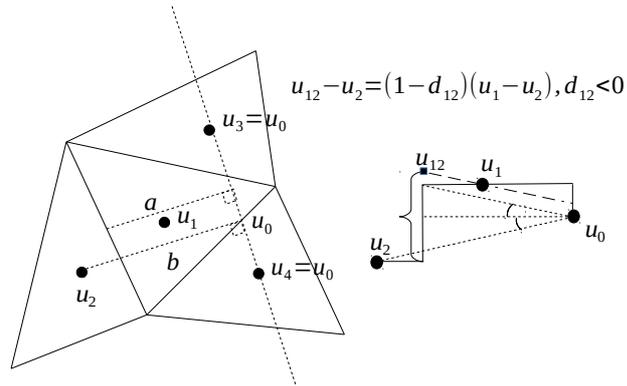


Рис. 1. Случай одинаковых наклонов центрального и одностороннего градиентов в схеме MUSCL на плоском элементе с индексом 1

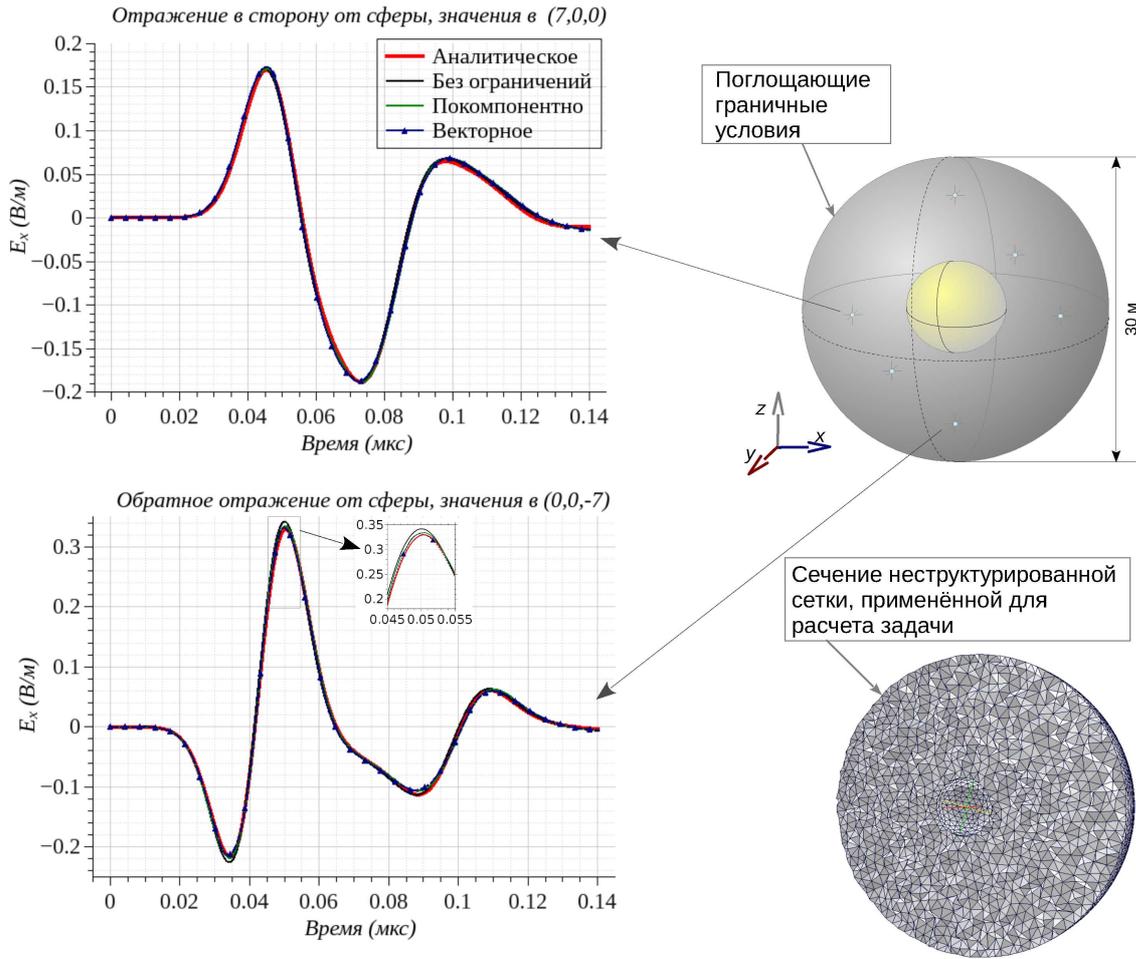


Рис. 2. Сравнение результатов расчетов для различных способов реализации ограничений на потоки

женного электрического поля от металлической сферы, состоящей из идеального проводника. Для такой задачи существует аналитическое решение, записываемое в виде суммы специальных функций [18]. Более того, эта задача интересна тем, что эффекты столкновения отраженных от сферы волн не приводят к разрушению решения без применения ограничений даже при  $\Delta t^1$ . Именно поэтому проведем сравнение отраженных волн электрической составляющей поля для двух типов ограничений на потоки с решением, полученным без каких бы то ни было ограничений.

Итак, сравним результаты расчетов отражения от металлической сферы с аналитическим решением. Для этого рассчитаем поле в точках  $(7, 0, 0)$  и  $(0, 0, -7)$ , отраженное от металлической сферы радиуса 3 метра с центром в начале координат, заключенной в расчетную область радиусом 15 метров. Начальное поле задано плоской волной, направленной вдоль оси  $\hat{n}_2 = (0, 1, 0)$ , и задано в виде  $E_1 = f'(t)$ ,  $E_2 = E_3 = 0$  и  $f(t) = \theta(t)a \exp(-(t - t_0)^2/b^2)$  с коэффициентами  $a = 1.33 \times 10^{-7}$ ,  $b = 1.14 \times 10^{-8}$ . Пусть  $t_0 = 3.5 \times 10^{-8}$  сек. Решение было получено для двух точек на неструктурированной сетке, состоящей из тетраэдров со средним размером грани около 0.75 метра, дающей незначительный разброс шагов по времени  $\Delta t^h \leq t_i \leq 3.4\Delta t^h$  для всех  $i$ , который возникает за счет неравномерности элементов. Для интерполяции значений между соседними элементами использовалась схема MUSCL. Расчеты отраженного поля в тестовых точках были проведены с применением различных способов ограничения, в которых использовалось условие  $\delta_i = \delta$  для всех  $i$ , что позволяет оценить влияние ограничения на потоки между соседними элементами на решение задачи. Полученные результаты, представленные на рис. 2, показывают, что введение ограничения на потоки между соседними элементами даже без использования (54) для гладких решений практически не влияет на качество полученных результатов. Незначительная диссипативность решения, вызванная ограничителями, является следствием неравномерности сетки и столкновения волн. Использование индивидуальных  $\delta_i$ , рассчитываемых по формуле (54) для каждого из расчетных элементов, позволяет в значительной степени исключить влияние ограничителя на потоки на большинстве

элементов неструктурированной сетки и существенно повысить качество расчетов.

Последнее показывает правильность полученных оценок для схем высокого порядка точности и позволяет с уверенностью применять полученную оценку на шаг по времени и ограничения на потоки между соседними элементами для схем высокого порядка точности без опасения получить искажения результатов в областях с гладкими решениями. При этом качество результатов в элементах, локализирующих разрыв, зависит от качества аппроксимации и жесткости ограничения на потоки между соседними элементами.

**Заключение.** Вычисляемый шаг по времени был применен к широкому кругу численных задач с использованием программы, описанной в работе [14] и примененной, например, к расчету задач магнито-резонансной томографии [19]. Применение принципа максимума или ограничения на величины потоков между соседними элементами позволяет гарантировать устойчивость схемы в районе расположения разрывов. Введение ограничения  $1 \leq \delta \leq 2$  на величину потоков между соседними элементами позволяет ввести критерий сходимости для схем высокого порядка точности  $\Delta t^h = \Delta t^1 / \delta$ , при котором величина шага по времени хоть и уменьшается, но схема практически не теряет точности и остается устойчивой. Более того, за счет применения (54) ограничение  $\delta$  в существенной мере снижает свою жесткость на основной части элементов неструктурированных сеток и не снижает точности решения. В случае же применения схем высокого порядка точности к областям решения, не имеющим разрывов, сохраняется критерий сходимости для схем первого порядка точности. Поэтому коэффициент  $\delta$  рекомендуется брать зависящим от задачи. В задачах, в которых решение не имеет разрывов, рекомендуется брать  $\delta = 1$ . В задачах же со значительным числом разрывов было достаточно ввести ограничение  $\delta = 1.4$ . В случае отказа от ограничений на разрывы рекомендуется вычислять локальные значения  $\delta_i$  либо брать шаг по времени в два раза меньший, чем для схемы первого порядка точности. Необходимо отметить, что критерий устойчивости для схем первого порядка точности позволяет получать устойчивые решения и для аппроксимаций высокого порядка точности на достаточно широком круге задач. Такие вычисления с шагами по времени для схем первого порядка точности, соответствующие  $\delta = 1$ , но без использования принципа максимума, проводились для неструктурированной и структурированной сеток. Для неструктурированной сетки было получено 5%–15% увеличение шага по времени по сравнению с критерием, вычисленным по формуле (2) из работы [3]. Для структурированной сетки, состоящей из кубических элементов и равносторонних тетраэдров, этот критерий дает тот же самый результат, что и критерий фон Неймана: в 1.5 раза больше, чем результат формулы (2) на кубических элементах. Последнее дает возможность заявить, что в работе приведен обобщенный, универсальный критерий устойчивости, применимый как на неструктурированных, так и на структурированных сетках.

#### СПИСОК ЛИТЕРАТУРЫ

1. Fumeaux C., Baumann D., Leuchtman P., Vahldieck R. A generalized local time-step scheme for efficient FDTD simulations in strongly inhomogeneous meshes // IEEE Transactions on Microwave Theory and Techniques. 2004. **52**, N 3. 1067–1076.
2. Lauritzen P.H. A stability analysis of finite-volume advection schemes permitting long time steps // Monthly Weather Reviews. 2007. **135**, N 7. 2658–2673.
3. Piperno S.  $L_2$ -stability of the upwind first order finite volume scheme for the Maxwell equations in two and three dimensions on arbitrary unstructured meshes // Mathematical Modelling and Numerical Analysis. 2000. **34**, N 1. 139–158.
4. Piperno S. Symplectic local time-stepping in non-dissipative DGTG methods applied to wave propagation problems // Mathematical Modelling and Numerical Analysis. 2006. **40**, N 5. 815–841.
5. Calgato C., Chane-Kane E., Creusé E., Goudon T.  $L_\infty$ -stability of vertex-based MUSCL finite volume schemes on unstructured grids: simulation of incompressible flows with high density ratios // Journal of Computational Physics. 2010. **229**, N 17. 6027–6046.
6. Coudière Y., Pierre C. Stability and convergence of a finite volume method for two systems of reaction-diffusion equations in electro-cardiology // Nonlinear Analysis: Real World Applications. 2006. **7**, N 4. 916–935.
7. Шурина Э.П., Великая М.Ю., Федорук М.П. Об алгоритмах решения уравнений Максвелла на неструктурированных сетках // Вычислительные технологии. 2000. **5**, № 6. 99–116.
8. Firsov D., Lo Vetri J. New stability criterion for unstructured mesh upwinding FDTD schemes for Maxwell's equations // ACES Journal. 2008. **23**, N 3. 193–199.
9. Gottlieb S., Shu C.-W., Tadmor E. Strong stability-preserving high-order time discretization methods // SIAM Review. 2001. **43**, N 1. 89–112.
10. Лебедев А.С., Федорук М.П., Штырина О.В. Решение нестационарных уравнений Максвелла для сред с неоднородными свойствами методом конечных объемов // Вычислительные технологии. 2005. **10**, № 2. 60–73.
11. Bonnet P., Ferrieres X., Michielsen B.L., Klotz P. Finite-volume time domain method // Time Domain Electromagnetics. New York: Academic Press, 1999. 307–367.

12. Chung E.T., Engquist B. Convergence analysis of fully discrete finite volume methods for Maxwell's equations in nonhomogeneous media // SIAM Journal on Numerical Analysis. 2006. **43**, N 1. 303–317.
13. Barth T., Ohlberger M. Finite volume methods: foundation and analysis // Encyclopedia of Computational Mechanics. Vol. 1. New York: Wiley, 2004. 1–57.
14. Firsov D., LoVetri J., Jeffrey I., Okhmatovski V., Gilmore C., Chamma W. High-order FVTD on unstructured grids using an object-oriented computational engine // ACES Journal. 2007. **22**, N 1. 71–82.
15. Firsov D., LoVetri J. New stability criterion for unstructured mesh upwinding FVTD schemes for Maxwell's equations // Proc. of the 23th Annual Review of Progress in Applied Computational Electromagnetics. Verona: ACES Press, 2007. 401–408.
16. Hubbard M.E. Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids // Journal of Computational Physics. 1999. **155**, N 1. 54–74.
17. Clain S. Finite volume maximum principle for hyperbolic scalar problems // SIAM J. Numer. Anal. 2013. **51**, N 1. 467–490.
18. Harrington R.F. Time-harmonic electromagnetic fields. New York: McGraw-Hill, 1961.
19. Kaye C., Gilmore C., Mojabi P., Firsov D., LoVetri J. Development of a resonant chamber microwave tomography system // Ultra-Wideband, Short Pulse Electromagnetics. Vol. 9. New York: Springer, 2010. 481–488.

Поступила в редакцию  
23.06.2013

---

## Stability of Explicit Schemes for Solving Maxwell's Equations by High-Order Finite Volume Methods

D. K. Firsov<sup>1</sup>

<sup>1</sup> *Geomodeling Technology Corp.; 1100-665 8 Street SW, Suite 1100, Calgary, AB T2P 3K7, Canada; Ph.D., Numerical Programmer, e-mail: d.k.firsov@gmail.com*

Received June 23, 2013

**Abstract:** A new stability criterion of explicit schemes for solving Maxwell's equations by high-order finite volume methods is proposed. The proof is based on a generalization of the stability criterion for the first-order finite volume scheme to the case of high-order schemes. The effect of discontinuities of the solution on the stability of high-order schemes is evaluated. The maximum principle for the finite volume approximations of vector conservation laws is discussed.

**Keywords:** Maxwell's equations, finite volume method, stability of explicit schemes, high-order accuracy, partial differential equations.

### References

1. C. Fumeaux, D. Baumann, P. Leuchtman, and R. Vahldieck, "A Generalized Local Time-Step Scheme for Efficient FVTD Simulations in Strongly Inhomogeneous Meshes," IEEE Trans. Microwave Theory Tech. **52** (3), 1067–1076 (2004).
2. P. H. Lauritzen, "A Stability Analysis of Finite-Volume Advection Schemes Permitting Long Time Steps," Monthly Weather Rev. **135** (7), 2658–2673 (2007).
3. S. Piperno, " $L_2$ -Stability of the Upwind First Order Finite Volume Scheme for the Maxwell Equations in Two and Three Dimensions on Arbitrary Unstructured Meshes," Math. Model. Numer. Anal. **34** (1), 139–158 (2000).
4. S. Piperno, "Symplectic Local Time-Stepping in Non-Dissipative DGTG Methods Applied to Wave Propagation Problems," Math. Model. Numer. Anal. **40** (5), 815–841 (2006).
5. C. Calgaro, E. Chane-Kane, E. Creusé, and T. Goudon, " $L_\infty$ -Stability of Vertex-Based MUSCL Finite Volume Schemes on Unstructured Grids: Simulation of Incompressible Flows with High Density Ratios," J. Comput. Phys. **229** (17), 6027–6046 (2010).
6. Y. Coudière and C. Pierre, "Stability and Convergence of a Finite Volume Method for Two Systems of Reaction-Diffusion Equations in Electro-Cardiology," Nonlinear Anal. Real World Appl. **7** (4), 916–935 (2006).
7. E. P. Shurina, M. Y. Velikaya, and M. P. Fedoruk, "On Algorithms for the Solution of Maxwell's Equations on Non-Structured Grids," Vychisl. Tekhnol. **5** (6), 99–116 (2000).

8. D. Firsov and J. LoVetri, "New Stability Criterion for Unstructured Mesh Upwinding FVTD Schemes for Maxwell's Equations," *Appl. Comput. Electromagn. Soc. J.* **23** (3), 193–199 (2008).
9. S. Gottlieb, C.-W. Shu, and E. Tadmor, "Strong Stability-Preserving High-Order Time Discretization Methods," *SIAM Rev.* **43** (1), 89–112 (2001).
10. A. S. Lebedev, M. P. Fedoruk, and O. V. Shtyrina, "Solution of Unsteady Maxwell's Equations for Inhomogeneous Media by Finite Volume Method," *Vychisl. Tekhnol.* **10** (2), 60–73 (2005).
11. P. Bonnet, X. Ferrieres, B. L. Michielsen, and P. Klotz, "Finite-Volume Time Domain Method," in *Time Domain Electromagnetics* (Academic, New York, 1999), pp. 307–367.
12. E. T. Chung and B. Engquist, "Convergence Analysis of Fully Discrete Finite Volume Methods for Maxwell's Equations in Nonhomogeneous Media," *SIAM J. Numer. Anal.* **43** (1), 303–317 (2006).
13. T. Barth and M. Ohlberger, "Finite Volume Methods: Foundation and Analysis," in *Encyclopedia of Computational Mechanics* (Wiley, New York, 2004), Vol. 1, pp. 1–57.
14. D. Firsov, J. LoVetri, I. Jeffrey, et al., "High-Order FVTD on Unstructured Grids Using an Object-Oriented Computational Engine," *Appl. Comput. Electromagn. Soc. J.* **22** (1), 71–82 (2007).
15. D. Firsov and J. LoVetri, "New Stability Criterion for Unstructured Mesh Upwinding FVTD Schemes for Maxwell's Equations," in *Proc. 23rd Annual Review of Progress in Applied Computational Electromagnetics Conf., Verona, March 19–23, 2007* (ACES Press, Verona, 2007), pp. 401–408.
16. M. E. Hubbard, "Multidimensional Slope Limiters for MUSCL-Type Finite Volume Schemes on Unstructured Grids," *J. Comput. Phys.* **155** (1), 54–74 (1999).
17. S. Clain, "Finite Volume Maximum Principle for Hyperbolic Scalar Problems," *SIAM J. Numer. Anal.* **51** (1), 467–490 (2013).
18. R. F. Harrington, *Time-Harmonic Electromagnetic Fields* (McGraw-Hill, New York, 1961).
19. C. Kaye, C. Gilmore, P. Mojabi, et al., "Development of a Resonant Chamber Microwave Tomography System," in *Ultra-Wideband, Short Pulse Electromagnetics* (Springer, New York, 2010), Vol. 9, pp. 481–488.