УДК 621.396.4

МЕТОД ДИФФЕРЕНЦИАЛЬНО-ОРТОГОНАЛЬНОГО ПОИСКА В ЗАДАЧЕ ОЦЕНКИ MIMO КАНАЛА

Ляшев В.А., *к.т.н., доцент кафедры мультимедийных технологий и телекоммуникаций, Московский физико-технический институт (национальный исследовательский университет), e-mail: lyashev.va@mipt.ru*

A METHOD OF DIFFERENTIAL ORTHOGONAL MATCHING PURSUIT FOR MIMO CHANNEL ESTIMATION

Lyashev V.A.

Today, active antenna arrays are increasingly being used in wireless communication. The systems become highly selective in space, which makes them sensitive to errors in channel estimations. The paper considers the main problems of MIMO systems in which it is not possible to evaluate the channel using direct methods based on pilot signals. Implicit estimation methods require a compressed representation of channel state information, which imposes limitations on their accuracy. A method of iterative recovery of channel information from a compressed representation is proposed, which has an improved efficiency of channel matrix acquisition and the effect of error smoothing.

Key words: MIMO-OFDM, vector quantization, channel estimation, spatial multiplexing.

Ключевые слова: MIMO-OFDM, векторное квантование, оценка канала, пространственное мультиплексирование, сжатые измерения, дифференциальный метод.

Сегодня активные антенные решетки находят все большее применение в задачах беспроводной связи. Системы связи становятся высокоизбирательными по пространству, что делает их чувствительными к ошибкам в оценках канала. В работе рассматриваются основные проблемы MIMO систем, в которых нет возможности оценить канал прямыми методами по пилотным сигналам. Неявные методы оценки требуют сжатого представления канальной информации, что накладывает ограничения на их точность. Предлагается метод итеративного восстановления канальной информации из сжатого представления, который обладает повышенной точностью восполнения канальной информации и эффектом сглаживания возможных ошибок в поступившей информации.

Введение

Современные сети подвижной связи требуют эффективных алгоритмов обработки сигналов как в восходящем канале (uplink, UL), так и в нисходящем канале (downlink, DL). Для одновременной работы этих каналов применяются различные методы множественного доступа, включая систему множественного доступа с частотным разделением каналов (FDMA) [1], которая позволяет мультиплексировать данные пользователей по разным частотным ресурсам.

Для поддержания связи в UL и DL также используются дуплекс с временным разделением (TDD) и дуплекс с частотным разделением (FDD) [1]. Последний обеспечивает полудуплексную и полнодуплексную передачу данных. Полнодуплексный режим позволяет одновременно передавать и принимать данные на разных частотах f_{UL} и f_{DL}, что делает его предпочтительным вариантом и как следствие – самым распространённым на сегодня.

Чтобы обеспечить эффективность систем с высоким коэффициентом разнесения несущих частот восходящего и нисходящего каналов, базовые станции (БС) должны получать информацию о состоянии нисходящего канала от мобильных станций (МС). Для этого БС передают опорные символы (пилоты), которые используются МС для измерений. Полученная оценка канала затем передается обратно на БС с помощью отчетов о состоянии канала (channel state information, CSI) – такой процесс называют обратной связью в оценке канала (OC-CSI).

Сжатие CSI перед отправкой на БС позволяет уменьшить объем служебной информации, необходимой для передачи состояния канала, при сохранении достаточной точности для управления системой. Этот процесс включает в себя преобразование непрерывных значений CSI в дискретные уровни в многомерном пространстве малого ранга, что позволяет сократить количество бит, необходимых для их представления. Данная информация является служебной и обеспечивает работоспособность нисходящего канала, поэтому иногда объем служебной информации относят к накладным расходам информационной емкости восходящего канала (накладные расходы алгоритма) и называют отчетами о состоянии канала.

Обширный обзор в [2] показывает разнообразие методов, учитывающих ограниченный канал OC-CSI. В качестве одного из решений предлагается сопоставлять состояния канала DL с кодовыми словами из словаря известного как БС, так и МС. Точность и размер CSI будет определяться размером словаря и методом поиска кодового слова, наилучшим образом согласованного с текущим состоянием канала. В данной работе предлагается рассмотреть итеративный метод подбора кодовых слов, которые эффективно и компактно представят информацию о канале для последующей передачи ее на БС.

Качество оценки канала влияет на точность пространственных фильтров, которые применяются для пространственного мультиплексирования пользователей. Стандарт поддерживает пространственное мультиплексирование до 12 МС, и каждая МС обслуживается несколькими потоками данных. Для словаря второго типа [3] средняя суммарная пропускная способность одной ячейки показана на рис. 1. Одинаковый тип линий соответствует одной детализации отчетов CSI в частотной области (пунктирная линий: одна оценка канала на полосу 720 кГц; сплошная линия – одна оценка канала на полосу 1.44 МГц). Линии с круглыми маркерами обозначают не квантованные CSI, а линии с треугольными маркерами - квантованные CSI. Рис. 1 позволяет оценить влияние квантования с использованием кодовых книг на производительность системы.

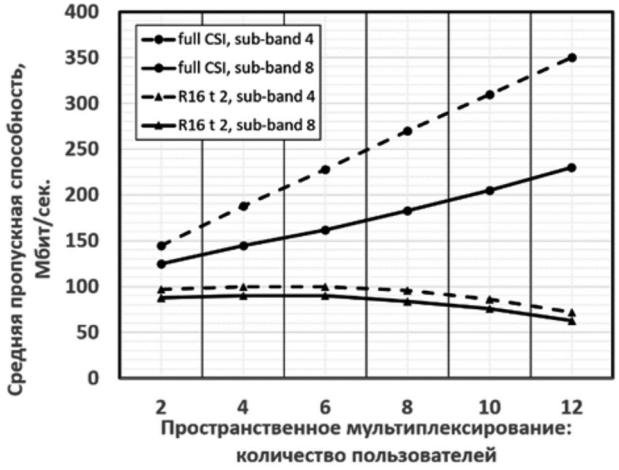


Рис. 1. Несжатый CSI (●) и CSI на основе словаря типа II (▲)

Видно, что сплошные кривые имеют экстремум, который объясняется тем, что недостаточно точное описание канальной информации посредством словарей из стандарта приводит к взаимной интерференции между мультиплексированными пользователями, суммарный уровень которой начинает доминировать начиная с 5-6 мультиплексированных МС. В то же время, полная информация о канале позволяет и далее наращивать производительность системы, пропускная способность которой будет ограничиваться лишь доступной мощностью передатчика, наличием МС для мультиплексирования и вычислительными возможностями БС. Данные характеристики можно рассматривать как верхнюю границу, которая обеспечивает максимально достижимую производительность MU-MIMO системы.

Сегодня 3GPP стандарт поддерживает 12 МС, но при этом словари второго типа из 16 и 18 версий стандарта работают в 2-3 раза хуже, чем предполагает верхняя граница (пунктирный график на рис. 1), что указывает на потребность дальнейшего улучшения методов представления и сжатия пространственно-временных оценок каналов MIMO-систем.

В работе предлагается улучшенная конструкция словаря, которая позволяет сократить разрыв между верхней границей и рассматриваемым решением.

Постановка задачи

Конструкции кодовых книг оптимальны только для определенной конфигурации системы. В литературе большинство словарей разработаны для систем с несколькими входами и одним выходом (multiple input – single output, MISO). Пусть антенная система БС состоит из N_{TX} элементов, а МС из $N_{RX}=1$. Ограничение на одну приемную антенну подразумевает один пространственный канал r = 1. Для простоты предлагается рассматривать аналогичный случай. Обобщение полученных результатов на $r>1$ будет сделано в последующих работах.

Для MISO систем принятый сигнал $y(t)\in\mathbb{C}$ в момент времени t для каждой МС записывают следующим образом [4] [5]

$$y(t)=\int \mathbf{h}^{H}(t,\tau)\mathbf{x}(t-\tau)d\tau+n(t)= \\ =\iint h(t,\tau,\mathbf{b})\mathbf{g}^{H}(\mathbf{b})x(t-\tau)d\tau d\mathbf{b}+n(t), \quad (1)$$

где $(\cdot)^{\mathrm{H}}$ – самосопряженный оператор. Вектор канала $\mathbf{h}\in\mathbb{C}^{N_{TX}}$ представляет собой произведение коэффициентов передаточной характеристики канала $h(t,\tau,\mathbf{b})\in\mathbb{C}$ в каждый момент времени t и для каждой задержки τ импульсной характеристики канала с диаграммой направленности антенной системы $\mathbf{g}(\mathbf{b})\in\mathbb{C}^{N_{TX}}$. Передающая антенна БС может быть описана набором параметров $\mathbf{b}\in\mathcal{P}$, который определен на многообразии $\mathcal{G}=\{\mathbf{g}(\mathbf{b}),\ \forall\mathbf{b}\in\mathcal{P}\}$. Из-за многолучевого распространения сигнала $\mathbf{x}(t)\in\mathbb{C}^{N_{TX}}$ в среде, МС всегда наблюдает сдвинутые копии сигнала на задержку τ. Более того, передаваемый сигнал представляет собой многомерный пространственный сигнал, порожденный пространственным фильтром $\mathbf{p}\in\mathbb{C}^{N_{TX}}$ и комплексным квадратурным сигналом $s(t)$

$$\mathbf{x}(t)=\mathbf{p}(t)s(t). \quad (2)$$

Квадратурный сигнал $s(t)$ характеризуется средней мощностью $P_S=\mathbb{E}(|s(t)|^2)$, здесь $\mathbb{E}(\cdot)$ – оператор оценки математического ожидания случайного процесса. Шум $n(t)\in\mathcal{CN}(0,\sigma_n^2)$ – комплексный нормальный случайный процесс с нулевым математическим ожиданием и дисперсией σ_n^2. Выражение (1) сформулировано во временной области и не зависит от вида сигнала. Однако, в данной работе будет рассматриваться система связи с ортогональным частотным мультиплексированием (orthogonal frequency division multiplexing, OFDM), что наиболее характерно для современных систем беспроводной связи [1].

Отчеты информации о канале представляют собой кодовые слова, которые более всего схожи с измеренным нисходящим каналом на МС. На основе таких отчетов, БС восстанавливает информацию о канале и формирует пространственный согласованный фильтр $\mathbf{p}$ (ПСФ). Зная точно матрицу пространственной корреляции канала, можно получить оптимальный вектор $\mathbf{p}$ через разложение матрицы на сингулярные числа и вектора [6]. Тогда первые значимые сингулярные вектора будут обеспечивать оптимальную пространственную фильтрацию.

Пусть словарь кодовых слов $\mathcal{C}$

$\mathcal{C} = (\mathbf{c}_j : \mathbf{c}_{j2} = 1,\ 0 \le j \le |\mathcal{C}| - 1),\ \mathbf{c}_j \in \mathbb{C}^{N_{TX}}$,

будем использовать N вместо N_{RX} для простоты, подразумевая $N_{RX} = 1$.

Размер словаря определяется количеством кодовых слов и обозначается $|\mathcal{C}|$. Очевидно, что, зная словарь, достаточно сообщить номер кодового слова, тогда можно определить количество бит B, которые необходимо использовать для информирования БС: $|\mathcal{C}| = 2^B$.

Задача МС заключается в быстром выборе наилучшего кодового слова из словаря, которое обеспечивает наименьшие потери от такого способа представления канала. Для системной модели (1) можно записать отношение сигнал/шум [7]

$$\gamma = \frac{P_S}{\sigma_n^2} \left| \mathbf{h}^H \mathbf{p} \right|^2. \quad (3)$$

МС в состоянии оценить γ и решить задачу поиска оптимального вектора $p = p_{opt}$

$$\mathbf{p}_{opt} = \arg\max_{\mathbf{p} \in \mathcal{C}} \frac{\left| \mathbf{h}^H \mathbf{p} \right|^2}{\mathbf{h}_2^2} = \arg\max_{\mathbf{p} \in \mathcal{C}} \left| \bar{\mathbf{h}}^H \mathbf{p} \right|^2, \quad (4)$$

где $\bar{\mathrm{h}} = \mathrm{h} / \|\mathrm{h}\|_2$ нормированный вектор канала. В этом случае, МС определит вектор

$\mathbf{c}_{opt} = \min_{j \subset [0\ |\mathcal{C}|-1]} \mathbf{c}_j - \bar{\mathbf{h}}_2$,

который будет передан на БС. В данной статье предложен иной подход к построению таких словарей.

Структурированные словари

По сравнению с неструктурированными словарями, в этом разделе описывается категория структурированных словарей. Это означает, что кодовые слова имеют специальное правило построения. Здесь вводятся две подкатегории:

Первая подкатегория по-прежнему включает в себя вероятностные методы для получения словарей, такие как ДПФ-словари, которые широко используются на практике, начиная с 3GPP стандарта десятой версии.

Во второй подкатегории правила построения словарей основываются на структурированности матриц канала и принципах линейной алгебры только. Этой категории и будет уделено основное внимание в данном разделе.

Сформулируем задачу разреженных измерений в каноническом виде. Пусть необходимо восстановить разреженный вектор $\mathbf{x}$ заданной длины $|\mathcal{C}|$ из вектора измерений $\mathbf{y}$, используя матрицу-словарь $\mathbf{C} \in \mathbb{C}^{N \times |\mathcal{C}|}$

$$\mathbf{C}\mathbf{x} = \mathbf{y}. \quad (5)$$

Здесь количество измерений в $\mathbf{y}$ намного меньше, чем длина разреженного вектора $\mathbf{x}$. Ограничение задачи на поиск 1-мерного разреженного $\mathbf{x}$ для приближения к $\mathbf{y}$ эквивалентно поиску наилучшего столбца матрицы-словаря $\mathbf{c}$, т.е. наилучшего кодового слова, описывающего $\mathbf{y}$. Эту проблему представления часто называют разреженным измерением или разреженным кодированием. Такие подходы имеют границу сверху, которая определяется условием Уэлча-Рэнкина. Для комплекснозначных словарей справедлива граница ([8], теорема 5.10).

$$|\mathcal{C}| \le N^2. \quad (6)$$

Для представления вектора измерений y в виде разреженного вектора с незначительными искажениями, грубая оценка (6) неприменима. Используя границы Уэлча-Рэнкина ([8], стр. 114), можно записать ее через минимальное расстояние d_{min} между кодовыми словами в матрице-словаре:

$$0 \le d_{min}(\mathcal{C}) \le \sqrt{1 - \frac{1}{N+1}}. \quad (7)$$

Известно, что для определенных измерений в некоторых словарях достигается именно эта граница. Здесь основное внимание уделяется мягкой оценке сверху, которая достигается в кодовых книгах Аллтопа. Так, для простых чисел $N \ge 5$, последовательности Аллтопа [9] достигают мягкой оценки сверху ([8], стр. 121).

$$d_{min}^{Alltop}(\mathcal{C}) = \sqrt{1 - \frac{1}{N}} < d_{min}. \quad (8)$$

Предположим, что матрица-словарь $\mathbf{C} \in \mathbb{C}^{N \times N^2}$ может быть разделена на N блоков размера $N \times N$

$$\mathbf{C} = (\mathbf{C}^{(0)} \quad \mathbf{C}^{(1)} \quad \dots \quad \mathbf{C}^{(N-1)}). \quad (9)$$

Тогда блоки $\mathbf{C}^{(k)}$ $k = 1, 2, \dots, N-1$ можно сформировать посредством матрицы сдвига $\mathbf{D}_k$ и матрицей модуляции $\mathbf{G}_N$ [10] размерностей $N \times N$: $\mathbf{C}^{(k)} = \mathbf{D}_k \mathbf{G}_N$, где последовательность Аллтопа [9] $f[q] = \frac{1}{\sqrt{N}} e^{j 2\pi q^3 / N}$, $q = 0, 1, \dots, N-1$ может быть сдвинута на k отсчетов в матрице сдвига $\mathbf{D}_k = \mathrm{diag}\{f_k, \dots, f_{N-1}, f_0, \dots, f_{k-1}\}$, а матрица модуляции $\mathbf{G}_N = (g_N^{(pq)})_{p,q=0}^{N-1}$, $g = e^{j 2\pi / N}$.

Соответствующая структура кодовой книги Аллтопа представлена на рис. 2, где показана матрица Грамма. Главная диагональ, выделенная желтым цветом, соответствует автокорреляции кодового слова. Блочная структура словаря Аллтопа хорошо видна. Кодовые слова в блоке $\mathbf{C}^{(k)}$ не имеют корреляции друг с другом, т.е. они ортогональны. Корреляции между кодовыми словами блока $\mathbf{C}^{(k)}$ и другими блоками $\mathbf{C}^{(l)}$, $l \ne k$ имеют постоянное значение $1/\sqrt{N}$.

На рис. 3 показано минимальное расстояние, доступное для кодовых книг Аллтопа. Хотя, граница Уэлча-Рэнкина не достигается с помощью этой конструкции, но приведенная выше мягкая оценка границы Аллтопа (8) полностью достижима (Alltop bound).

Нужно заметить, что для реализации книг Аллтопа необходимый размер словаря квадратично зависит от размерности измеряемого пространства $|\mathcal{C}| \approx N^2$.

Таким образом, в кодовых книгах Аллтопа кодовые слова формируются на основе последовательностей с определенными корреляционными свойствами и не учитывают особенности распределений измеряемых вели-

чин. Они не привязаны к критерию Уэлча-Рэнкина, но конструкции близки к этому критерию. Количество элементов велико из-за правил построения, что подразумевает более высокие накладные расходы на обратную связь, но при этом обеспечивают небольшие ошибки аппроксимации. В целом, такие словари могут рассматриваться как достижимая граница по точности представления канальной информации. Любые последующие подходы – это попытки снизить накладные расходы в петле обратной связи OC-CSI.

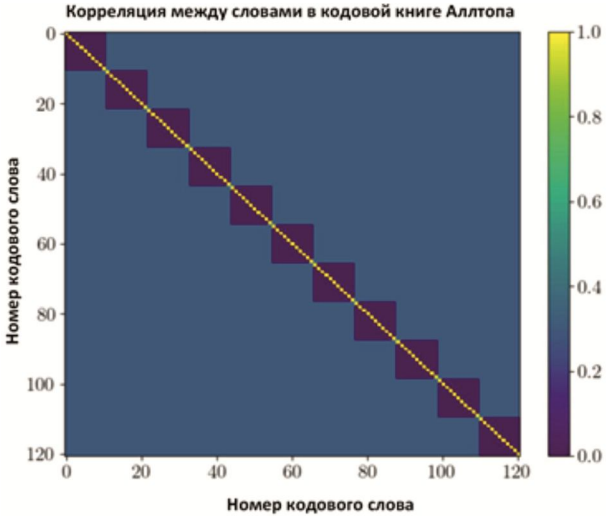


Рис. 2. Структура кодовой книги Аллтопа

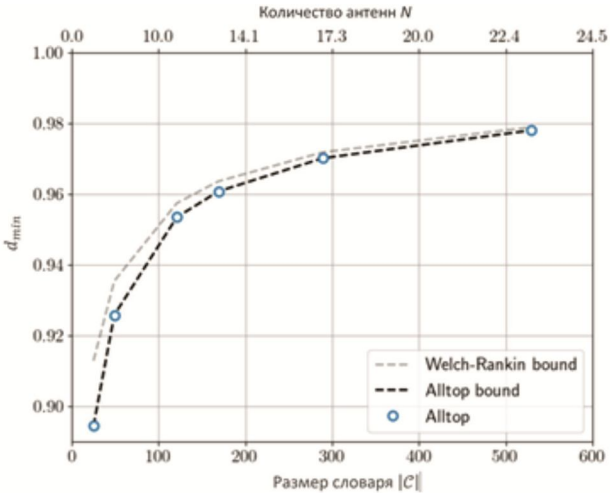


Рис. 3. Минимальное расстояние между кодовыми словами в словаре Аллтопа

Итеративные словари

Итеративные словари основываются на грассмановом многообразии. Чтобы продолжить введение грассманова многообразия, удобно кратко изложить решение задачи нахождения оптимального вектора пространственного фильтра относительно максимального SNR, которое задается формулой (4). В задаче рассматриваются решения, которые приводят к максимуму в смысле абсолютных значений. В результате вектор $\mathbf{p}$ из (4) не является уникальным, поскольку любой повернутый вектор $e^{j\alpha}\mathbf{p}$, $\alpha \in [0\ 2\pi]$ также находится в наборе решений. Все векторы в этом наборе имеют одну и ту же линию в $\mathbb{C}^N$. Комплексное многообразие Грассмана $G(K,\mathbb{C}^d)$ описывает множество K -мерных подпространств $\mathbb{C}^d$. Установив $K=1$ и $d=N$, многообразие Грассмана $G(1,\mathbb{C}^N)$ представляет собой набор всех одномерных подпространств (линий), проходящих через начало координат, и может быть использовано для математического описания оптимального набора векторов $\mathbf{p}$.

Словари должны охватывать большинство наборов решений, которые могут возникнуть. Однако ограничение накладных расходов на обратную связь ограничивает размер словаря $|\mathcal{C}|$. Чтобы эффективно охватить все наборы и ограничить потери при квантовании, необходимо задачу свести к поиску линий из $G(1,\mathbb{C}^N)$ расположенных на максимальном расстоянии друг от друга. В литературе такую задачу называют проблемой упаковки грассмановых линий или задачей плотной упаковки. Для измерения расстояния между линиями на многообразии Грассмана, чаще всего применяют сферическую метрику

$$d_{ch}(\mathbb{C}^d) = \sqrt{\sin^2\theta_1 + \ldots + \sin^2\theta_K},$$

которую обычно называют хордальным расстоянием. В условиях применения теории к задаче измерения точности представления канала векторами кодовой книги $G(K,\mathbb{C}^d) \to G(1,\mathbb{C}^N)$ сферическая метрика вырождается в

$$d_{ch}(\mathbb{C}^N) = d_{ch} = \sin\theta. \qquad (10)$$

и принимает значения в пределах [0,1]. Угол между двумя грассмановыми линиями обозначается как θ. В терминах линейной алгебры можно определить линию как множество точек, которые принадлежат комплексной N-мерной гиперплоскости (алгебраическая кривая в $\mathbb{C}^N$). Тогда метрика (10) есть угол между двумя гиперплоскостями в пространстве $\mathbb{C}^N$.

Алгоритмической основой таких подходов является алгоритм жадного ортогонального поиска (Orthogonal Matching Pursuit, OMP) [11], хорошо известный в области разреженных измерений/аппроксимаций, который представляет собой алгоритм, определяющий собственные векторы путем итеративных приближений.

Ранее, словарь Аллтопа был представлен в контексте разреженных измерений. Рассмотрим алгоритм OMP - это итеративный алгоритм, который использует одно кодовое слово за один итерационный шаг $k = 1,2,\ldots,K$. После выполнения максимум $k = K$ шагов найдено K кодовых слов, которые складываются в новую матрицу кодовой книги $\tilde{\mathbf{C}}^{(k)} \in \mathbb{C}^{N\times(k-1)}$. Индексы $j_1, j_2, \ldots, j_{k-1} \in \in \{1,2,3\ldots,|\mathcal{C}|\}$ соответствуют выбранным колонкам в исходной измерительной матрице $\mathbf{c}$ такие, что $\tilde{\mathbf{C}}^{(k)} = (\mathbf{c}_{j_1}\ \ \mathbf{c}_{j_2}\ \ \ldots\ \ \mathbf{c}_{j_{k-1}})$. Набор индексов обозначает измеряемые элементы разреженного вектора $\mathbf{x}$, и от итерации к итерации таких элементов в векторе $\mathbf{x}$ будет становиться все больше и больше. Таким образом, алгоритмы, в основе которых лежит метод OMP, должны передать K индексов соответствующих столбцов матрицы $\mathbf{C}$ для восстановления вектора $\mathbf{x}$. Для стационарного канала это относительно приемлемое решение.

Однако в динамике оно становится непригодным, так как итерации распределены во времени, а также данный метод работает эффективно для первого сингулярного вектора канальной матрицы, но малоприменим для векторов более высоких порядков.

Рассмотрим дифференциальный подход, который позволяет получать дополнительные обновления CSI для $K \ll N$. Вместо инициализации вычислений OMP случайным вектором $\mathbf{x}$ – инициализируем его вектором невязки.

Вектор невязки – это разница между текущим собственным вектором $\mathbf{v}_1(t_s)$ и результатом K OMP итераций, в которых найдены кодовые слова $\tilde{\mathbf{C}}(t_{s-1})$ и коэффициенты $\mathbf{x}(t_{s-1})$

$$\mathbf{r}(t_s) = \mathbf{v}_1(t_s) - \tilde{\mathbf{C}}(t_{s-1})\mathbf{x}(t_{s-1}). \quad (11)$$

Первоначально невязка определяется только собственным вектором. На каждом шаге квантуется только неизвестное. По сравнению с традиционным OMP, дифференциальный метод обеспечивает повышенную гибкость, поскольку качество обратной связи может регулироваться параметрами K и t. Таким образом, можно выбрать между несколькими кодовыми словами и постоянным размером накладных расходов в петле обратной связи. Поскольку параметр K может быть меньше, чем N, дифференциальный подход существенно экономит накладные расходы. Принцип интеграции алгоритма в систему связи схемотично изображен на рис. 4.

Кратко, алгоритм можно описать следующим образом (см. алгоритм 1).

Рис. 4. Принцип применения дифференциального алгоритма ортогонального жадного перебора (ДАП)

Здесь $S^{(k)}$ – подмножество доступных для выбора колонок в измерительной матрице $\tilde{\boldsymbol{C}}^{(k)}$.

При каждом вызове OMP на шаге 3 будут выбираться кодовые слова из полного набора кодовой книги, т.е. нет ограничений на базисные вектора словаря с течением времени. Следовательно, векторы ошибок e в OMP ортогональны, а вектор невязки $\mathbf{r} = 0$. БС суммирует отчеты по формуле $\mathbf{v}_1 = \mathbf{C}(t_s)\mathbf{x}(t_s)$, чтобы каждый отчет улучшал оценку. Кроме того, можно установить критерий остановки, чтобы не нужно было отправлять коэффициенты на БС, если вектор невязки достиг малых значений, например $\| \mathbf{r} \| < \sigma_n$.

Алгоритм 1 – дифференциальный алгоритм ортогонального жадного перебора (ДАП).

1 Инициализация: $s = 0$

2 Вычислить вектор невязки между текущим собственным вектором и результатом OMP на предыдущем временном шаге.

$\boldsymbol{r}(t_s) = \boldsymbol{v}_1(t_s) - \tilde{\boldsymbol{C}}(t_{s-1})\boldsymbol{x}(t_{s-1})$

Замечание:

$\boldsymbol{r}(t_0) = \boldsymbol{v}_1(t_0)$.

3 Выполнить K итераций OMP алгоритма

a) $k = 0, S^{(0)} = 0,\ \boldsymbol{x}^{(0)} = 0, \boldsymbol{e}^{(0)} = \boldsymbol{r}(t_s)$.

b) Выбрать кодовое слово согласно условию

$j_k = \arg \max_{j_k \in S^{(k-1)}} \left| \boldsymbol{c}_{j_k}^H \boldsymbol{e}^{(k-1)} \right|$

c) Обновить матрицу $\tilde{\boldsymbol{C}}^{(k)} = (\tilde{\boldsymbol{C}}^{(k-1)} \quad \boldsymbol{c}_{j_k})$

$S^{(k)} = S^{(k-1)} \bigcup \{j_k\}$

d) Обновить оценку вектора $\boldsymbol{x}^{(k)}$ путем решения задачи наименьших квадратов

$\tilde{\boldsymbol{x}} = \arg \min_{z \in \mathbb{C}^k} r(t_s) - \tilde{\boldsymbol{C}}^{(k)} \boldsymbol{z}_2^2$

e) $\tilde{\boldsymbol{x}} \overset{S^{(k)}}{\rightarrow} \boldsymbol{x}^{(k)}$

f) Обновить вектор невязки

$\boldsymbol{e}^{(k)} = \boldsymbol{r}(t_s) - \boldsymbol{C}\boldsymbol{x}^{(k)}$

g) $k = k + 1$ и перейти к шагу b) пока $k \neq K$

4 Передать на БС индексы кодовых слов из $\tilde{\boldsymbol{C}}(t_s)$ и коэффициенты вектора $\boldsymbol{x}(t_s)$

5 Перейти к следующему временному шагу

$s = s + 1$ и вернуться на шаг 2.

Ниже представлены различные результаты тестирования ДАП. На рис. 5 показана нормированная корреляция алгоритма ДАП с кодовой книгой Аллтопа для разных значений K с течением времени. Графики на рис. 5 и рис. 6 относятся к одним условиям моделирования и отличаются только рассматриваемым периодом времени. Цель состоит в том, чтобы достичь значения корреляции, равного единице за минимальное число итераций. Из рисунка видно, что значение K влияет на наклон кривой. Таким образом, для больших значений K, т.е. для большего количества кодовых слов, передаваемых на БС за один раз, требуется меньше итераций, а для меньших значений требуется больше итераций. При $K=1$, т.е. при каждом сообщении передается одно кодовое слово, усредненная корреляция меньше 1. Это означает, что изменения в собственном векторе приводят к большим ошибкам, которые невозможно квантовать с помощью одной итерации OMP.

На рис. 7 показана корреляция во времени для различных значений скорости, где $K=1$ для левого рисунка и $K=4$ для правого. При $K=1$ более высокие скорости приводят к снижению производительности. Соответственно, ДАП с $K=1$ применим только для каналов с очень медленным затуханием. В противном случае интервалы обратной связи следует выбирать короче, что не всегда возможно. Напротив, при более высоких значениях K, таких как $K=4$, как показано на рис. 7, *б*, можно полностью компенсировать снижение скорости.

В заключение, параметр K следует скорректировать в соответствии со сценарием, при этом для обеспечения высокой мобильности требуются более высокие значения, в то время как для сценариев с низкой мобильностью достаточно меньших значений.

Далее исследуется эффективность обратной связи ДАП и поведение при различных конструкциях кодовых книг, что показано на рис. 8. Рассмотрены следующие конструкции кодовых книг: случайное векторное квантование (RVQ), ДПФ словарь из 5G стандарта (NR-DFT), а также словарь Аллтопа. На рис. 8, *а* показана нормированная корреляция для $K=1$ и скорости 10 км/ч. В среднем влияние кодовых книг на корреляцию невелико, хотя у словарей Аллтопа есть небольшое преимущество. Напротив, максимальное и минимальное значения различаются значительно. Причиной этого является соответствующее минимальное расстояние между кодовыми словами $d_{\min}$ для различных конструкций кодовых книг. Это значение $d_{\min}=0,95$ для кодовой книги Аллтопа, $d_{\min}=0,7$ для RVQ и $d_{\min}=0,16$ для NR-DFT, ко-

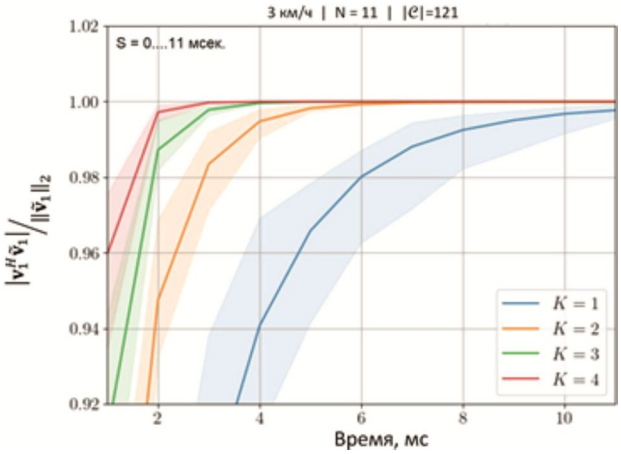


Рис. 5. Сходимость алгоритма ДАП для разных K, начальный участок времени

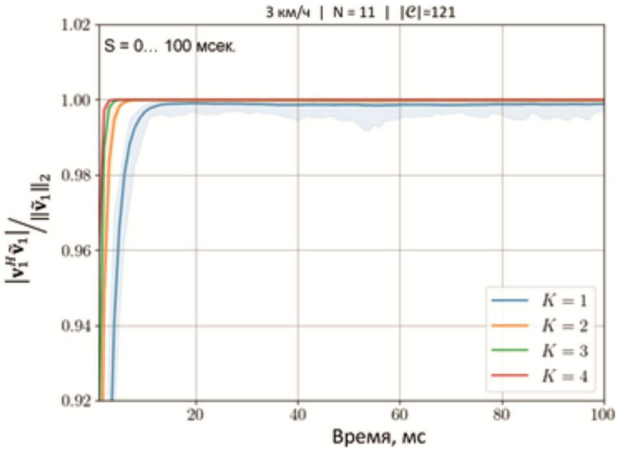
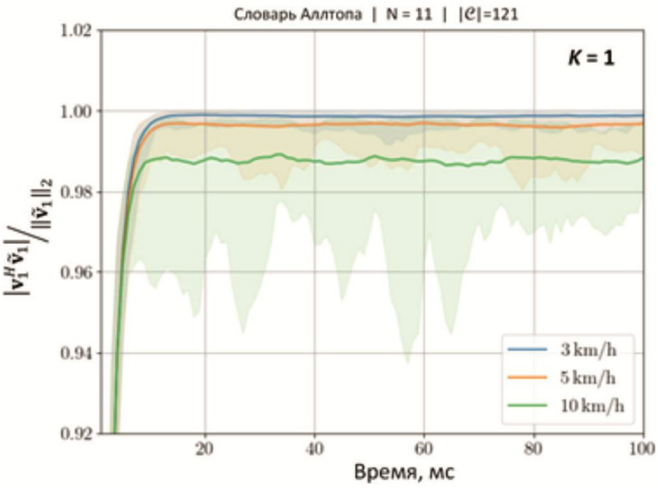
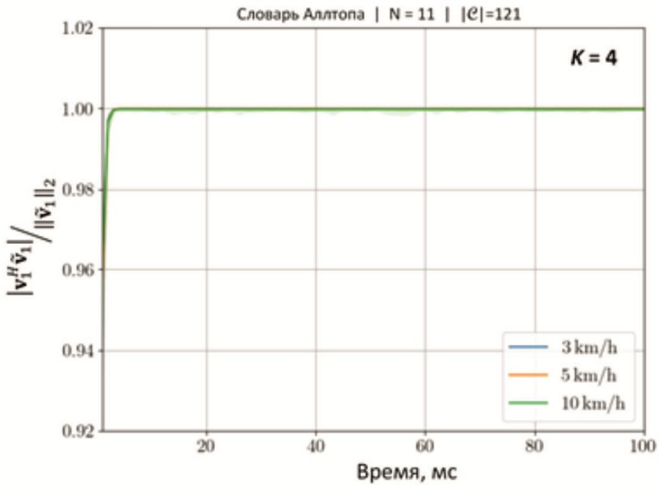


Рис. 6. Сходимость алгоритма ДАП для разных K

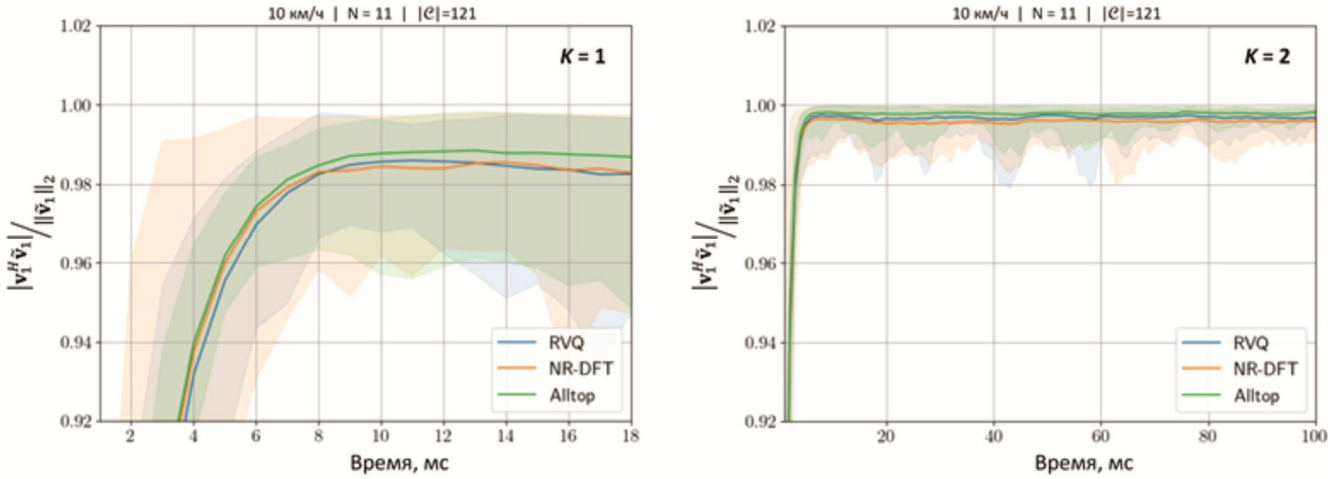


а) $K=1$



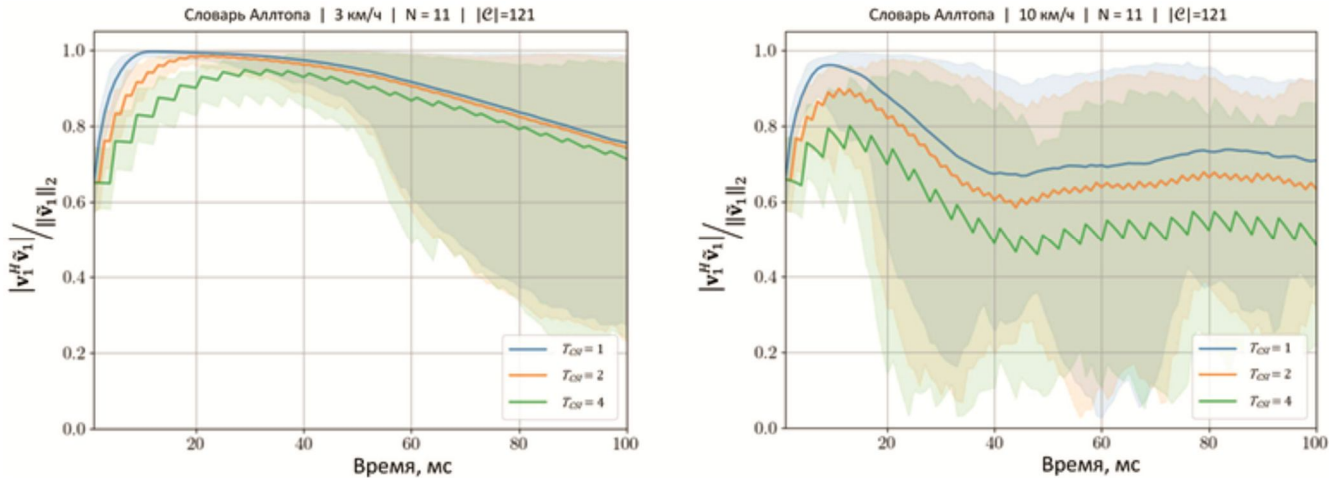
б) $K=4$

Рис. 7. Нормированная корреляция во времени с использованием алгоритма ДАП и кодовой книги Alltop для различных значений скорости. Сплошными линиями обозначено усредненное значение по 30 реализациям



а) $K=1$ *б)* $K=2$

Рис. 8. Нормированная корреляция во времени с использованием алгоритма ДАП для различных словарей и различных значений K*. Сплошные линии обозначают среднее значение по 30 реализациям*



а) скорость МС 3 км/ч *б) скорость МС 10 км/ч*

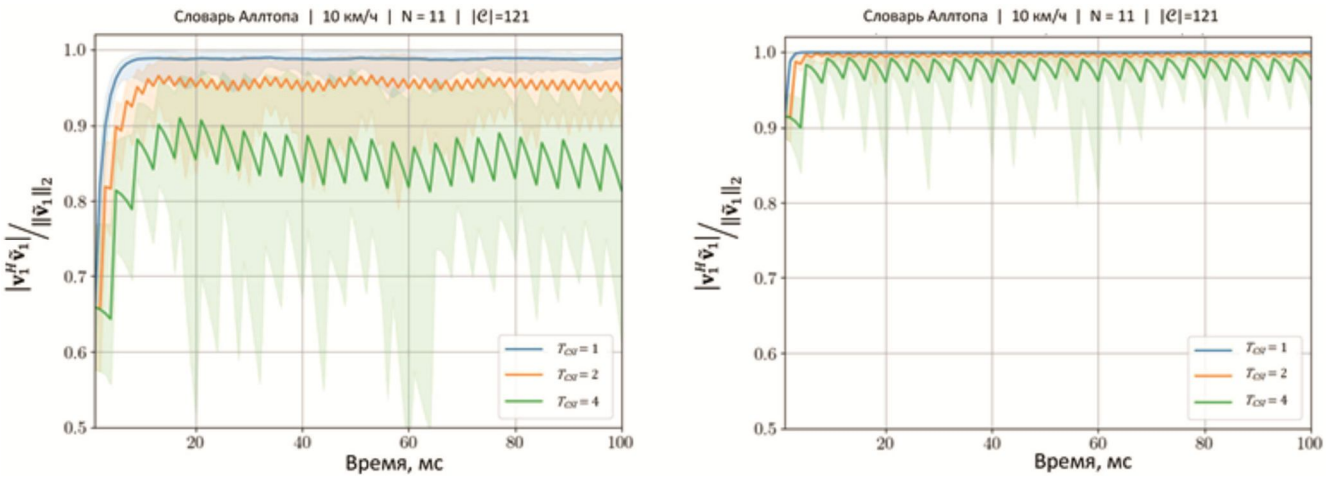
Рис. 9. Нормализованная корреляция во времени с использованием алгоритма MP и кодовой книги Аллтопа. Сплошными линиями обозначено среднее значение за 30 реализаций

торый используется в 3GPP NR версии 16. Кодовые слова NR-DFT неравномерно распределены по единичной сфере и имеют наименьшее минимальное расстояние между кодовыми словами. Следовательно, некоторые реализации соответствуют друг другу очень хорошо и достигают высокой корреляции, в то время как для других это не так.

Вычислительный эксперимент и сравнительный анализ

Ниже представлены различные результаты по ранее представленным алгоритмам. За основу будут взяты два алгоритма: алгоритм жадного поиска и ДАП. В основе каждого алгоритма лежат итерации OMP, которые были рассмотрены ранее в статье. Алгоритм жадного поиска (Matching Pursuit, MP) известен с 1970-х годов. Одна из первых работ по обработке сигналов с использованием MP алгоритма [12] использует словарь Габборовских функций. В данной статье используется словарь Аллтопа (Alltop) и словарь ДПФ высокого разрешения (O-DFT) для сравнения результатов.

Необходимо проанализировать устойчивость алгоритмов к периоду измерений. На всех графиках выше предполагалось, что оценка и квантование канала выполняются каждую миллисекунду. На самом деле в протоколе физического уровня есть ограничения на передачу канальной информации в канале обратной связи. Графики на рис. 9 показывают, как продолжительность отчетных интервалов T_{CSI} влияет на корреляцию во времени для различных скоростей {3, 10} км/ч и отчетных интервалов $T_{CSI}=\{1,\ 2,4\}$ мс. $T_{CSI}>1$ мс показывает пилообразное поведение нормированной корреляции на всех скоростях. Это происходит из-за смещения векторов канала, которое снижает корреляцию, когда канальная информация не передается (канал устаревает). Следовательно, следующая обратная связь должна компенсировать смещение в дополнение к остаточному квантованию. Соответственно, нормированная корреляция для более длинных интервалов обратной связи ниже, чем для более коротких. Для $T_{CSI}=4$, значение корреляции, равное единице, не может быть достигнуто. Таким образом, алгоритм обратной связи MP требует более коротких интервалов обратной связи. При более высоких скоростях дрейф векторов канала влияет на производительность, см. рис. 10, *б*. В результате MP не может компенсировать изменения канала, и набор ба-



а) $K=1$ *б)* $K=3$

Рис. 10. Нормализованная корреляция во времени с использованием алгоритма обратной связи ДАП и кодовой книги Аллтопа. Сплошными линиями обозначено среднее значение за 30 реализаций

зисных векторов словаря необходимо обновить.

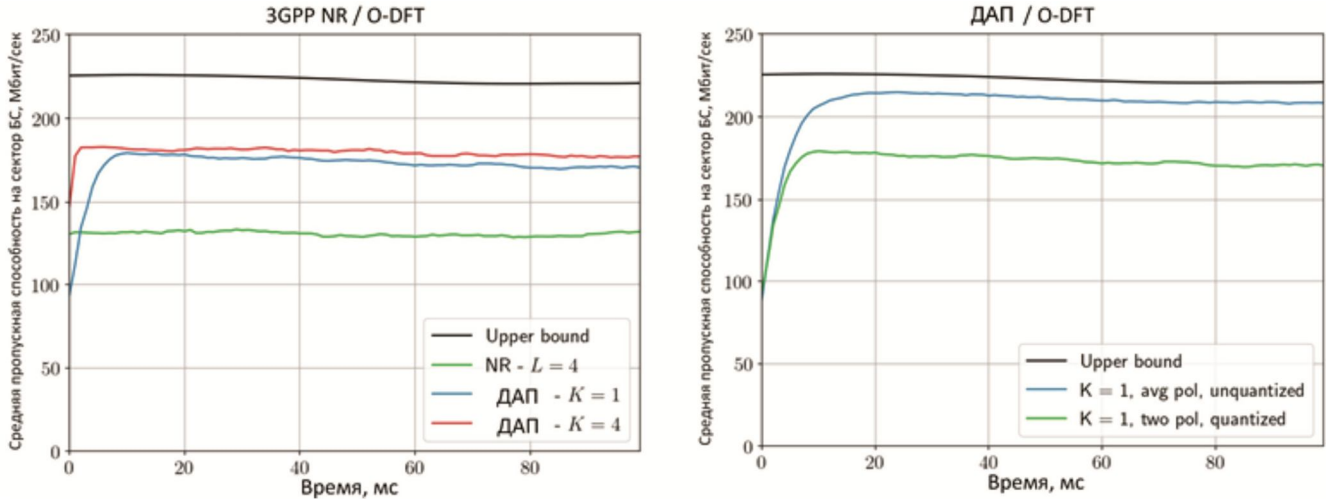
На рис. 10 показаны аналогичные результаты для ДАП алгоритма. Отдельные графики отличаются значением $K=\{1,3\}$. Как и в случае с МР, ДАП также демонстрирует пилообразное поведение при $T_{CSI}>1$ мс. Однако со временем производительность не снижается, поскольку базисные вектора словаря могут изменяться. Кроме того, можно видеть, что повторная передача нескольких кодовых слов на одной итерации может увеличивать или уменьшать значение корреляции. Представленные здесь результаты на скорости 10 км/ч позволяют предположить, что для обеспечения мобильности необходим ДАП более высокого порядка $K>1$. В дальнейшем эффективность алгоритмов обратной связи, основанная на величине корреляции, рассматриваться не будет. Для выбранных конфигураций будет выполнено полноценное системное моделирование с параметрами из табл. 1.

Рассматривается пространственное мультиплексирование 4 МС с 1 антенной, скорость движения 3 км/ч. На рис. 11, *а*, сравниваются характеристики метода обратной связи 3GPP NR на основе O-DFT с алгоритмом ДАП в петле OC-CSI, который также использует кодовую книгу O-DFT. В каждом случае предположим, что размер кодовой книги равен $|\mathcal{C}|=64$. Верхняя граница постоянна и составляет 225 Мбит/с. Метод OC-CSI согласно 3GPP NR версии 16 достигает скорости 130 Мбит/с, которая практически не изменяется во времени. Если метод OC-CSI основан на подходе ДАП, то производительность повышается до 175 Мбит/с после некоторого времени адаптации алгоритма. Кроме того, показаны кривые для $K=1$ и $K=4$. Увеличение числа итераций OMP и, следовательно, количество кодовых слов и коэффициентов в одном отчете вряд ли улучшат производительность. Ключевым моментом для этого является не количество кодовых слов, а сжатие информации в канале обратной связи, что показано на рис. 11, *б*. На рисунке представлены характеристики для неквантованных и квантованных OC-CSI с использованием ДАП. В первом случае используется среднее значение для двух поляризаций, чтобы найти кодовые слова, усредненные по полосе. Таким образом, коэффициенты поддиапазона для этой средней поляризации вычисляются количественно, и в конечном итоге обратная связь используется для обеих поляризаций в БС. В другой реализации ДАП среднее значение двух поляризаций также используется для выбора кодовых слов, усредненных по полосе. Однако коэффициенты подполосы определяются отдельно для обеих поляризаций и квантуются аналогично NR. Хорошо видно, что квантование коэффициентов приводит к значительным потерям.

Таблица 1. Основные параметры модельного эксперимента

Модель канала	Quadriga
Сценарий	3D-3GPP 38.901 Uma NLOS outdoor
Несущая частота	1,85 ГГц
Расстояние между поднесущими частотами	15 кГц
Расстояние между БС	500 м
Высота БС	25 м
Конфигурация антенной системы БС	16 колонок по 4 элемента в каждой колонке, 2 поляризации на каждом элементе. Итого 128 элементов. Расстояние между элементами – по вертикали $0,9\lambda$ – по горизонтали $0,5\lambda$
Наклон антенной системы	наклон панели 3° наклон диаграммы направленности с помощью фазовращателей 7°
Антенные порты	4 антенных элемента в одной колонке объединены в один порт
Минимальное расстояние БС-МС	35 м
Высота МС	1,5 м
МС антенна	Всенаправленная
МС скорость	3 – 10 км/ч
T_{CSI}	1 мс

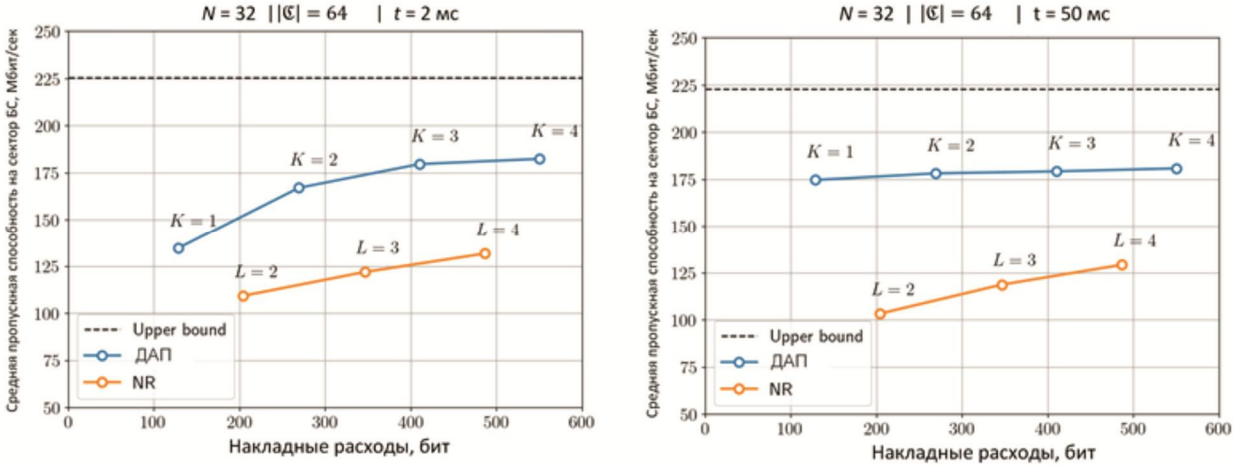
На рис. 12 показана пропускная способность сектора БС в зависимости от параметров конфигурации в



а) Сравнение 3GPP NR и ДАП подходов

б) Сравнение ДАП с неквантованными коэффициентами для среднего значения поляризаций и с квантованными коэффициентами для каждой поляризации в отдельности

Рис. 11. Пропускная способность БС при использовании ДАП алгоритма



а) $t = 2$ *мс*

б) $t = 50$ *мс*

Рис. 12. Сравнение накладных расходов на обратную связь между 3GPP NR и ДАП подходами

момент времени t = 2 мс от начала моделирования (рис. 12б *а*) и для t = 50 мс на рис. 13, *б*. Здесь под верхней границей следует понимать максимальную скорость, которая не зависит от накладных расходов (идеальный канал).

В целом, можно видеть, что с помощью решения ДАП достигается значительное повышение эффективности в соотношении производительность-накладные расходы по сравнению с алгоритмом OC-CSI, использующимся в 3GPP NR стандарте. Соответственно, при заданном объеме доступных накладных расходов более высокая скорость может быть достигнута с помощью ДАП. Кроме того, для алгоритма обратной связи NR в оба момента времени увеличение L приводит к сравнительно равному увеличению производительности. Конечно, для больших значений L требуется больше бит. В случае ДАП это также относится к увеличению K. Однако увеличение K приводит к значительному увеличению производительности только при небольших значениях t, см. также рис. 11, *а*. Следовательно, на этапе запуска ДАП использует меньшее количество кодовых слов, что позволяет поддерживать высокую производительность и при этом значительно сокращать накладные расходы на пересылку информации о состоянии канала.

Заключение

В работе приводится модифицированный алгоритм квантования и обратной связи канальной информации. Предложенный алгоритм показал большую надежность и устойчивость к ошибкам в оценке канала, обеспечил прирост пропускной способности системы на 50 % по сравнению с методами, на которых основан стандарт 3GPP. При этом накладные расходы на передачу канальной информации в восходящем канале сопоставимы с NR-DFT методом из 3GPP.

Алгоритм ДАП показал возможность сглаженного решения в случае резких выбросов в оценках канала, таким образом он неявным образом обозначил лучшую устойчивость к такого рода возмущениям во входных данных. Имея предшествующее знание о канале, можно легко адаптировать алгоритм к динамическим каналам с допплеровскими замираниями сигнала, что позволит построить словарь, адаптированный к динамике в канале, и является предметом дальнейшего исследования.

Литература

1. E. Dahlman, S. Parkvall, and J. Skold. 5G NR: The Next Generation Wireless Access Technology. Elsevier Science, 2018. ISBN: 9780128143230. URL: https://books.google.de/books?id=IcSLswEACAAJ.

2. D. J. Love et al. «An overview of limited feedback in wireless communication systems». In: IEEE Journal on Selected Areas in Communications 26.8 (2008), pp. 1341-1365. DOI: 10.1109/JSAC.2008.081002.

3. Physical layer procedures for data. 3GPP Technical Specification 38.214, ver. 18.2.0, 29 March 2024. https://www.3gpp.org/ftp/Specs/archive/38_series/38.214/38214-i20.zip

4. Arogyaswami Paulraj et al. Introduction to Space-Time Wireless Communications. Cambridge: Cambridge University Press, 2003. ISBN: 978-0-521-82615-0.

5. Gerald Matz and Franz Hlawatsch. Wireless Communications Over Rapidly Time-Varying Channels. Ed. by Franz Hlawatsch and Gerald Matz. Oxford: Academic Press, 2011. ISBN: 978-0-12-374483-8. DOI: https://doi.org/10.1016/C2009-0-18506-9.

6. N. Solomennikova, A. Sherstobitov and V. Lyashev, «Frequency Selective MIMO precoding in Time Domain», 2021 International Conference Engineering and Telecommunication (En&T), Dolgoprudny, Russian Federation, 2021, pp. 1-6, doi: 10.1109/EnT50460.2021.9681749.

7. Chun Kin Au Yeung and D. J. Love. «Performance Analysis of Random Vector Quantization Limited Feedback Beamforming!, Conference Record of the Thirty-Ninth Asilomar Conference on Signals, Systems and Computers, 2005., Pacific Grove, CA, USA, 2005, pp. 408-412, doi: 10.1109/ACSSC.2005.1599778.

8. Foucart S. and Rauhut H. A Mathematical Introduction to Compressive Sensing, in series «Applied and Numerical Harmonic Analysis», Springer, New York, 2013, 625 p.

9. W. Alltop. «Complex sequences with low periodic correlations (Corresp.)», in IEEE Transactions on Information Theory, vol. 26, no. 3, pp. 350-354, May 1980, doi: 10.1109/TIT.1980.1056185.

10. M.A. Herman and T. Strohmer. «High-Resolution Radar via Compressed Sensing», in IEEE Transactions on Signal Processing, vol. 57, no. 6, pp. 2275-2284, June 2009, doi: 10.1109/TSP.2009.2014277.

11. Лившиц Е.Д. «Об эффективности ортогонального жадного алгоритма в задаче о сжатых измерениях», Матем. сб., 203:2, 2012, с. 33-44, DOI: https://doi.org/10.4213/sm7827.

12. S.G. Mallat and Zhifeng Zhang. «Matching pursuits with time-frequency dictionaries», in IEEE Transactions on Signal Processing, vol. 41, no. 12, pp. 3397-3415, Dec. 1993, doi: 10.1109/78.258082.