Повышение эффективности методов кодирования для волоконно‑оптических линий связи.

**Человек:** Статья посвящена актуальным проблемам кодирования в линиях связи в присутствии паттерн‑эффекта – эффекта зависимости частоты ошибок при передаче информации от вида самой информации. Объектом исследования являются методы кодирования информации. В статье приведен общий обзор методов кодирования для линий связи с паттерн-эффектом. Исследуются недостатки метода адаптивного блочного кодирования предложенного Скидиным А. С. и предлагаются способы их устранения. Автор рассматривает применение методов кодирования информации на примере волоконно-оптической линии связи. Проведена симуляция адаптивного блочного кодирования и предложенных улучшений для многоканальной высокоскоростной линии связи с амплитудным методом модуляции сигнала. Был произведен анализ полученных результатов и сделаны выводы о применимости данных кодов. Были разработаны новые методы кодирования на основе адаптивного блочного кодирования предложенного Скидиным А. С. для канала связи с паттерн-эффектом. Для этих методов было проведено сравнительное исследование, показавшее, что новые методы при той же избыточности и том же способе кодирования и декодирования устраняют большее количество ошибок. Проведённая работа по реализации методов кодирования позволяет использовать их не только в предложенной ВОЛС, но и в любой другой, где наблюдается явление паттерн-эффекта, что говорит о широкой применимости данной работы.

**Key words:** волоконно-оптические линии связи, нелинейные эффекты, паттерн-эффект, межсимвольное взаимодействие, четырёхволновое взаимодействие, фазовая кроссмодуляция, теория кодирования, адаптивное блочное кодирование, коды с ограничениями, методы кодирования информации

=================================

**FastText\_KMeans\_Clean:** Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%. Зато для для блока размером 13 разница составляет уже 17%, что показывает преимущество АБКС "среднее" над "чистым" АБКС. На рис.5 мы видим результат этого сравнения. При этом, графики имеют резкие переходы, так как предложенный способ для поиска вероятности того, что среди выбранных для кодирования слов в таблице слово будет начинаться (заканчиваться) с , не является оптимальным. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.5

=================================

**FastText\_KMeans\_Raw/:** Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%. На рис. 4 мы видим, что для АБКС "среднее" с увеличением размера блока ошибок падает количество ошибок. На рис.5 мы видим результат этого сравнения. При этом, графики имеют резкие переходы, так как предложенный способ для поиска вероятности того, что среди выбранных для кодирования слов в таблице слово будет начинаться (заканчиваться) с , не является оптимальным. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.5

=================================

**FastText\_PageRank\_Clean/:** Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Поэтому чем больше размер блока, тем больше разница между результатами. На рис.5 мы видим результат этого сравнения. Для оценки улучшения АБКС "оптимальное" на заданной ЛС была написана его реализация на языке С++. Результаты почти совпадают с результатами АБКС "среднее", но лучше его примерно на 0.3% – 2.5%.

**Key words part:** 0.3125

=================================

**FastText\_PageRank\_Raw/:** Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Поэтому чем больше размер блока, тем больше разница между результатами. На рис.5 мы видим результат этого сравнения. Для оценки улучшения АБКС "оптимальное" на заданной ЛС была написана его реализация на языке С++. Рисунок 9: Сравнение АБКС "оптимальное" и АБКС для длины линии связи 6500 км

**Key words part:** 0.375

=================================

**Mixed\_ML\_TR/:** Для "оптимального" метода возникает следующая проблема — нужно выбрать подмножество слов в таблице (), чтобы среди них считать вероятность того, что слово будет начинаться (заканчиваться) с . Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%. Рис. 4: Коэффициент уменьшения ошибок для АБКС "среднее" при длине линии связи 2500 км. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. Зато для для блока размером 13 разница составляет уже 17%, что показывает преимущество АБКС "среднее" над "чистым" АБКС. Для иллюстрации этого явления, мы выберем для длин линий связи 2500 км, 4500 км и 6500 км АБКС с тем размером блока ошибки, который показал самый лучший результат 11, 11 и 13 бит, соответственно. Преимущество улучшения "среднее" над "чистым" АБКС состоит в том, что оно позволяет использовать блоки ошибок большей длины, устраняя при этом большее количество ошибок. Рис. 6: Разница между результатами АБКС"оптимальное" и АБКС "среднее" для длины линии связи 2500 км.

**Key words part:** 0.4375

=================================

**MultiLingual\_KMeans/:** Для "оптимального" метода возникает следующая проблема — нужно выбрать подмножество слов в таблице (), чтобы среди них считать вероятность того, что слово будет начинаться (заканчиваться) с . Таким образом для выбора этого подмножества мы будем. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. При этом видно, что каждое следующее увеличение блока на два бита дает больший выигрыш, чем для "чистого" АБКС. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. Для иллюстрации этого явления, мы выберем для длин линий связи 2500 км, 4500 км и 6500 км АБКС с тем размером блока ошибки, который показал самый лучший результат 11, 11 и 13 бит, соответственно. Преимущество улучшения "среднее" над "чистым" АБКС состоит в том, что оно позволяет использовать блоки ошибок большей длины, устраняя при этом большее количество ошибок. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.40625

=================================

**Multilingual\_PageRank/:** Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Зато для для блока размером 13 разница составляет уже 17%, что показывает преимущество АБКС "среднее" над "чистым" АБКС. На рис.5 мы видим результат этого сравнения. Для оценки улучшения АБКС "оптимальное" на заданной ЛС была написана его реализация на языке С++. При этом, графики имеют резкие переходы, так как предложенный способ для поиска вероятности того, что среди выбранных для кодирования слов в таблице слово будет начинаться (заканчиваться) с , не является оптимальным. Рис. 8: Сравнение АБКС "оптимальное" и АБКС для длины линии связи 4500 км.

**Key words part:** 0.46875

=================================

**RuBERT\_KMeans\_Without\_ST/:** Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. На рис.5 мы видим результат этого сравнения. Рис. 5: Коэффициент уменьшения ошибок для АБКС "среднее" для линий связи разной длины. Преимущество улучшения "среднее" над "чистым" АБКС состоит в том, что оно позволяет использовать блоки ошибок большей длины, устраняя при этом большее количество ошибок. Рис. 8: Сравнение АБКС "оптимальное" и АБКС для длины линии связи 4500 км.

**Key words part:** 0.375

=================================

**RuBERT\_KMeans\_With\_ST/:** Для "оптимального" метода возникает следующая проблема — нужно выбрать подмножество слов в таблице (), чтобы среди них считать вероятность того, что слово будет начинаться (заканчиваться) с . Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. Для оценки улучшения АБКС "среднее" на заданной ЛС была написана его реализация на языке С++. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%. На рис. 4 мы видим, что для АБКС "среднее" с увеличением размера блока ошибок падает количество ошибок. Зато для для блока размером 13 разница составляет уже 17%, что показывает преимущество АБКС "среднее" над "чистым" АБКС. На рис.5 мы видим результат этого сравнения. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.4375

=================================

**RUBERT\_page\_rank\_Without\_ST/:** Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Поэтому чем больше размер блока, тем больше разница между результатами. На рис.5 мы видим результат этого сравнения. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%.

**Key words part:** 0.34375

=================================

**RUBERT\_page\_rank\_With\_ST/:** Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. На рис.5 мы видим результат этого сравнения. Преимущество улучшения "среднее" над "чистым" АБКС состоит в том, что оно позволяет использовать блоки ошибок большей длины, устраняя при этом большее количество ошибок. Рисунок 9: Сравнение АБКС "оптимальное" и АБКС для длины линии связи 6500 км

**Key words part:** 0.375

=================================

**RUSBERT\_KMeans\_Without\_ST/:** Для "оптимального" метода возникает следующая проблема — нужно выбрать подмножество слов в таблице (), чтобы среди них считать вероятность того, что слово будет начинаться (заканчиваться) с . Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. При этом в подмножестве с лучшим результатом с большей вероятностью там окажутся те слова, которые имеют меньшую вероятность ошибки, если оценивать их по формуле 1. Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. Для оценки улучшения АБКС "оптимальное" на заданной ЛС была написана его реализация на языке С++. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.40625

=================================

**RUSBERT\_KMeans\_With\_ST/:** Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%. Т. к. график АБКС "оптимальное" будет почти совпадать с графиком АБКС "среднее", то мы построим диаграмму разности этих улучшений. На рис. 6 можно увидеть положительную разность, что указывает на то, что АБКС"оптимальное" исправляет больше ошибок, чем АБКС "среднее". Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.40625

=================================

**RUSBERT\_page\_rank\_Without\_ST/:** Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. Таким образом для выбора этого подмножества мы будем. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. На рис.5 мы видим результат этого сравнения. Результаты почти совпадают с результатами АБКС "среднее", но лучше его примерно на 0.3% – 2.5%.

**Key words part:** 0.3125

=================================

**RUSBERT\_page\_rank\_With\_ST/:** Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. Таким образом для выбора этого подмножества мы будем. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. На рис.5 мы видим результат этого сравнения. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.375

=================================

**Simple\_PageRank/:** На рис. 4 мы видим, что для АБКС "среднее" с увеличением размера блока ошибок падает количество ошибок. Зато для для блока размером 13 разница составляет уже 17%, что показывает преимущество АБКС "среднее" над "чистым" АБКС. Для иллюстрации этого явления, мы выберем для длин линий связи 2500 км, 4500 км и 6500 км АБКС с тем размером блока ошибки, который показал самый лучший результат 11, 11 и 13 бит, соответственно. Результаты почти совпадают с результатами АБКС "среднее", но лучше его примерно на 0.3% – 2.5%. На рис. 6 можно увидеть положительную разность, что указывает на то, что АБКС"оптимальное" исправляет больше ошибок, чем АБКС "среднее". При этом, графики имеют резкие переходы, так как предложенный способ для поиска вероятности того, что среди выбранных для кодирования слов в таблице слово будет начинаться (заканчиваться) с , не является оптимальным.

**Key words part:** 0.46875

=================================

**TextRank/:** Рис. 4: Коэффициент уменьшения ошибок для АБКС "среднее" при длине линии связи 2500 км. На рис. 4 мы видим, что для АБКС "среднее" с увеличением размера блока ошибок падает количество ошибок. Для иллюстрации этого явления, мы выберем для длин линий связи 2500 км, 4500 км и 6500 км АБКС с тем размером блока ошибки, который показал самый лучший результат 11, 11 и 13 бит, соответственно. Рис. 5: Коэффициент уменьшения ошибок для АБКС "среднее" для линий связи разной длины. Преимущество улучшения "среднее" над "чистым" АБКС состоит в том, что оно позволяет использовать блоки ошибок большей длины, устраняя при этом большее количество ошибок. Рис. 6: Разница между результатами АБКС"оптимальное" и АБКС "среднее" для длины линии связи 2500 км.

**Key words part:** 0.375

=================================

**TF-IDF\_KMeans/:** Для "оптимального" метода возникает следующая проблема — нужно выбрать подмножество слов в таблице (), чтобы среди них считать вероятность того, что слово будет начинаться (заканчиваться) с . Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%. На рис. 4 мы видим, что для АБКС "среднее" с увеличением размера блока ошибок падает количество ошибок. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС "среднее" - 0.04%. Для оценки улучшения АБКС "оптимальное" на заданной ЛС была написана его реализация на языке С++. Рис. 7: Сравнение АБКС"оптимальное" и АБКС для длины линии связи 2500 км.

**Key words part:** 0.4375

=================================

**Текст:** (1). . (2). Для «оптимального» метода возникает следующая проблема — нужно выбрать подмножество слов в таблице (), чтобы среди них считать вероятность того, что слово будет начинаться (заканчиваться) с . Количество всех подмножеств составит число сочетаний из по , что делает полный перебор всех таблиц для и невозможным. При этом в подмножестве с лучшим результатом с большей вероятностью там окажутся те слова, которые имеют меньшую вероятность ошибки, если оценивать их по формуле 1. Таким образом для выбора этого подмножества мы будем. сначала использовать для построения таблицы АБКС, а потом брать первые слов.. . Для оценки улучшения АБКС «среднее» на заданной ЛС была написана его реализация на языке С++. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%.. . Рис. 4: Коэффициент уменьшения ошибок для АБКС «среднее» при длине линии связи 2500 км. На рис. 4 мы видим, что для АБКС «среднее» с увеличением размера блока ошибок падает количество ошибок. При этом видно, что каждое следующее увеличение блока на два бита дает больший выигрыш, чем для «чистого» АБКС. Это можно объяснить описанной в п. 2.2 разницей в составлении таблицы ошибок — АБКС «среднее» учитывает образованные на границе слов другие подслова. Поэтому чем больше размер блока, тем больше разница между результатами. Например, для блока размером 3 с избыточностью в 5% разница между процентом исправленных ошибок у АБКС и у АБКС «среднее» - 0.04%. Зато для для блока размером 13 разница составляет уже 17%, что показывает преимущество АБКС «среднее» над «чистым» АБКС.. Аналогично АБКС, с ростом длины линии связи АБКС «среднее» исправляет все меньше ошибок. Для иллюстрации этого явления, мы выберем для длин линий связи 2500 км, 4500 км и 6500 км АБКС с тем размером блока ошибки, который показал самый лучший результат 11, 11 и 13 бит, соответственно. На рис.5 мы видим результат этого сравнения.. . Рис. 5: Коэффициент уменьшения ошибок для АБКС «среднее» для линий связи разной длины. Преимущество улучшения «среднее» над «чистым» АБКС состоит в том, что оно позволяет использовать блоки ошибок большей длины, устраняя при этом большее количество ошибок.. Для оценки улучшения АБКС «оптимальное» на заданной ЛС была написана его реализация на языке С++. Мы будем рассматривать этот код с размером блока ошибок от 3 до 13 и избыточностью 4.3% до 100%.. Результаты почти совпадают с результатами АБКС «среднее», но лучше его примерно на 0.3% – 2.5%. Т. к. график АБКС «оптимальное» будет почти совпадать с графиком АБКС «среднее», то мы построим диаграмму разности этих улучшений.. На рис. 6 можно увидеть положительную разность, что указывает на то, что АБКС«оптимальное» исправляет больше ошибок, чем АБКС «среднее». При этом, графики имеют резкие переходы, так как предложенный способ для поиска вероятности того, что среди выбранных для кодирования слов в таблице слово будет начинаться (заканчиваться) с , не является оптимальным.. . Рис. 6: Разница между результатами АБКС«оптимальное» и АБКС «среднее» для длины линии связи 2500 км. . Рис. 7: Сравнение АБКС«оптимальное» и АБКС для длины линии связи 2500 км. . Рис. 8: Сравнение АБКС «оптимальное» и АБКС для длины линии связи 4500 км. . . Рисунок 9: Сравнение АБКС «оптимальное» и АБКС для длины линии связи 6500 км