



# Math-Net.Ru

Общероссийский математический портал

А. Н. Трофимов, Ф. А. Таубин, Анализ эффективности каскадного кодирования для повышения выносливости многоуровневой NAND флеш-памяти, *Информатика и автоматизация*, 2023, выпуск 22, том 2, 316–348

DOI: 10.15622/ia.22.2.4

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 3.143.214.178

1 ноября 2024 г., 13:19:53



А.Н. ТРОФИМОВ, Ф.А. ТАУБИН  
**АНАЛИЗ ЭФФЕКТИВНОСТИ КАСКАДНОГО КОДИРОВАНИЯ  
ДЛЯ ПОВЫШЕНИЯ ВЫНОСЛИВОСТИ МНОГОУРОВНЕВОЙ  
NAND ФЛЕШ-ПАМЯТИ**

*Трофимов А.Н., Таубин Ф.А. Анализ эффективности каскадного кодирования для повышения выносливости многоуровневой NAND флеш-памяти.*

**Аннотация.** Повышение плотности записи в современных чипах NAND флеш-памяти, достигаемое как за счет уменьшающегося физического размера ячейки, так и благодаря возрастающему количеству используемых состояний ячейки, сопровождается снижением надежности хранения данных – вероятности ошибки, выносливости (числа циклов перезаписи) и времени хранения. Стандартным решением, позволяющим повысить надежность хранения данных в многоуровневой флеш-памяти, является введение помехоустойчивого кодирования. Эффективность введения помехоустойчивого кодирования в существенной степени определяется адекватностью модели, формализующей основные процессы, связанные с записью и чтением данных. В работе приводится описание основных искажений, сопровождающих процесс записи/считывания в NAND флеш-памяти, и явный вид плотностей распределения результирующего шума. В качестве аппроксимации полученных плотностей распределения результирующего шума рассматривается модель на основе композиции гауссова распределения и распределения Лапласа, достаточно адекватно отражающая плотности распределения результирующего шума при большом числе циклов перезаписи. Для этой модели проводится анализ помехоустойчивости каскадных кодовых конструкций с внешним кодом Рида-Соломона и внутренним многоуровневым кодом, состоящим из двоичных компонентных кодов. Выполненный анализ позволяет получить обменные соотношения между вероятностью ошибки, плотностью записи и числом циклов перезаписи. Полученные обменные соотношения показывают, что предложенные конструкции позволяют за счет очень незначительного снижения плотности записи обеспечить увеличение граничного значения числа циклов перезаписи (определяемого производителем) в 2–2.5 раза при сохранении требуемого значения вероятности ошибки на бит.

**Ключевые слова:** многоуровневая NAND флеш-память, модель искажений в канале записи/считывания, NL распределение, каскадное кодирование, анализ помехоустойчивости, выносливость.

**1. Введение.** Энергонезависимая память (non-volatile memory, NVM) – класс памяти, который поддерживает сохраненные данные даже после отключения от источника питания. Современная энергонезависимая память обеспечивают более быстрый доступ к данным, пониженное энергопотребление и улучшенную отказоустойчивость по сравнению с обычными жесткими дисками. Как результат, устройства на основе энергонезависимой памяти стали основной заменой жестких дисков для различных новых приложений, связанных с хранением данных, включая аэрокосмические системы, персональную электронику, мобильные вычисления, автономные

транспортные средства, корпоративные хранилища, дата-центры, вычислительные системы с интенсивным использованием данных, медицинские и IoT устройства [1 – 9]. Существует множество типов энергонезависимой памяти, различающихся по своим потребительским свойствам, обусловленным способом построения запоминающих ячеек, и сферам применения. В настоящее время наиболее популярной технологией энергонезависимой памяти является NAND флеш-память, представляющая собой разновидность технологии электрически перепрограммируемой памяти и поддерживающая чтение, программирование и стирание данных в качестве основных операций. Ключевыми достоинствами NAND флеш-памяти являются компактность, дешевизна, механическая прочность, высокая емкость и скорость работы, низкое энергопотребление. Одной из основных целевых областей применения NAND флеш-памяти являются приложения, в которых данные постоянно записываются и стираются. Диапазон таких приложений весьма широк – от потребительской электроники до корпоративных центров обработки данных для облачных приложений и социальных сетей. Эффективное функционирование таких центров требует наличия быстрых и надежных механизмов хранения/обновления громадных объемов данных, и эта потребность удовлетворяется за счет использования флеш-памяти в виде твердотельных накопителей (SSD) [1, 6, 9]. Улучшения в технологии изготовления NAND флеш-памяти за последнее десятилетие привели к снижению удельной стоимости и увеличению плотности записи данных и, таким образом, сделали NAND флеш-память практически безальтернативным решением для ряда современных приложений.

Чип NAND флеш-памяти представляет собой микросборку из нескольких кристаллов. Каждый кристалл, в свою очередь, представляет собой массив ячеек, каждая из которых является полевым транзистором с плавающим затвором. Такая ячейка характеризуется одним из нескольких конечных состояний, определяемых величиной заряда в плавающем затворе. Флеш-память с числом состояний  $q > 2$  называется многоуровневой (multi-level cell, MLC), в отличие от одноуровневой памяти (single-level cell, SLC), где возможен лишь один ненулевой уровень. Наиболее распространенными в настоящее время являются чипы многоуровневой NAND флеш-памяти с 4 и 8 уровнями хранимого электрического заряда в ячейке. На стадии разработки находятся чипы, позволяющие хранить в ячейке 4 бита [10, 11] и 5 бит [12].

Известно, что повышение плотности записи сопровождается как уменьшением максимально возможного числа циклов программирования и стирания (P/E cycling), так и снижением надежности хранения данных [13, 14]. В результате в типичных условиях эксплуатации срок службы чипа многоуровневой флеш-памяти вследствие повышенного износа может составить всего несколько месяцев [15]. Хорошо известно, что в результате повышения плотности записи, характеристики основных параметров многоуровневой NAND флеш-памяти – вероятность ошибки, выносливость (flash cell endurance) и долговечность хранения (data retention), как правило, оказываются неприемлемыми. Так, исходная вероятность ошибки в четырехуровневой флеш-памяти (raw error rate), даже при умеренных значениях числа циклов перезаписи и времени хранения, составляет порядка  $10^{-4} \dots 10^{-3}$  и более [16], тогда как требуемая вероятность ошибки лежит в диапазоне  $10^{-12} \dots 10^{-16}$ . Например, согласно рекомендациям Joint Electron Device Engineering Council (JEDEC), стандартное значение требуемой вероятности ошибки в твердотельных накопителях составляет  $10^{-15}$  для клиентских приложений и  $10^{-16}$  для корпоративных решений [17, 18]. Далее, твердотельные накопители на базе SLC флеш-памяти могли выдерживать 150 000 циклов программирования и стирания, тогда как современные твердотельные накопители с четырехуровневой флеш-памятью на основе техпроцесса 1x (т.е. 15–19 нм) могут выдержать только 3000 циклов программирования и стирания [19, 20]. При этом требуемая величина выносливости, т.е. допустимого числа циклов перезаписи, может достигать значения  $10^5$ , а требуемое время хранения данных составляет до 10 лет. В результате, производители и пользователи многоуровневой флеш-памяти (и в частности, твердотельных накопителей) сталкиваются с трудным выбором в соотношении цены, производительности, емкости и надежности, включающем в качестве важного компонента обеспечение увеличения выносливости при сохранении требуемой вероятности ошибки.

Важным фактором, влияющим на выносливость флеш-памяти, является то, что конструктивно чип NAND флеш-памяти состоит, как правило, из нескольких десятков расположенных друг над другом слоев двумерных массивов ячеек, называемых флеш-блоками. Флеш-блок есть своего рода неделимая единица при выполнении операции стирания (block granularity), т.е. в нулевое состояние переводятся одновременно все ячейки блока. Каждый флеш-блок делится на страницы, которые являются минимальными (по размеру) адресуемыми элементами при выполнении операций записи и/или

чтения (page granularity). В результате, при перезаписи некоторой страницы происходит ненужное (unneded) стирание остальных страниц флеш-блока (с последующим восстановлением стертых данных). Этот эффект – write amplification, который количественно определяется как отношение числа физически выполненных операций записи к числу логически выполненных записей и может достигать величин порядка 3...4, приводит, очевидно, к снижению производительности памяти и её выносливости. Многообещающим инструментом, позволяющим существенно снизить число «паразитных» циклов перезаписи, является использование перезаписываемых (re-write) кодов, и, в частности, WOM кодов [21, 22, 23]. WOM коды характеризуются тем, что при перезаписи только увеличивают уровень заряда в ячейке; в результате страница может быть перезаписана несколько раз (по меньшей мере дважды) перед тем, как потребуются стереть флеш-блок. Вместе с тем, известные WOM коды не позволяют обеспечить требуемую вероятность ошибки порядка  $10^{-12}...10^{-16}$ , а потенциально перспективное объединение перезаписываемых кодов и кодов, исправляющих ошибки, пока находится в начальной стадии исследования [24, 25].

Поэтому в настоящее время решением, позволяющим улучшить потребительские характеристики многоуровневой флеш-памяти, является введение помехоустойчивого кодирования [26 – 33]. Одна из основных проблем, возникающих при реализации помехоустойчивого кодирования в NAND флеш-памяти, связана с необходимостью учета наличия нескольких источников шума, характеристики которых изменяются при изменении как количества выполненных циклов программирования/стирания, так и времени хранения данных в памяти. Кроме того, параметры источников шума зависят от данных, так как увеличение величины заряда в ячейке сопровождается потенциально более интенсивным процессом утечки заряда. Таким образом, эффективность введения помехоустойчивого кодирования в существенной степени определяется адекватностью модели, формализующей основные процессы, связанные с записью и чтением данных. В разделе 2 приводится описание основных искажений, сопровождающих процесс записи/считывания в NAND флеш-памяти, и явный вид плотностей распределения результирующего шума. В качестве аппроксимации полученных плотностей распределения результирующего шума, рассматривается модель на основе композиции гауссова распределения и распределения Лапласа, достаточно адекватно отражающая

плотности распределения результирующего шума при большом числе циклов перезаписи. Применительно к этой модели искажений сигнала в NAND флеш-памяти, для ряда каскадных кодовых конструкций из [30, 31, 33], кратко представленных в разделе 3, в разделе 4 проводится анализ помехоустойчивости рассматриваемых кодовых конструкций, позволяющий получить обменные соотношения между вероятностью ошибки, плотностью записи и числом циклов перезаписи. Полученные обменные соотношения позволяют оценить эффективность использования рассматриваемых вариантов помехоустойчивого кодирования для увеличения выносливости NAND флеш-памяти при поддержании требуемой вероятности ошибки.

Некоторые предварительные результаты данной работы были частично представлены на XXV Международной научной конференции «Волновая электроника и инфокоммуникационные системы», Санкт-Петербург, 2022 [34].

**2. Модель искажений в NAND флеш-памяти.** К основным факторам, определяющим надежность флеш-памяти, обычно относят:

- 1) случайные изменения величины заряда, вводимого в плавающий затвор транзистора в процессе записи (program disturbance);
- 2) повторяющиеся в процессе эксплуатации циклы записи/стирания – P/E cycling;
- 3) взаимную интерференцию ячеек (cell-to-cell interference);
- 4) утечку заряда плавающего затвора с течением времени (retention problem).

Один из подходов, позволяющих формализовать влияние указанных факторов, состоит в следующем [35]. Случайное отклонение реального порогового уровня напряжения от целевого уровня (target value) рассматривается как сумма шумов (случайных величин), порождаемых процессами стирания (erasure noise), записи в ячейку (programming noise), перезаписи (random telegraph noise), старения памяти (retention noise) и взаимной интерференцией ячеек (cell-to-cell interference, CCI). Обозначим упорядоченную по возрастанию последовательность целевых уровней (состояний) напряжения в ячейке как  $x_0, x_1, \dots, x_{q-1}$ , где  $q$  – число уровней записи; наименьший по величине целевой уровень  $x_0$  соответствует состоянию «стерто».

Плотность распределения шума  $\xi_e$ , возникающего при считывании состояния  $x_0$  («стерто»), хорошо аппроксимируется, как показывают многочисленные эксперименты [35, 36], выражением:

$$f_e(\xi_e) = \frac{1}{\sqrt{2\pi\sigma_e^2}} \exp\left(-\frac{\xi_e^2}{2\sigma_e^2}\right), \quad (1)$$

где  $\sigma_e^2$  – дисперсия шума в состоянии «стерто».

Плотность распределения шума  $\xi_p$ , возникающего в процессе записи, полагается равномерной на интервале  $[0, \Delta]$ , где  $\Delta$  – величина шага приращения напряжения (the incremental step pulse programming, ISPP) при записи целевого уровня:

$$f_p(\xi_p) = \frac{1}{\Delta}, \quad 0 \leq \xi_p \leq \Delta. \quad (2)$$

Отметим, что при нахождении ячейки в состоянии «стерто» шум  $\xi_p = 0$ .

Плотность распределения шума  $\xi_t$ , возникающего вследствие перезаписи, хорошо аппроксимируется, согласно многочисленным измерениям [36], стандартным распределением Лапласа с нулевым средним:

$$f_t(\xi_t) = \frac{1}{2\theta} e^{-\frac{|\xi_t|}{\theta}}, \quad (3)$$

где значение параметра  $\theta$  определяются числом циклов записи/стирания  $N$  в виде:  $\theta = K_\theta N^\alpha$ , где  $K_\theta$  и  $\alpha$  – константы, определяемые конкретной технологией изготовления чипа [36].

Плотность распределения шума  $\xi_r$ , возникающего вследствие старения памяти, может быть с разумной точностью аппроксимирована гауссовым распределением [35, 36], параметры которого – среднее значение и дисперсия, зависят от целевого уровня напряжения (состояния) ячейки. Условная плотность распределения шума  $\xi_r$ , возникающего вследствие старения памяти, при считывании уровня  $x_i$ ,  $i = 1, 2, \dots, q-1$ , имеет вид:

$$f_r(\xi_r | x_i) = \frac{1}{\sqrt{2\pi\sigma_r^2(x_i)}} \exp\left(-\frac{(\xi_r - \mu_r(x_i))^2}{2\sigma_r^2(x_i)}\right), \quad (4)$$

где:

$$\mu_r(x_i) = -K_s(x_i - x_0)K_d N^{0.5} \ln\left(1 + \frac{T}{t_0}\right),$$

$$\sigma_r^2(x_i) = K_s(x_i - x_0)K_m N^{0.6} \ln\left(1 + \frac{T}{t_0}\right),$$

$T$  – длительность хранения данных,  $K_s$ ,  $K_d$ ,  $K_m$ ,  $t_0$  – константы, определяемые конкретной технологией изготовления чипа [36]. Отметим, что среднее значение шума  $\xi_r$ : 1) отлично от нуля и отрицательно, что можно интерпретировать как «сжатие» диапазона целевых уровней, 2) различно для различных целевых уровней и 3) возрастает при увеличении целевого уровня. При нахождении ячейки в состоянии  $x_0$  («стерто»), шум  $\xi_r=0$ .

Взаимная интерференция ячеек проявляется в том, величина заряда подвергающейся воздействию интерференции ячейки (victim cell) изменяется при выполнении операций записи или стирания в соседних ячейках (aggressor cells). В частности, при выполнении операции записи в соседних ячейках интерферирующая помеха увеличивает величину заряда в ячейке-жертве, тогда как при выполнении операции стирания (в соседних ячейках) величина заряда в ячейке-жертве уменьшается вследствие интерференции. Для ослабления влияния взаимной интерференции (CCI) используется компенсация интерференционной помехи в ячейке-жертве на основе удаления/добавления корректирующего заряда, величина которого есть линейная комбинация величин зарядов соседних ячеек с весами, определяемыми емкостными связями между ячейками [37]. Плотность распределения остаточного (после компенсации) интерференционного шума  $\xi_{cci}$ , как показывают результаты моделирования в [37], хорошо аппроксимируется «вплоть до хвостов» (down to the tails) гауссовым распределением с нулевым средним. Таким образом, плотность распределения интерференционного шума  $\xi_{cci}$  моделируется как:

$$f_{cci}(\xi_{cci}) = \frac{1}{\sqrt{2\pi\sigma_{cci}^2}} \exp\left(-\frac{\xi_{cci}^2}{2\sigma_{cci}^2}\right), \quad (5)$$



где  $\sigma_{cci}^2$  – дисперсия остаточного (после компенсации) интерференционного шума.

Экспериментальные исследования [38, 39] показывают, что при эффективной компенсации взаимной интерференции в многоуровневой флеш-памяти доминирующим является распределение ошибок, близкое к равномерному на множестве блока ячеек, и при этом не наблюдается заметной тенденции к пакетированию. Последнее обстоятельство может служить аргументом для формального описания рассматриваемого канала записи/считывания блока ячеек флеш-памяти как канала без памяти. Поэтому, в рамках принятого допущения, описание модели сводится к отысканию плотности распределения сигнала, считываемого из ячейки флеш-памяти.

Указанные факторы, определяющие надежность флеш-памяти, характерны, главным образом, для 2D (планарной) NAND флеш-памяти. Архитектура 3D NAND флеш-памяти – трехмерное расположение ячеек чипа, и используемая в 3D NAND технология ловушки заряда (charge trap), т.е. хранение заряда в изолированной области ячейки из непроводящего материала типа нитрида кремния (вместо помещения заряда в плавающий затвор), позволяют существенно повысить плотность записи и снизить вероятность утечки заряда, но при этом порождают ряд дополнительных особенностей. К ним относятся, прежде всего: а) появление существенного отличия динамики изменения среднего значения и дисперсии шума  $\xi_r$ , возникающего вследствие старения памяти (early retention loss), и б) возникновение зависимости скорости утечки заряда из ячейки от величин зарядов, хранящихся в соседних ячейках (retention interference) [40, 41, 42]. Моделирование интерференционных явлений между «charge trap»-ячейками и оценки влияния интерференции на надежность памяти оказывается весьма сложной задачей. Поэтому формализация и апробирование модели шума старения в 3D NAND флеш-памяти пока, как представляется, ещё находятся в стадии анализа и накопления экспериментальных данных.

Приведенные выше источники шумов полагаются аддитивными и независимыми, поэтому условные плотности распределений  $p_{y|x_0}(y|x_0)$ ,  $p_{y|x_1}(y|x_1)$ , ...,  $p_{y|x_{q-1}}(y|x_{q-1})$  считываемого из ячейки сигнала  $y$  можно вычислить с использованием свертки распределений (1)–(5) отдельных слагаемых суммарного шума. Если ячейка находится в состоянии  $x_0$  («стерто»), то считываемый из ячейки сигнал:

$$y = x_0 + \xi_e + \xi_t + \xi_{cci}, \quad (6)$$

т.е. суммарный шум равен  $\xi_e + \xi_t + \xi_{cci}$ , поэтому плотность распределения  $p_{y|x_0}(y|x_0)$  считываемого из ячейки сигнала (6) вычисляется с использованием свертки распределений (1), (3) и (5). В результате получаем:

$$p_{y|x_0}(y|x_0) = \frac{1}{2\theta} e^{\frac{\delta_0^2}{2\theta^2}} \left( e^{-\frac{y-x_0}{\theta}} Q\left(\frac{\delta_0}{\theta} - \frac{y-x_0}{\delta_0}\right) + e^{\frac{y-x_0}{\theta}} Q\left(\frac{\delta_0}{\theta} + \frac{y-x_0}{\delta_0}\right) \right), \quad (7)$$

где  $Q(z) = (1/\sqrt{2\pi}) \int_z^\infty \exp(-t^2/2) dt$ ,  $\delta_0^2 = \sigma_e^2 + \sigma_{cci}^2$ . Если ячейка находится в программируемом состоянии  $x_i$ ,  $i = 1, 2, \dots, q-1$ , то считываемый из ячейки сигнал:

$$y = x_i + \xi_p + \xi_t + \xi_r + \xi_{cci}, \quad (8)$$

т.е. суммарный шум равен  $\xi_p + \xi_t + \xi_r + \xi_{cci}$ , поэтому плотность распределения  $p_{y|x_i}(y|x_i)$  считываемого из ячейки сигнала (8) вычисляется с использованием свертки распределений (2) – (5). В результате получаем:

$$\begin{aligned} p_{y|x_i}(y|x_i) = & \frac{1}{2\Delta} \exp\left(\frac{\delta_i^2}{2\theta^2}\right) \times \\ & \left( \exp\left(\frac{y-x_i-\mu_r(x_i)}{\theta}\right) Q\left(\frac{\delta_i}{\theta} + \frac{y-x_i-\mu_r(x_i)}{\delta_i}\right) - \right. \\ & - \exp\left(-\frac{y-x_i-\mu_r(x_i)}{\theta}\right) Q\left(\frac{\delta_i}{\theta} - \frac{y-x_i-\mu_r(x_i)}{\delta_i}\right) \Big) + \\ & + \exp\left(-\frac{y-x_i-\Delta-\mu_r(x_i)}{\theta}\right) Q\left(\frac{\delta_i}{\theta} - \frac{y-x_i-\Delta-\mu_r(x_i)}{\delta_i}\right) - \\ & - \exp\left(\frac{y-x_i-\Delta-\mu_r(x_i)}{\theta}\right) Q\left(\frac{\delta_i}{\theta} + \frac{y-x_i-\Delta-\mu_r(x_i)}{\delta_i}\right) \Big) + \\ & + \frac{1}{\Delta} \left( Q\left(\frac{y-x_i-\Delta-\mu_r(x_i)}{\delta_i}\right) + Q\left(\frac{y-x_i-\mu_r(x_i)}{\delta_i}\right) \right), \end{aligned} \quad (9)$$

где  $\delta_i^2 = \sigma_r^2(x_i) + \sigma_{cci}^2$ .

Приведенные выражения (7) и (9), определяющие в замкнутой форме точное (в рамках принятой модели) аналитическое представление условных плотностей распределения  $p_{y|x_i}(y|x_i)$ ,  $i = 0, 1, \dots, q-1$ , являются весьма громоздкими, что затрудняет их непосредственное использование как при анализе кодовых конструкций, так и при реализации процедур декодирования с использованием мягких решений. Указанное обстоятельство диктует необходимость достаточно простых аппроксимаций для условных плотностей распределения  $p_{y|x_i}(y|x_i)$ ,  $i = 0, 1, \dots, q-1$ , хорошо согласующихся, при этом, с точными выражениями (7) и (9). Так, например, в работе [35] были рассмотрены возможности аппроксимации плотности распределения суммарного (результатирующего) шума гауссовым распределением, бета-распределением, гамма-распределением, логнормальным распределением и распределением Вейбулла. Среди этих распределений для рассмотренных в [35] сценариев работы флеш-памяти с умеренным числом циклов перезаписи гауссово распределение оказалось в большинстве случаев наиболее предпочтительным вариантом аппроксимации.

*Гауссовская аппроксимация.* Простейшей (и наиболее распространенной) аппроксимацией является аппроксимация выражений (7) и (9) гауссовым распределением [35, 38, 43, 44]. При этом средние значения и дисперсии аппроксимирующих гауссовых распределений полностью определяются первым и вторым моментами распределений (1)–(5) отдельных слагаемых суммарного шума. Так, для состояния  $x_0$  имеем: среднее значение равно  $x_0$ , дисперсия составляет  $\sigma_e^2 + 2\theta^2 + \sigma_{cci}^2$ . Для состояния  $x_i$ ,  $i = 1, 2, \dots, q-1$  (программируемое состояние) среднее значение равно  $x_i + \Delta/2 - \mu_r(x_i)$ , дисперсия составляет  $2\theta^2 + \Delta^2/12 + \sigma_r^2(x_i) + \sigma_{cci}^2$ . Таким образом, гауссова аппроксимация условных плотностей распределения сигнала, считываемого из ячейки, имеет следующий вид. Если ячейка находится в состоянии  $x_0$  («стерто»), то:

$$p_{y|x_0}^{(g)}(y|x_0) = \frac{1}{\sqrt{2\pi(\sigma_e^2 + 2\theta^2 + \sigma_{cci}^2)}} \exp\left(-\frac{(y-x_0)^2}{2(\sigma_e^2 + 2\theta^2 + \sigma_{cci}^2)}\right). \quad (10)$$

Если ячейка находится в состоянии  $x_i, i = 1, 2, \dots, q - 1$ , (одном из программируемых состояний), то:

$$p_{y|x_i}^{(g)}(y|x_i) = \frac{1}{\sqrt{2\pi(2\theta^2 + \Delta^2/12 + \sigma_r^2(x_i) + \sigma_{cci}^2)}} \times \exp\left(-\frac{(y - x_i - \Delta/2 + \mu_r(x_i))^2}{2(2\theta^2 + \Delta^2/12 + \sigma_r^2(x_i) + \sigma_{cci}^2)}\right). \quad (11)$$

Отметим, что среднее значение и дисперсия сигнала, считываемого из ячейки, находящейся в одном из программируемых состояний, зависят от этого состояния, поэтому рассматриваемая аппроксимация с условной плотностью распределения, задаваемой соотношениями (10) и (11), представляет собой гауссовскую модель с дисперсией аддитивного шума, зависящей от записанного значения (input-dependent additive Gaussian noise, ID-AGN).

Характеризовать в целом адекватность гауссовской аппроксимации, с учетом факторов, ограничивающие её применение, можно следующим образом. Среди перечисленных выше искажений, возникающих в процессе записи/считывания, искажения  $\xi_p$ , порождаемые при записи в ячейку (programming noise), и искажения  $\xi_t$ , порождаемые накопленными циклами перезаписи (random telegraph noise), имеют распределение, существенно отличающееся от гауссова. При малых значениях шага приращения напряжения  $\Delta$ , что достаточно типично, влияние первого из этих двух видов искажений  $\xi_p$ , не является существенным. Влияние же искажений  $\xi_t$ , вызванных накопленным количеством циклов перезаписи  $N$ , довольно быстро усиливается с ростом  $N$ ; так, дисперсия случайной величины  $\xi_t$  возрастает пропорционально  $N^{2\alpha}$ , где показатель  $\alpha = 0.5 \dots 1$ . Таким образом, можно полагать, что гауссовская аппроксимация вполне приемлема для сценария, при котором влияние искажений  $\xi_t$  (оцениваемое, например, относительным весом дисперсии  $2\theta^2$  в суммарной дисперсии  $2\theta^2 + \Delta^2/12 + \sigma_r^2(x_i) + \sigma_{cci}^2, i = 1, 2, \dots, q - 1$ ) сравнительно невелико. По мере увеличения накопленного количества циклов перезаписи  $N$  (и приближения к предельному значению, заявленному производителем), качество гауссовской аппроксимации плотностей (7) и (9) существенно снижается, что объясняется, главным

образом, появлением асимметрии и «утяжелением хвостов» в плотностях (7) и (9) [45]. Для учета указанных факторов – асимметрии и «утяжеления хвостов» (вследствие большего влияния распределения Лапласа в свертках соответствующих функций плотности вероятностей), в работе [46] было предложено использовать новую модель, представляющую собой композицию гауссова распределения и распределения Лапласа – NL распределение (normal-Laplace distribution).

*Аппроксимация NL распределением.* NL распределение представляет собой свертку гауссова распределения и распределения Лапласа. Это распределение «почти» гауссово в средней области, тогда как на краях области определения NL распределение «почти» совпадает с распределением Лапласа. Асимметрия может быть введена в NL распределение посредством использования в свертке перекошенного, или асимметричного, распределения Лапласа (skew-Laplace distribution).

Общее выражение для плотности NL распределения (с введенной асимметрией) имеет следующий вид. Пусть  $\xi(m, \sigma)$  – гауссова случайная величина с плотностью распределения:

$$f_g(\xi; m, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\xi - m)^2}{2\sigma^2}\right), \quad (12)$$

где  $m$  – среднее значение,  $\sigma^2$  – дисперсия. Пусть  $\zeta(\lambda, \mu)$  – случайная величина с асимметричной плотностью распределения Лапласа:

$$f_{sl}(\zeta; \lambda, \mu) = \begin{cases} \frac{\lambda\mu}{\lambda + \mu} \exp(-\lambda\zeta), & \zeta > 0, \\ \frac{\lambda\mu}{\lambda + \mu} \exp(\mu\zeta), & \zeta \leq 0, \end{cases} \quad (13)$$

где  $\lambda > 0$ ,  $\mu > 0$  – параметры перекошенного распределения Лапласа. Очевидно, что при  $\lambda = \mu$  получаем обычное (симметричное) распределение Лапласа вида (3). Другие частные случаи  $\mu = \infty$  и  $\lambda = \infty$  приводят к экспоненциальным плотностям вида:

$$f_{sl}(\zeta; \lambda, \infty) = \begin{cases} \lambda \exp(-\lambda\zeta), & \zeta > 0, \\ 0, & \zeta \leq 0, \end{cases}$$

и:

$$f_{sl}(\zeta; \infty, \mu) = \begin{cases} 0, & \zeta > 0, \\ \mu \exp(\mu\zeta), & \zeta \leq 0. \end{cases}$$

Пусть  $y = \xi(m, \sigma) + \zeta(\lambda, \mu)$ ; тогда случайная величина  $y$  имеет NL распределение с плотностью:

$$f_{NL}(y; m, \sigma, \lambda, \mu) = \frac{\lambda\mu}{\lambda + \mu} \left( e^{\frac{\lambda(-2y+2m+\lambda\sigma^2)}{2}} Q\left(-\frac{y-m-\lambda\sigma^2}{\sigma}\right) + e^{\frac{\mu(2y-2m+\mu\sigma^2)}{2}} Q\left(\frac{y-m+\mu\sigma^2}{\sigma}\right) \right). \quad (14)$$

В работе [46] предложена два варианта модели четырехуровневой ячейки флеш-памяти, названные «идеальный» и «смешанный». В идеальном варианте выходные значения ячейки памяти заданы следующим равенством:

$$y = \begin{cases} x_0 + \zeta(\lambda_0, \infty), & \text{если было записано значение } x_0, \\ x_1 + \xi(0, \sigma_1) + \zeta(\lambda_1, \mu), & \text{если было записано значение } x_1, \\ x_2 + \xi(0, \sigma_2) + \zeta(\lambda_2, \mu), & \text{если было записано значение } x_2, \\ x_3 + \zeta(\infty, \mu), & \text{если было записано значение } x_3. \end{cases}$$

С учетом того, что  $x_1 + \xi(0, \sigma_1) = \xi(x_1, \sigma_1)$  и  $x_2 + \xi(0, \sigma_2) = \xi(x_2, \sigma_2)$  можно записать, что:

$$y = \begin{cases} x_0 + \zeta(\lambda_0, \infty), & \text{если было записано значение } x_0, \\ \xi(x_1, \sigma_1) + \zeta(\lambda_1, \mu), & \text{если было записано значение } x_1, \\ \xi(x_2, \sigma_2) + \zeta(\lambda_2, \mu), & \text{если было записано значение } x_2, \\ x_3 + \zeta(\infty, \mu), & \text{если было записано значение } x_3, \end{cases} \quad (15)$$

где обозначения  $\xi(m, \sigma)$  и  $\zeta(\lambda, \mu)$  используются для указания случайных величин, распределенных в соответствии с плотностями вероятности, заданными равенствами (12) и (13) соответственно. Из равенств (15) следует, что условные функции плотности вероятности, определяющие «идеальную» модель, заданы как:

$$\begin{aligned} P_{y|x_0}^{(ideal)}(y|x_0) &= f_{sl}(y-x_0; \lambda_0, \infty), \\ P_{y|x_1}^{(ideal)}(y|x_1) &= f_{NL}(y; x_1, \sigma_1, \lambda_1, \mu), \\ P_{y|x_2}^{(ideal)}(y|x_2) &= f_{NL}(y; x_2, \sigma_2, \lambda_2, \mu), \\ P_{y|x_3}^{(ideal)}(y|x_3) &= f_{sl}(y-x_3; \infty, \mu). \end{aligned} \quad (16)$$

В работе [46] указано, что распределение, дающее более точное соответствие с результатами экспериментальных измерений, может быть получено как смесь распределений, задаваемых равенствами (16), вида:

$$\begin{bmatrix} P_{y|x_0}(y|x_0) \\ P_{y|x_1}(y|x_1) \\ P_{y|x_2}(y|x_2) \\ P_{y|x_3}(y|x_3) \end{bmatrix} = \begin{bmatrix} k_{00} & 0 & k_{02} & k_{03} \\ 0 & k_{11} & k_{12} & k_{13} \\ 0 & 0 & k_{22} & k_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} P_{y|x_0}^{(ideal)}(y|x_0) \\ P_{y|x_1}^{(ideal)}(y|x_1) \\ P_{y|x_2}^{(ideal)}(y|x_2) \\ P_{y|x_3}^{(ideal)}(y|x_3) \end{bmatrix}, \quad (17)$$

где все величины  $k_{00}, k_{01}, \dots, k_{33}$  положительны, и кроме того  $k_{00} + k_{02} + k_{03} = 1, k_{11} + k_{12} + k_{13} = 1, k_{22} + k_{23} = 1$  и  $k_{33} = 1$ . Такое описание канала называется смешанной NL (mixed normal-Laplace) моделью. Конкретный вид функций плотности вероятности вида (17) определяется значениями пятнадцати параметров: четырех значений уровней  $x_0, \dots, x_3$ , шести значений параметров плотностей (16)  $\lambda_0, \sigma_1, \lambda_1, \mu, \sigma_2, \lambda_2$  и пяти весовых коэффициентов  $k_{02}, k_{03}, k_{12}, k_{13}, k_{23}$  в правой части равенства (17). В работе [46] указаны значения этих параметров в зависимости от отношения  $N/N_{\text{ном}}$ , где  $N$  – число циклов перезаписи,  $N_{\text{ном}}$  – число циклов перезаписи, гарантируемое производителем устройства памяти, или номинальное число циклов перезаписи. Эти зависимости указаны в таблице 1. Графики условных ф.п.в (17) в линейном и логарифмическом масштабе представлены на рисунках 1 и 2.

Таблица 1. Параметры смешанной NL модели ячейки флеш-памяти

Параметр модели	Функциональная зависимость от $z$ , $z=N/N_{\text{nom}}$	$c_2$	$c_1$	$c_0$
$k_{02}$ $k_{03}$ $k_{12}$ $k_{13}$ $k_{23}$	$\exp(c_1 z + c_0)$		0.87 1.41 1.63 0.73 1.50	-11.89 -19.82 -19.22 -11.67 -17.69
$x_0$ $x_1$ $x_2$ $x_3$	$c_1 z^{c_2} + c_0$	1 0.17 0.26 0.39	1.52 12.96 11.69 6.83	-150.81 61.12 186.86 319.38
$\sigma_1$ $\sigma_2$	$c_0$			14.95 11.62
$\lambda_0$ $\lambda_1$ $\lambda_2$ $\mu$	$c_1 z + c_0$		1.84 0.65 0.53 0.37	14.28 6.36 6.48 5.13

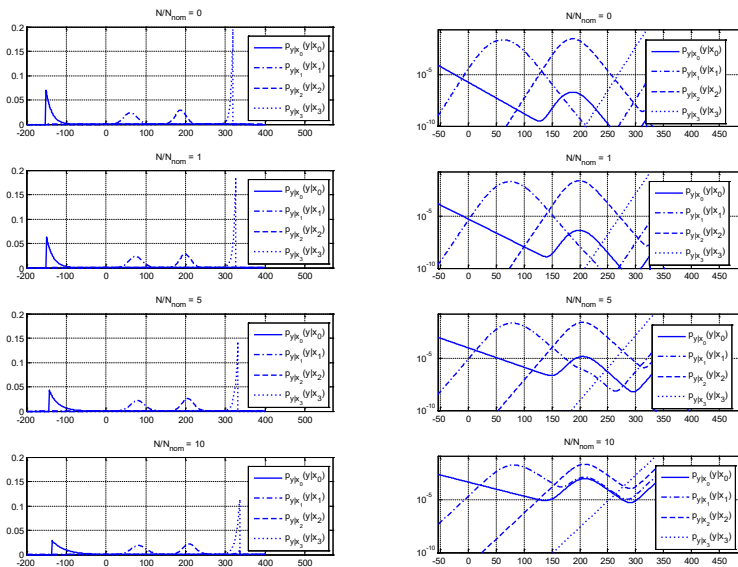


Рис. 1. Графики условных функций плотности вероятности для смешанной NL модели в линейном и логарифмическом масштабах



**3. Краткое описание рассматриваемой каскадной кодовой конструкции.** Один из возможных эффективных подходов к организации введения помехоустойчивого кодирования связан с использованием каскадных конструкций [24]. Основной мотив целесообразности использования каскадной схемы кодирования для многоуровневой флеш-памяти состоит в том, чтобы с помощью подходящего сравнительно простого внутреннего кода обеспечить лучшее согласование мощного внешнего кода с расширенным каналом, включающим кодер и декодер внутреннего кода. Среди возможных вариантов внутреннего кода следует выделить многомерные сигнальные множества на основе целочисленных решеток, обладающие гибкой структурой с широким диапазоном варьирования параметров и допускающие сравнительно простую организацию мягкого декодирования, что позволяет существенно повысить эффективность внешнего кодирования. Каскадное кодирование (с многомерным сигнальным множеством в качестве внутреннего кода) рассматривалось в работах [47, 48]. В этих публикациях, рассматривавших в качестве возможных внешних кодов, соответственно, коды Рида-Соломона и LDPC коды, была продемонстрирована существенная эффективность такого рода каскадных конструкций. Весьма важным достоинством каскадной схемы в плане реализации является гибкость ее архитектуры, что позволяет, в частности, реализовать внутреннюю ступень кода, имеющую небольшую сложность, непосредственно на странице кристалла (чипа) флеш-памяти, содержащей защищаемые данные (on-chip implementation). Расширение класса приемлемых конструкций внутреннего кода с сопутствующим существенным расширением диапазона обменных соотношений «помехоустойчивость – плотность записи – сложность реализации», достигаемое путем использования произвольных mod-4 решеток и многоуровневых конструкций было представлено в работах [25, 26, 28].

Анализируемый далее класс каскадных многокомпонентных кодовых конструкций был предложен в работах [30, 31, 33]. Для удобства и связности изложения приведем далее краткое описание рассматриваемой класса каскадных многокомпонентных кодовых конструкций. В рассматриваемой схеме каскадного кодирования в качестве внешнего кода выбран расширенный (удлиненный на один символ)  $(N_1, K)$  код Рида-Соломона (РС) над полем  $\mathbb{F}_{2^s}$ ,  $2^s = N_1$ . Внутренний код, обозначим его  $B_0$ , представляет собой множество последовательностей, состоящих из  $n$  символов над алфавитом  $A = \{0, 1, \dots, q-1\}$ , где  $q$  – число уровней записи ячейки флеш-памяти.

Элементы алфавита  $A$  связаны с входными уровнями (уровнями записи) ячейки памяти посредством взаимно однозначного отображения  $I$  множества  $A$  на множество  $\{x_0, x_1, \dots, x_{q-1}\}$  вида  $I(i) = x_i$ ,  $i = 0, 1, \dots, q-1$ . Иными словами, алфавит  $A$  представляет собой множество индексов уровней записи  $q$ . Будем полагать далее, что  $m = \log_2 q$  целое число, т.е. ячейка памяти может хранить  $m$  бит данных. Для исключения пакетирования ошибок на внешней ступени декодирования, кодовые символы внешнего кода (при необходимости) подвергаются блоковому перемежению (рисунок 2), а именно:  $h$  последовательных слов кода Рида-Соломона записываются в прямоугольную таблицу, содержащую  $h$  строк, после чего последовательно считываются по столбцам; каждый столбец этой таблицы отображается в один символ внутреннего кода.

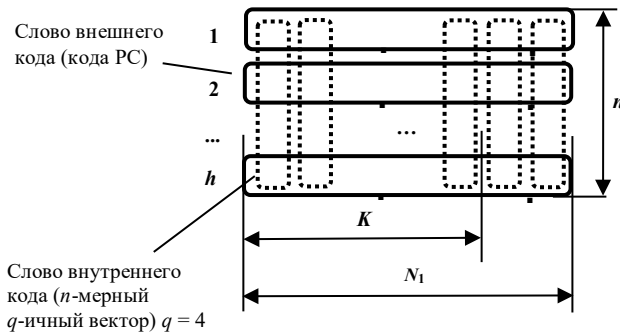


Рис. 2. Структура каскадного кода с перемежением между ступенями

Параметр  $h$  определяется объёмом внутреннего кода, обозначим его  $|B_0|$ , и размером алфавита кода Рида-Соломона, совпадающим с его длиной  $N_1$ , в соответствии со следующим соотношением:  $N_1^h = |B_0|$ . При  $h = 1$  один символ внешнего кода отображается в одно слово внутреннего кода, то есть имеет место обычная каскадная схема кодирования. Обозначим через  $R$  и  $R_0$  скорости внешнего и внутреннего кода, соответственно; очевидно, что  $R = K/N_1$ ,  $R_0 = \log_2 |B_0| / n$ . Общая скорость каскадного кода  $R_{общ}$ , определяющая значение плотности записи в ячейку флеш-памяти, есть произведение  $R$  и  $R_0$ , т.е.  $R_{общ} = RR_0 = K \log_2 |B_0| / (N_1 n)$  бит/ячейка.

Внутренний код  $B_0$  определяется следующей формулой:

$$B_0 = \sum_{j=0}^{L-1} 2^j \psi \left( \left\{ \mathbf{a}_i^{(j)} \mathbf{G}_j \mid 1 \leq i \leq \exp_2(k_j) \right\} \right) + \sum_{j=L}^{m-1} 2^j \psi \left( \left\{ \mathbf{a}_i^{(j)} \mid 1 \leq i \leq \exp_2(n) \right\} \right),$$

где  $\psi$  – естественное вложение ( $\psi$ -вложение)  $\mathbb{F}_2^n$  в  $n$ -мерную целочисленную решетку  $Z^n$ ,  $\left\{ \mathbf{a}_i^{(j)} \mid 1 \leq i \leq \exp_2(k_j) \right\}$  – совокупность двоичных векторов размера  $k_j$ ,  $\left\{ \mathbf{a}_i^{(j)} \mid 1 \leq i \leq \exp_2(n) \right\}$  – совокупность двоичных векторов размера  $n$ ,  $\mathbf{G}_j$  – порождающая матрица линейного двоичного кода  $C_j$ , имеющая размер  $k_j \times n$ ,  $j = 0, 1, \dots, L$ . Скорость внутреннего многоуровневого кода  $R_0 = (m - L) + \sum_{j=0}^{L-1} k_j / n$ .

Компонентные коды  $C_0, C_1, \dots, C_{L-1}$  внутреннего кода  $B_0$  выбираются исходя из обменных соотношений между скоростью  $R_0$ , минимальным евклидовым расстоянием  $\delta_0$  и сложностью декодирования. Отметим, что при выполнении ряда условий, указанных в [30, 31], внутренний код  $B_0$  представляет собой конечное подмножество решетки Барнса-Уолла, что позволяет использовать для декодирования кода  $B_0$  достаточно эффективную (хоть и не оптимальную) процедуру – bounded-distance decoding (BDD). Ограничимся далее рассмотрением двухуровневых ( $L=2$ ) конструкций внутреннего кода для NAND флеш-памяти с четырьмя ранжированными по уровню состояниями.

Примеры каскадных конструкций с внутренним двухуровневым кодом  $B_0$  представлены в таблицах 2 и 3, которые частично повторяют содержимое соответствующих таблиц, приведенных в [30, 31] и [33]. В таблице 2 перечислены варианты каскадного кодирования, в которых внутренний код построен на основе решеток Барнса-Уолла, а в таблице 3 – на основе циклически усеченного сверточного кода (tail-biting code, TB)  $C_0$  и кода с проверкой на четность (single parity check code, SPC)  $C_1$ . Номер варианта в общей нумерации показан в первом столбце таблиц, а выбор конкретного варианта(-ов) из нескольких, если он возможен, выделен полужирным шрифтом. Описание структуры кодов, перечисленных в таблицах 2 и 3, более подробно изложено в [30, 31] и [33] соответственно.

**4. Анализ помехоустойчивости кодовой конструкции для смешанной NL модели.** Приводимый далее анализ помехоустойчивости рассматриваемой кодовой конструкции позволяет получить обменные соотношения между вероятностью ошибки, плотностью записи и числом циклов перезаписи.

Таблица 2. Параметры каскадных конструкций с внутренним кодом на основе решеток Барнса-Уолла

№	Размерность решётки $n$	Исходная решетка $\Lambda_0$	Коды $C_0$ и $C_1$	Объём внутреннего кода $ B_0 $	Возможные значения $h$ и длины кода РС $N_1$	Плотность записи, $R_{общ}$ , бит/ячейка
1,2	8	$E_8$	$C_0 = (8,4,4)$ $C_1 = (8,8,1)$	$2^{12}$	$N_1 = 16, h = 3$ $N_1 = 64, h = 2$ $N_1 = 4096, h=1$	1.5 R
3	8	$RE_8$	$C_0 = (8,1,8)$ $C_1 = (8,7,2)$	$2^8$	$N_1 = 256, h = 1$	1.0 R
4	16	$\Lambda_{16}$	$C_0 = (16,5,8)$ $C_1 = (16,15,2)$	$2^{20}$	$N_1 = 16, h = 5$ $N_1 = 32, h = 4$ $N_1 = 1024, h=2$	1.25 R

Таблица 3. Параметры каскадных конструкций с внутренним двухуровневым кодом  $B_0$  на основе циклически усечённого сверточного кода  $C_0$  и кода с проверкой на четность  $C_1$

№	Коды $C_0$ и $C_1$	Объём внутреннего кода $ B_0 $	Возможные значения $h$ и длин кода РС $N_1$	Плотность записи, $R_{общ}$ , бит/ячейка
5	$C_0 = (8,4,4)$ $C_1 = (8,7,2)$	$2^{11}$	$N_1=2048, h = 1$	1.375 R
6	$C_0 = (12,3,6)$ $C_1 = (12,11,2)$	$2^{14}$	$N_1 = 128, h = 2$	1.167 R
7	$C_0 = (18,9,6)$ $C_1 = (18,17,2)$	$2^{26}$	$N_1=8192, h = 2$	1.444 R

Полученные численным образом обменные соотношения позволяют оценить эффективность использования рассматриваемых вариантов помехоустойчивого кодирования для увеличения выносливости NAND флеш-памяти при поддержании требуемой вероятности ошибки. Оценка помехоустойчивости представленных каскадных кодовых конструкций включает в себя два этапа. Первый этап состоит в вычислении (оценке) вероятности ошибки декодирования по максимуму правдоподобия (МП) слова внутреннего кода. На втором этапе вычисляется (оценивается) вероятность ошибки декодирования слова внешнего кода (кода Рида-Соломона) с использованием результатов, полученных на первом этапе. Стандартная оценка вероятности ошибки на бит при декодировании недвоичного внешнего кода в канале с независимыми ошибками может быть записана как [49]:

$$P_b \leq \frac{1}{2} \sum_{i=t+1}^{N_1} \frac{i+t}{N_1} C_{N_1}^i P_e^i (1-p_e)^{N_1-i}, \quad (18)$$

где  $p_e$  – вероятность ошибки декодирования символа внутреннего кода по МП,  $N_1$  – длина внешнего кода,  $t$  – максимальное число ошибок, исправляемых кодом. Основная проблема при вычислении правой части оценки (18) состоит в вычислении аддитивной оценки вероятности ошибки декодирования внутреннего кода  $p_e$ . Она решается с использованием ранее разработанного подхода [30, 31, 50]. Итоговое выражение для верхней границы вероятности  $p_e$  при этом получается в виде:

$$p_e \leq \frac{1}{\pi M} \int_0^{\infty} \operatorname{Re} \frac{D(\alpha - j\beta) - M}{\beta + j\alpha} d\alpha, \quad (19)$$

где  $M$  – число слов внутреннего кода, и:

$$D(\omega) = \sum_x \sum_{x'} \prod_{l=1}^n c_z(\omega; x^{(l)}, x'^{(l)}). \quad (20)$$

Значение параметра  $\beta$  в правой части (19) выбирается в некоторых допустимых пределах; хорошим выбором для многих примеров может служить значение  $\beta = 1/2$  [30, 31, 50]. В выражении (20):

$$c_{z(\omega; x, x')} = \overline{\exp(j\omega z(x, x'))}, \quad (21)$$

где  $z(x, x') = \ln(p_{y|x'}(y|x') / p_{y|x}(y|x))$ , черта сверху здесь обозначает усреднение по распределению, заданному условной плотностью вероятности  $p_{y|x}(y|x)$ , а суммирование по  $x$  и  $x'$  в правой части (20) выполняется по всем словам внутреннего кода. В качестве технических приемов для вычисления границы (19) используются: 1) решетчатое представление внутреннего кода, 2) рассмотрение произведения этой решетки на себя, 3) разметка решетки-произведения с использованием характеристических функций вида (21), 4) получение значений функции  $D(\omega)$  (20), и 5) численное нахождение значения интеграла (19) при значении параметра  $\beta = 1/2$ . Подробное описание процесса вычисления значений функции  $D(\omega)$  дано в работе [50].

Как следует из описания приведенного подхода, для вычисления оценки вероятности  $p_e$  требуется найти значения характеристических функций (21), которые можно выразить как:

$$c_{z(\omega; x, x')} = \overline{\exp(j\omega z(x, x'))} = \int_{-\infty}^{\infty} p_{y|x'}(y|x')^{j\omega} p_{y|x}(y|x)^{1-j\omega} dy, \quad (22)$$

для  $\omega = \alpha - j\beta$ , где  $0 < \alpha < \infty$ ,  $\beta = 1/2$  для различных условных ф.п.в  $p_{y|x}(y|x)$ , задающих модель канала, и всех входных значений канала  $x, x'$ . Ранее [30, 31, 50] этот подход применялся для случаев, когда выражения для характеристических функций (22) удавалось получить аналитически в замкнутом виде. К таким случаям, в частности, относится и ID-AGN модель канала флеш-памяти, которая рассматривалась в работах [30, 31]. Для рассматриваемой в настоящем исследовании смешанной NL модели получить выражение для характеристических функций (22) в замкнутом виде не удастся. Поэтому их вычисление выполнялось численным методом для ряда значений переменной  $\alpha$  при соответствующем выборе пределов изменения этой величины. Набор значений характеристических функций (22), вычисленных для значений аргумента  $\omega = \alpha - j\beta$ ,  $\beta = 1/2$ , далее использовался при вычислениях по формулам (20), (19) и (18).

В качестве примеров применения этой техники оценки характеристик корректирующего кодирования для канала, заданного смешанной NL моделью, были рассмотрены варианты кодовых схем, перечисленных в таблицах 2 и 3. Результаты вычисления границ (19) и (18) для семи вариантов кодовых схем, указанных в таблицах 2 и 3, в зависимости от значений нормированного числа циклов перезаписи  $N/N_{\text{ном}}$  приведены на рисунках 3 – 6.

Приведенные на рисунках 3 – 6 графики в совокупности представляют собой искомые обменные соотношения между вероятностью ошибки  $P_b$ , плотностью записи  $R_{\text{обц}}$  и числом циклов перезаписи  $N$ . Из рассмотрения данных, представленных на рисунках 3 – 6, следует:

1. Для каждой из рассмотренных конструкций, при числе циклов перезаписи  $N$  в диапазоне от  $0.1N_{\text{ном}}$  до  $N_{\text{ном}}$ , незначительное увеличение плотности записи (это увеличение может быть даже несколько тысячных) сопровождается существенным ростом вероятности ошибки (на порядок и более).

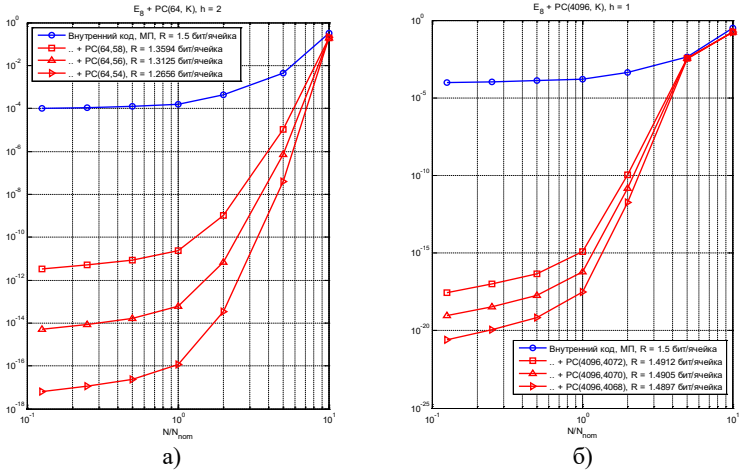


Рис. 3. Зависимость границы вероятности  $P_b$  от нормированного числа циклов перезаписи  $N/N_{\text{ном}}$ : а) кодовая схема № 1, б) кодовая схема № 2

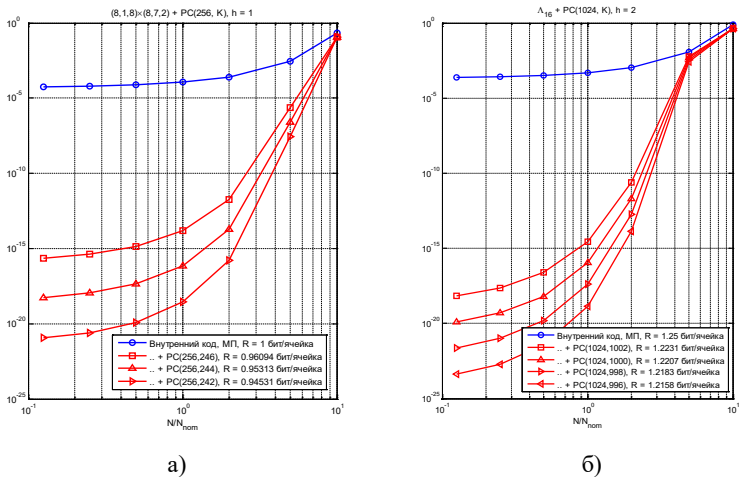


Рис. 4. Зависимость границы вероятности  $P_b$  от нормированного числа циклов перезаписи  $N/N_{\text{ном}}$ : а) кодовая схема № 3, б) кодовая схема № 4

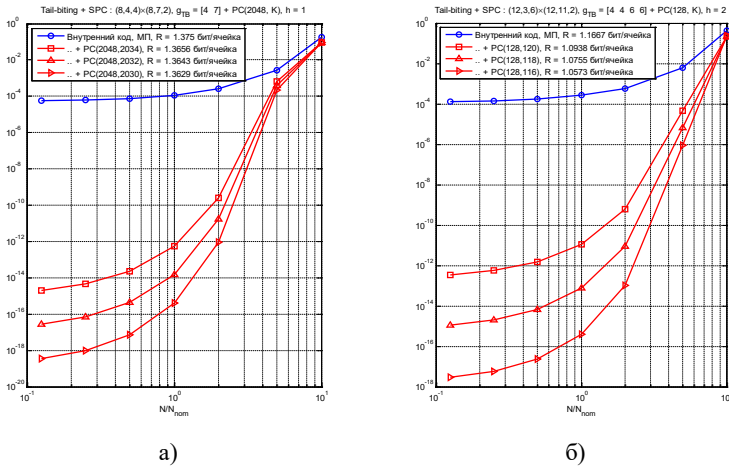


Рис. 5. Зависимость границы вероятности  $P_b$  от нормированного числа циклов перезаписи  $N/N_{nom}$ : а) кодовая схема № 5 (слева), б) кодовая схема № 6

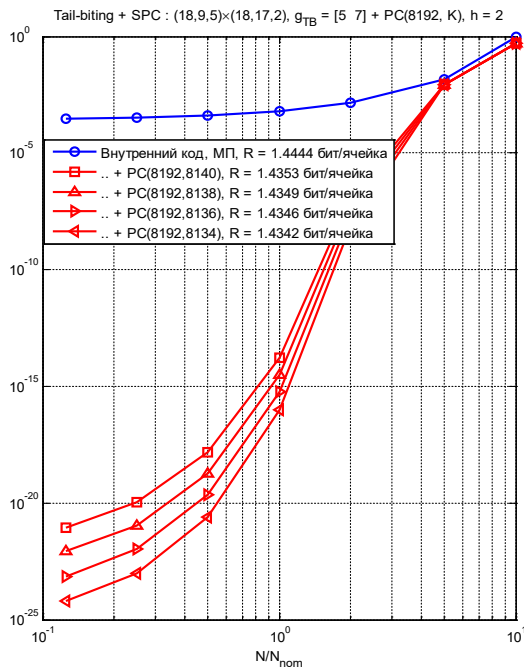


Рис. 6. Зависимость границы вероятности  $P_b$  от нормированного числа циклов перезаписи  $N/N_{nom}$ , кодовая схема № 7



2. Число циклов перезаписи  $N = N_{\text{ном}}$  может действительно рассматриваться, в большинстве случаев, как своего рода граничное значение. Судя по приведенным графикам, увеличение числа циклов перезаписи  $N$  на порядок – от  $0.1N_{\text{ном}}$  до  $1.0N_{\text{ном}}$ , приводит к росту вероятности ошибки на  $1.5...2$  порядка, тогда как дальнейшее увеличение числа циклов перезаписи  $N$  на порядок – до  $10N_{\text{ном}}$ , сопровождается ростом вероятности ошибки на  $12(!)$  и более порядков. Отметим, что при использовании гауссовской аппроксимации зависимость вероятности ошибки от числа циклов перезаписи  $N$  примерно линейна (в логарифмическом масштабе), поэтому выделить такого рода граничное значение в рамках гауссовой модели, видимо, не представляется возможным.

3. За счет очень незначительного снижения плотности записи  $R_{\text{обц}}$  можно обеспечить увеличение допустимого числа циклов перезаписи в  $2-2.5$  раза, относительно величины  $N_{\text{ном}}$ , при сохранении требуемого значения вероятности ошибки на бит. Ясно, что увеличение числа циклов перезаписи приводит к увеличению срока службы при неизменной интенсивности обменов с памятью, либо к увеличению интенсивности обмена данными при сохранении времени эксплуатации системы хранения. В обоих случаях получаем эффект увеличения выносливости флеш-памяти, что обеспечивает заметное улучшение эксплуатационных характеристик систем долговременного хранения данных. Одним из иллюстративных примеров такого обмена «плотность записи – число циклов перезаписи» может служить рисунок 5, из которого следует, что при снижении плотности записи с  $1.2231$  до  $1.2158$  бит/ячейка вероятность ошибки на бит остается в пределах  $\approx 2 \cdot 10^{-15}$  при значении  $N/N_{\text{ном}} = 2$ .

4. Конструкции внутреннего кода вида TB/SPC могут обеспечить внушительный выигрыш (до нескольких порядков) по приемлемому числу циклов перезаписи, относительно кодов на основе решеток Барнса-Уолла, при равной плотности записи и надежности хранения данных. Отметим, что этот выигрыш имеет место как при гауссовской аппроксимации искажений сигнала записи/считывания (это было показано в работе [28]), так и при смешанной NL модели, причем в рамках смешанной NL модели выигрыш может достигать нескольких порядков.

**5. Заключение.** В настоящей работе представлены результаты применения разработанных авторами подходов к оценке эффективности корректирующего кодирования для каналов хранения данных многоуровневой NAND флеш-памяти. Эффективность корректирующего кодирования оценивается через обменное

соотношение, связывающее повышение надежности воспроизведения данных при их передаче или хранении за счет снижения скорости кодирования. Применительно к рассматриваемой модели канала флеш-памяти эти показатели выражаются в повышении числа циклов перезаписи (выносливости ячейки флеш-памяти) за счет снижения плотности записи при сохранении требуемого значения вероятности ошибки. Каскадный принцип построения корректирующего кода позволяет строить длинные коды с высокой корректирующей способностью. При этом достигается высокая корректирующая способность при умеренной сложности реализации, что порой имеет решающее значение при реализации систем кодирования для флеш-памяти.

Отличительной особенностью настоящей публикации является использование смешанной гауссово-лапласовой модели (NL модели) для описания искажений, возникающих при считывании из ячейки NAND флеш-памяти. Такая модель используется наряду с ID-AGN моделью и в некоторых случаях оказывается более предпочтительной, так как достаточно адекватно отражает искажения сигнала записи/считывания для широкого диапазона изменений основных параметров – числа циклов перезаписи и длительности хранения. В качестве схемы кодирования рассмотрена становящаяся все более популярной каскадная схема кодирования. В ней внутренняя ступень представляет собой многоуровневый код, построенный либо на основе решеток Барнса-Уолла, либо на основе циклически усеченного сверточного кода и кода с проверкой на четность, а в качестве внешней ступени используется код Рида-Соломона. Для этой конфигурации подразумевается мягкое декодирование внутреннего кода по МП и алгебраическое декодирование внешнего кода с исправлением сравнительно небольшого числа ошибок. Разработанный ранее подход [30, 31, 50] к анализу помехоустойчивости декодера внутреннего кода был применен в настоящей работе для анализа в условиях принятой смешанной NL модели. В ходе исследования оказалось, что особенности, обусловленные природой используемой модели, не позволяют получить в явном виде выражения для характеристических функций, используемых при вычислении границы вероятности ошибки декодирования внутреннего кода по МП. Это привело к определенным трудностям вычислительного характера, которые были преодолены, и в итоге были получены в численной форме обменные соотношения между вероятностью ошибки, плотностью записи и числом циклов

перезаписи. Полученные обменные соотношения позволили оценить эффективность использования представленных вариантов каскадного кодирования для увеличения выносливости NAND флеш-памяти при поддержании требуемой вероятности ошибки. Среди представленных в разделе 4 конечных результатов можно отметить два результата, важных в прикладном плане. Во-первых, в рамках смешанной NL модели можно достаточно определенно выделить такую важную характеристику, как номинальное число циклов перезаписи, представляющего собой своего рода граничное значение, при превышении которого помехоустойчивость резко ухудшается. Во-вторых, предложенные конструкции позволяют за счет очень незначительного снижения плотности записи обеспечить увеличение этого граничного значения числа циклов перезаписи в 2–2.5 раза при сохранении требуемого значения вероятности ошибки на бит.

### Литература

1. Advances in Non-Volatile Memory and Storage Technology. Second Edition / Eds.: Magyari-Köpe B., Nishi Y. // Amsterdam.: Woodhead Publishing. 2019. 662 p.
2. Gao B. Emerging Non-Volatile Memories for Computation-in-Memory // 25th Asia and South Pacific Design Automation Conference (ASP-DAC). 2020. pp. 381–384. DOI: 10.1109/ASP-DAC47756.2020.9045394.
3. Ishimaru K. Future of Non-Volatile Memory — From Storage to Computing // IEEE International Electron Devices Meeting (IEDM). 2019. pp. 1.3.1–1.3.6. DOI: 10.1109/IEDM19573.2019.8993609.
4. Ishimaru K. Non-Volatile Memory Technology for Data Age // 14th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). 2018. pp. 1–4. DOI: 10.1109/ICSICT.2018.8564815.
5. Gerardin S., Paccagnella A. Present and future non-volatile memories for space // IEEE Transactions on Nuclear Science. 2010. vol. 57. no. 6. pp. 3016–3039. DOI: 10.1109/TNS.2010.2084101.
6. Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using Flash Memory Devices / Eds.: Brewer J., Jill M. // Wiley–IEEE Press. 2008. 792 p.
7. Kang J., Huang P., Han R., Xiang Y., Cui X., Liu X. Flash-based Computing in-Memory Scheme for IOT // Proceedings of the 2019 IEEE 13th International Conference on ASIC (ASICON). 2019. pp. 1–4. DOI: 10.1109/ASICON47005.2019.8983502.
8. Bennett S., Sullivan J. NAND Flash Memory and Its Place in IoT // Proceedings of the 2021 32nd Irish Signals and Systems Conference (ISSC). 2021. pp. 1–6. DOI: 10.1109/ISSC52156.2021.9467859.
9. Aritome S. NAND Flash Memory Technologies // Hoboken.: Wiley. 2016. 432 p.
10. Ohshima S.J. Empowering Next-Generation Applications through FLASH Innovation // Proceedings of the 2020 IEEE Symposium on VLSI Technology 2020. pp. 1–4. DOI: 10.1109/VLSITechnology18217.2020.9265031.
11. Janukowicz J. How New QLC SSDs Will Change the Storage Landscape. IDC White Paper. 2018. Available at: <https://www.micron.com/-/media/client/global/documents/products/white->

- paper/how\_new\_qlc\_ssds\_will\_change\_the\_storage\_landscape.pdf?la=en (accessed: 13.10.2022).
12. Goda A. Recent Progress on 3D NAND Flash Technologies // *Electronics*. 2021. vol. 10. no. 24. pp. 3156. DOI: 10.3390/electronics10243156.
  13. Luo Y., Ghose S., Cai Y., Haratsch E., Mutlu O. Enabling Accurate and Practical Online Flash Channel Modeling for Modern MLC NAND Flash Memory // *IEEE Journal on Selected Areas in Communications*. 2016. vol. 34. no. 9. pp. 2294–2311. DOI: 10.1109/JSAC.2016.2603608.
  14. Liu W. et al., Modeling of Threshold Voltage Distribution in 3D NAND Flash Memory // 2021 Design, Automation & Test in Europe Conference & Exhibition (DATE). 2021. pp. 1729–1732. DOI: 10.23919/DATE51398.2021.9473974.
  15. Grupp L., Davis J., Swanson S. The bleak future of NAND flash memory // *Proceedings of the 10th USENIX Conference on File and Storage Technologies (FAST'12)*. 2012. pp. 2.
  16. Mielke N. et al. Bit error rate in NAND flash memories // *Proceedings of IEEE International Reliability Physics Symposium*. 2008. pp. 9–19.
  17. Liu J., Hsu C., Wang I., Hou T. Categorization of multilevel-cell storage-class memory: an RRAM example // *IEEE Transactions on Electron Devices*. 2015. vol. 62. no. 8. pp. 2510–2516. DOI: 10.1109/TED.2015.2444663.
  18. Solid-State Drive (SSD) Requirements and Endurance Test Method (JESD218) // JEDEC Solid State Technology Association. 2010.
  19. Yoon J., Tressler G. Advanced Flash Technology Status, Scaling Trends & Implications to Enterprise SSD Technology Enablement // *Flash Memory Summit*. 2012.
  20. Maislos A. A New Era in Embedded Flash Memory // *Flash Memory Summit*. 2011.
  21. Fan B., Qin M., Siegel P. Enhancing the Expected Lifetime of NAND Flash by Short q-Ary WOM Codes // *IEEE Communications Letters*. 2018. vol. 22. no. 7. pp. 1302–1305. DOI: 10.1109/LCOMM.2017.2776200.
  22. Chee Y., Kiah H., Vardy A., Yaakobi E. Explicit and Efficient WOM Codes of Finite Length. // *IEEE Transactions on Information Theory*. 2020. vol. 66. no. 5. pp. 2669–2682. DOI: 10.1109/TIT.2019.2946483.
  23. Yaakobi E., Yucovich A., Maor G., Yadgar G. When do WOM codes improve the erasure factor in flash memories? *IEEE International Symposium on Information Theory (ISIT)*. 2015. pp. 2091–2095. DOI: 10.1109/ISIT.2015.7282824.
  24. Jiang A., Li Y., Gad E., Langberg M., Bruck J. Joint rewriting and error correction in write-once memories // *IEEE International Symposium on Information Theory (ISIT)*. 2013. pp. 1067–1071. DOI: 10.1109/ISIT.2013.6620390.
  25. Solomon A., Cassuto Y. Error-Correcting WOM Codes: Concatenation and Joint Design // *IEEE Transactions on Information Theory*. 2019. vol. 65. no. 9. pp. 5529–5546. DOI: 10.1109/TIT.2019.2917519.
  26. Micheloni R., Marelli A., Ravasio R. Error Correction Codes for Non-Volatile Memories // *Springer Science & Business Media*. 2008. 338 p.
  27. Li S., Zhang T. Improving multi-level NAND flash memory storage reliability using concatenated BCH- TCM coding // *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2010. vol. 18. no. 10. pp. 1412–1420. DOI: 10.1109/TVLSI.2009.2024154.
  28. Dong G., Xie N., Zhang T. On the Use of Soft-Decision Error-Correction Codes in NAND Flash Memory // *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2011. vol. 58. no. 2. pp. 429–439. DOI: 10.1109/TCSI.2010.2071990.
  29. Dolecek L., Cassuto Y. Channel coding for nonvolatile memory technologies: Theoretical advances and practical considerations // *Proceedings of the IEEE*. 2017. vol. 105. no. 9. pp. 1705–1724. DOI: 10.1109/JPROC.2017.2694613.

30. Таубин Ф.А., Трофимов А.Н. Каскадное кодирование на основе многомерных решеток и кодов Рида-Соломона для многоуровневой флеш-памяти // Труды СПИИРАН. 2018. Вып. 2(57). С. 75–103. DOI: 10.15622/sp.57.4.
31. Таубин Ф.А., Трофимов А.Н. Каскадное кодирование для многоуровневой флеш-памяти с исправлением ошибок малой кратности во внешней ступени // Труды СПИИРАН. 2019. Вып. 18(5). С. 1149–1181. DOI: 10.15622/sp.2019.18.5.1149-1181.
32. IEEE Std 1890-2018 // IEEE Standard for Error Correction Coding of Flash Memory Using Low-Density Parity Check Codes. 2019. pp. 1–51.
33. Таубин Ф.А., Трофимов А.Н. Каскадное кодирование с внутренним двухуровневым tail-biting/parity check кодом для многоуровневой flash памяти // XXIII международная научная конференция Волновая электроника и инфокоммуникационные системы: Сб. научн. тр. конференции. 2020. С. 354–361.
34. Трофимов А.Н., Таубин Ф.А. Анализ каскадного кодирования для многоуровневой флеш-памяти с использованием смешанной Normal-Laplace модели // XXV международная научная конференция Волновая электроника и инфокоммуникационные системы: Сб. научн. тр. конференции. 2022. С. 109–113.
35. Cai Y., Ghose S., Haratsch E., Luo Y., Mutlu O. Error characterization, mitigation, and recovery in Flash Memory-Based solid-state drives // Proceedings of IEEE. 2017. vol. 105. no. 9. pp. 1666–1704. DOI: 10.1109/JPROC.2017.2713127.
36. Dong G., Pan Y., Xie N., Varanasi C., Zhang T. Estimating information-theoretical NAND flash memory storage capacity and its implication to memory system design space exploration // IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2012. vol. 20. no. 9. pp. 1705–1714. DOI: 10.1109/TVLSI.2011.2160747.
37. Park S., Moon J. Characterization of Inter-Cell Interference in 3D NAND Flash Memory // IEEE Transactions on Circuits and Systems I: Regular Papers. 2021. vol. 68. no. 3. pp. 1183–1192. DOI: 10.1109/TCSI.2020.3047484.
38. Moon J., No J., Lee S., Kim S., Choi S., Song Y. Statistical Characterization of Noise and Interference in NAND Flash Memory // IEEE Transactions on Circuits and Systems I: Regular Papers. 2013. vol. 60. no. 8. pp. 2153–2164. DOI: 10.1109/TCSI.2013.2239116.
39. Wang X., Dong G., Pan L., Zhou R. Error Correction Codes and Signal Processing in Flash Memory. / Ed.: Igor Stievano // IntechOpen. 2011. pp. 57–82. Available at: [www.intechopen.com/books/flash-memories/error-correction-codes-and-signal-processing-in-flash-memory](http://www.intechopen.com/books/flash-memories/error-correction-codes-and-signal-processing-in-flash-memory) (accessed 13.10.2022).
40. Wang K., Du G., Lun Z., Liu X. Investigation of Retention Noise for 3-D TLC NAND Flash Memory // IEEE Journal of the Electron Devices Society. 2019. vol. 7. pp. 150–157. DOI: 10.1109/JEDS.2018.2886359.
41. Luo Y. et al. Improving 3D NAND Flash Memory Lifetime by Tolerating Early Retention Loss and Process Variation // Proceedings of the ACM on Measurement and Analysis of Computing Systems. 2018. vol. 2. no. 3. pp 1–48. DOI: 10.1145/3224432.
42. Liu W. et al. Characterization Summary of Performance, Reliability, and Threshold Voltage Distribution of 3D Charge-Trap NAND Flash Memory // ACM Transactions on Storage. 2022. vol. 18. no. 2. pp 1–25. DOI: 10.1145/3491230.
43. Cai Y., Haratsch E., Mutlu O., Mai K. Threshold voltage distribution in MLC NAND flash memory: Characterization, analysis, and modeling // Proceedings of Design, Automation and Test in Europe Conference. 2013. pp. 1285–1290. DOI: 10.7873/DATE.2013.266.
44. Li Q., Jiang A., Haratsch E. Noise modeling and capacity analysis for NAND flash memories // Proceedings of IEEE International Symposium on Information Theory. 2014. pp. 2262–2266. DOI: 10.1109/ISIT.2014.6875236.

45. Ashrafi R., Arslan S., Pusane A. On the distribution of the threshold voltage in multi-level cell flash memories // *Physical Communication*. 2019. vol. 36. no. 1–2. pp. 1–21. DOI: 10.1016/j.phycom.2019.100747.
46. Parnell T., Papandreou N., Mittelholzer T., Pozidis H. Modelling of the Threshold Voltage Distributions of Sub-20nm NAND Flash Memory // *IEEE Global Communications Conference*. 2014. pp. 2351–2356. DOI: 10.1109/GLOCOM.2014.7037159.
47. Xu Q., Gong P., Chen T.M. Concatenated LDPC-TCM coding for reliable storage in multi-level flash memories // *Proceedings of the 9th International Symposium on Communication System, Networks & Digital Signal Processing (CSNDSP' 2014)*. 2014. pp. 166–170.
48. Kurkoski B.M. Coded modulation using lattices and Reed-Solomon codes, with applications to flash memories // *IEEE Transactions on Selected Areas in Communications*. 2014. vol. 32. no. 5. pp. 900–908. DOI: 10.1109/JSAC.2014.140510.
49. Кларк Дж., Кейн Дж. Кодирование с исправлением ошибок в системах цифровой связи / Под ред. Б.С. Цыбакова // М.: Радио и связь. 1987. 392 с.
50. Трофимов А.Н., Таубин Ф.А. Вычисление аддитивной границы вероятности ошибки декодирования с использованием характеристических функций // *Информационно-управляющие системы*. 2021. № 4. С. 71–85. DOI:10.31799/1684-8853-2021-4-71-85.

**Трофимов Андрей Николаевич** — канд. техн. наук, доцент, институт радиотехники и инфокоммуникационных технологий, кафедра инфокоммуникационных технологий и систем связи, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: теория цифровой связи, теория информации, методы помехоустойчивого кодирования. Число научных публикаций — 66. [andrei.trofimov@vu.spb.ru](mailto:andrei.trofimov@vu.spb.ru); улица Большая Морская, 67, 190000, Санкт-Петербург, Россия; р.т.: +7(812)494-7052.

**Таубин Феликс Александрович** — д-р техн. наук, профессор, институт аэрокосмических приборов и систем, кафедра аэрокосмических компьютерных и программных систем, Санкт-Петербургский государственный университет аэрокосмического приборостроения (СПбГУАП). Область научных интересов: цифровые системы связи, методы помехоустойчивого кодирования, широкополосные системы, беспроводная связь. Число научных публикаций — 102. [ftaubin@yahoo.com](mailto:ftaubin@yahoo.com); улица Большая Морская, 67, 190000, Санкт-Петербург, Россия; р.т.: +7(812)494-7051.

**Поддержка исследований.** Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации, соглашение № FSRF-2020-0004.

A. TROFIMOV, F. TAUBIN  
**PERFORMANCE ANALYSIS OF CONCATENATED CODING TO  
INCREASE THE ENDURANCE OF MULTILEVEL NAND FLASH  
MEMORY**

*Trofimov A., Taubin F. Performance Analysis of Concatenated Coding to Increase the Endurance of Multilevel NAND Flash Memory.*

**Abstract.** The increasing storage density of modern NAND flash memory chips, achieved both due to scaling down the cell size, and due to the increasing number of used cell states, leads to a decrease in data storage reliability, namely, error probability, endurance (number of P/E cycling) and retention time. Error correction codes are often used to improve the reliability of data storage in multilevel flash memory. The effectiveness of using error correction codes is largely determined by the model accuracy that exhibits the basic processes associated with writing and reading data. The paper describes the main sources of disturbances for a flash cell that affect the threshold voltage of the cell in NAND flash memory, and represents an explicit form of the threshold voltage distribution. As an approximation of the obtained threshold voltage distribution, a Normal-Laplace mixture model was shown to be a good fit in multilevel flash memories for a large number of rewriting cycles. For this model, a performance analysis of the concatenated coding scheme with an outer Reed-Solomon code and an inner multilevel code consisting of binary component codes is carried out. The performed analysis makes it possible to obtain tradeoffs between the error probability, storage density, and the number of P/E cycling. The resulting tradeoffs show that the considered concatenated coding schemes allow, due to a very slight decrease in the storage density, to increase the number of P/E cycling up to 2–2.5 times than their nominal endurance specification while maintaining the required value of the bit error probability.

**Keywords:** multilevel NAND flash memory, threshold voltage distribution, Normal-Laplace mixture model, concatenated coding, performance analysis, endurance.

## References

1. Advances in Non-Volatile Memory and Storage Technology. Second Edition. Eds.: Magyari-Köpe B., Nishi Y. Amsterdam.: Woodhead Publishing. 2019. 662 p.
2. Gao B. Emerging Non-Volatile Memories for Computation-in-Memory. 25th Asia and South Pacific Design Automation Conference (ASP-DAC). 2020. pp. 381–384. DOI: 10.1109/ASP-DAC47756.2020.9045394.
3. Ishimaru K. Future of Non-Volatile Memory – From Storage to Computing. IEEE International Electron Devices Meeting (IEDM). 2019. pp. 1.3.1–1.3.6. DOI: 10.1109/IEDM19573.2019.8993609.
4. Ishimaru K. Non-Volatile Memory Technology for Data Age. 14th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). 2018. pp. 1–4. DOI: 10.1109/ICSICT.2018.8564815.
5. Gerardin S., Paccagnella A. Present and future non-volatile memories for space. IEEE Transactions on Nuclear Science. 2010. vol. 57. no. 6. pp. 3016–3039. DOI: 10.1109/TNS.2010.2084101.
6. Nonvolatile Memory Technologies with Emphasis on Flash: A Comprehensive Guide to Understanding and Using Flash Memory Devices. Eds.: Brewer J., Jill M. Wiley–IEEE Press. 2008. 792 p.
7. Kang J., Huang P., Han R., Xiang Y., Cui X., Liu X. Flash-based Computing in-Memory Scheme for IOT. Proceedings of the 2019 IEEE 13th International

- Conference on ASIC (ASICON). 2019. pp. 1–4. DOI: 10.1109/ASIC47005.2019.8983502.
8. Bennett S., Sullivan J. NAND Flash Memory and Its Place in IoT. Proceedings of the 2021 32nd Irish Signals and Systems Conference (ISSC). 2021. pp. 1–6. DOI: 10.1109/ISSC52156.2021.9467859.
  9. Aritome S. NAND Flash Memory Technologies. Hoboken.: Wiley. 2016. 432 p.
  10. Ohshima S.J. Empowering Next-Generation Applications through FLASH Innovation. Proceedings of the 2020 IEEE Symposium on VLSI Technology 2020. pp. 1–4. DOI: 10.1109/VLSITechnology18217.2020.9265031.
  11. Janukowicz J. How New QLC SSDs Will Change the Storage Landscape. IDC White Paper. 2018. Available at: [https://www.micron.com/-/media/client/global/documents/products/white-paper/how\\_new\\_qlc\\_ssds\\_will\\_change\\_the\\_storage\\_landscape.pdf?la=en](https://www.micron.com/-/media/client/global/documents/products/white-paper/how_new_qlc_ssds_will_change_the_storage_landscape.pdf?la=en) (accessed 13.10.2022).
  12. Goda A. Recent Progress on 3D NAND Flash Technologies. Electronics. 2021. vol. 10. no. 24. pp. 3156. DOI: 10.3390/electronics10243156.
  13. Luo Y., Ghose S., Cai Y., Haratsch E., Mutlu O. Enabling Accurate and Practical Online Flash Channel Modeling for Modern MLC NAND Flash Memory. IEEE Journal on Selected Areas in Communications. 2016. vol. 34. no. 9. pp. 2294–2311. DOI: 10.1109/JSAC.2016.2603608.
  14. Liu W. et al. Modeling of Threshold Voltage Distribution in 3D NAND Flash Memory. Design, Automation & Test in Europe Conference & Exhibition (DATE). 2021. pp. 1729–1732. DOI: 10.23919/DATE51398.2021.9473974.
  15. Grupp L., Davis J., Swanson S. The bleak future of NAND flash memory. Proceedings of the 10th USENIX Conference on File and Storage Technologies (FAST'12). 2012. pp. 2.
  16. Mielke N. et al. Bit error rate in NAND flash memories. Proceedings of IEEE International Reliability Physics Symposium. 2008. pp. 9–19.
  17. Liu J., Hsu C., Wang L., Hou T. Categorization of multilevel-cell storage-class memory: an RRAM example. IEEE Transactions on Electron Devices. 2015. vol. 62. no. 8. pp. 2510–2516. DOI: 10.1109/TED.2015.2444663.
  18. Solid-State Drive (SSD) Requirements and Endurance Test Method (JESD218). JEDEC Solid State Technology Association. 2010.
  19. Yoon J., Tressler G. Advanced Flash Technology Status, Scaling Trends & Implications to Enterprise SSD Technology Enablement. Flash Memory Summit. 2012.
  20. Maislos A. A New Era in Embedded Flash Memory. Flash Memory Summit. 2011.
  21. Fan B., Qin M., Siegel P. Enhancing the Expected Lifetime of NAND Flash by Short q-Ary WOM Codes. IEEE Communications Letters. 2018. vol. 22. no. 7. pp. 1302–1305. DOI: 10.1109/LCOMM.2017.2776200.
  22. Chee Y., Kiah H., Vardy A., Yaakobi E. Explicit and Efficient WOM Codes of Finite Length. IEEE Transactions on Information Theory. 2020. vol. 66. no. 5. pp. 2669–2682. DOI: 10.1109/TIT.2019.2946483.
  23. Yaakobi E., Yuovich A., Maor G., Yadgar G. When do WOM codes improve the erasure factor in flash memories? IEEE International Symposium on Information Theory (ISIT). 2015. pp. 2091–2095. DOI: 10.1109/ISIT.2015.7282824.
  24. Jiang A., Li Y., Gad E., Langberg M., Bruck J. Joint rewriting and error correction in write-once memories. IEEE International Symposium on Information Theory (ISIT). 2013. pp. 1067–1071. DOI: 10.1109/ISIT.2013.6620390.
  25. Solomon A., Cassuto Y. Error-Correcting WOM Codes: Concatenation and Joint Design. IEEE Transactions on Information Theory. 2019. vol. 65. no. 9. pp. 5529–5546. DOI: 10.1109/TIT.2019.2917519.
  26. Micheloni R., Marelli A., Ravasio R. Error Correction Codes for Non-Volatile Memories. Springer Science & Business Media. 2008. 338 p.



27. Li S., Zhang T. Improving multi-level NAND flash memory storage reliability using concatenated BCH-TCM coding. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2010. vol. 18. no. 10. pp. 1412–1420. DOI: 10.1109/TVLSI.2009.2024154.
28. Dong G., Xie N., Zhang T. On the Use of Soft-Decision Error-Correction Codes in NAND Flash Memory. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2011. vol. 58. no. 2. pp. 429–439. DOI: 10.1109/TCSI.2010.2071990.
29. Dolecek L., Cassuto Y. Channel coding for nonvolatile memory technologies: Theoretical advances and practical considerations. *Proceedings of the IEEE*. 2017. vol. 105. no. 9. pp. 1705–1724. DOI: 10.1109/JPROC.2017.2694613.
30. Taubin F.A., Trofimov A.N. [Concatenated Reed–Solomon/lattice coding for multi-level flash memory]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2018. vol. 2(57). pp. 75–103. DOI: 10.15622/sp.57.4. (In Russ.).
31. Taubin F.A., Trofimov A.N. [Concatenated coding for multi-level flash memory with low error correction capabilities in outer stage]. *Trudy SPIIRAN – SPIIRAS Proceedings*. 2019. vol. 18. no. 5. pp. 1149–1181. DOI: 10.15622/sp.2019.18.5.1149-1181. (In Russ.).
32. IEEE Std 1890-2018. *IEEE Standard for Error Correction Coding of Flash Memory Using Low-Density Parity Check Codes*. 2019. pp. 1–51.
33. Taubin F.A., Trofimov A.N. [Concatenated coding with inner two-level TB/SPC code for multi-level flash memory]. *Volnovaja jelektronika i infokommunikacionnye sistemy: Sb. nauchn. tr. XXIII mezhdunarodnoj konf. [XXIII International Conference Wave electronics and infocommunication systems: Collected papers]*. 2020. pp. 354–361. (In Russ.).
34. Trofimov A.N., Taubin F.A. [Analysis of concatenated coding scheme for the multilevel flash memory using normal-Laplace mixture]. *Volnovaja jelektronika i infokommunikacionnye sistemy: Sb. nauchn. tr. XXV mezhdunarodnoj konf. [XXV International Conference Wave electronics and infocommunication systems: Collected papers]*. 2022. pp. 109–113. (In Russ.).
35. Cai Y., Ghose S., Haratsch E., Luo Y., Mutlu O. Error characterization, mitigation, and recovery in Flash Memory-Based solid-state drives. *Proceedings of IEEE*. 2017. vol. 105. no. 9. pp. 1666–1704. DOI: 10.1109/JPROC.2017.2713127.
36. Dong G., Pan Y., Xie N., Varanasi C., Zhang T. Estimating information-theoretical NAND flash memory storage capacity and its implication to memory system design space exploration. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2012. vol. 20. no. 9. pp. 1705–1714. DOI: 10.1109/TVLSI.2011.2160747.
37. Park S., Moon J. Characterization of Inter-Cell Interference in 3D NAND Flash Memory. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2021. vol. 68. no. 3. pp. 1183–1192. DOI: 10.1109/TCSI.2020.3047484.
38. Moon J., No J., Lee S., Kim S., Choi S., Song Y. Statistical Characterization of Noise and Interference in NAND Flash Memory. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2013. vol. 60. no. 8. pp. 2153–2164. DOI: 10.1109/TCSI.2013.2239116.
39. Wang X., Dong G., Pan L., Zhou R. *Error Correction Codes and Signal Processing in Flash Memory*. Ed.: Igor Stievano. IntechOpen. 2011. pp. 57–82. Available at [www.intechopen.com/books/flash-memories/error-correction-codes-and-signal-processing-in-flash-memory](http://www.intechopen.com/books/flash-memories/error-correction-codes-and-signal-processing-in-flash-memory) (accessed 13.10.2022).
40. Wang K., Du G., Lun Z., Liu X. Investigation of Retention Noise for 3-D TLC NAND Flash Memory. *IEEE Journal of the Electron Devices Society*. 2019. vol. 7. pp. 150–157. DOI: 10.1109/JEDS.2018.2886359.
41. Luo Y. et al. Improving 3D NAND Flash Memory Lifetime by Tolerating Early Retention Loss and Process Variation. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*. 2018. vol. 2. no. 3. pp. 1–48. DOI: 10.1145/3224432.

42. Liu W. et al. Characterization Summary of Performance, Reliability, and Threshold Voltage Distribution of 3D Charge-Trap NAND Flash Memory. *ACM Transactions on Storage*. 2022. vol. 18. no. 2. pp 1–25. DOI: 10.1145/3491230.
43. Cai Y., Haratsch E., Mutlu O., Mai K. Threshold voltage distribution in MLC NAND flash memory: Characterization, analysis, and modeling. *Proceedings of Design, Automatin and Test in Europe Conference*. 2013. pp. 1285–1290. DOI: 10.7873/DATE.2013.266.
44. Li Q., Jiang A., Haratsch E. Noise modeling and capacity analysis for NAND flash memories. *Proceedings of IEEE International Symposium on Information Theory*. 2014. pp. 2262–2266. DOI: 10.1109/ISIT.2014.6875236.
45. Ashrafi R., Arslan S., Pusane A. On the distribution of the threshold voltage in multi-level cell flash memories. *Physical Communication*. 2019. vol. 36. no. 1–2. pp. 1–21. DOI: 10.1016/j.phycom.2019.100747.
46. Parnell T., Papandreou N., Mittelholz T., Pozidis H. Modelling of the Threshold Voltage Distributions of Sub-20nm NAND Flash Memory. *IEEE Global Communications Conference*. 2014. pp. 2351–2356. DOI: 10.1109/GLOCOM.2014.7037159.
47. Xu Q., Gong P., Chen T. Concatenated LDPC-TCM coding for reliable storage in multi-level flash memories. *Proceedings of the 9th International Symposium on Communication System, Networks & Digital Signal Processing (CSNDSP' 2014)*. 2014. pp. 166–170.
48. Kurkoski B. Coded modulation using lattices and Reed-Solomon codes, with applications to flash memories. *IEEE Transactions on Selected Areas in Communications*. 2014. vol. 32. no. 5. pp. 900–908. DOI: 10.1109/JSAC.2014.140510.
49. Clark G., Cain J. *Error-Correction Coding for Digital Communications*. Plenum Press, 1982. 432 p. (Russ. ed.: Klark Dzh., Kejn Dzh. Kodirovanie s isprav-leniem oshibok v sistemah cifrovoy svyazi. Moscow, Radio i svyaz' Publ. 1987. 392 p.)
50. Trofimov A.N., Taubin F.A. [Evaluation of the union bound for the decoding error probability using characteristic functions]. *Informatsionno-upravliaushchie sistemy – Information and Control Systems*. 2021. no. 4. pp. 71–85. DOI: 10.31799/1684-8853-2021-4-71-85. (In Russ.).

**Trofimov Andrey** — Ph.D., Associate professor, Department of infocommunication technologies and communication systems, institute of radio engineering and infocommunication technologies, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: communication theory, error-correcting coding, information theory. The number of publications — 66. andrei.trofimov@vu.spb.ru; 67, Bolshaya Morskaya St., 190000, St. Petersburg, Russia; office phone: +7(812)494-7052.

**Taubin Feliks** — Ph.D., Dr.Sci., Professor, Institute of aerospace instruments and systems, department of aerospace computer and software systems, Saint Petersburg State University of Aerospace Instrumentation (SUAI). Research interests: communication theory, error-correcting coding, spread spectrum systems, wireless communication. The number of publications — 102. ftaubin@yahoo.com; 67, Bolshaya Morskaya St., 190000, St. Petersburg, Russia; office phone: +7(812)494-7051.

**Acknowledgements.** This research is supported by the Ministry of Science and Higher Education of the Russian Federation, agreement no FSRF-2020-0004.