СБОЕУСТОЙЧИВАЯ ОБРАТИМАЯ ЛОГИКА

С. И. Гуров, Т. Д. Жукова

e-mail: *sgur@cs.msu.ru*, *zhukova_t@ippm.ru*

XXVI Международная конференция
«Математика. Экономика. Образование»
Пансионат Моряк, пос. Дюрсо
27 мая – 3 июня 2018 г.

План выступления

Эффект Неймана–Ландауэра (1960) —

следствие второго закона термодинамики

в любой вычислительной системе, независимо от её физической реализации, при потере 1 бита информации выделяется теплота в количестве не менее $\varepsilon_0 = kT \ln 2$ *Дж (*k *— постоянная Больцмана,* T *— абсолютная температура)*.

«*Информация осязаема*» (Ландауэр)

В 2012 г. эффект удалось обнаружить на практике.

При комнатной температуре
$\varepsilon_0 \approx 3$ зДж$= 3 \cdot 10^{-21}$ Дж

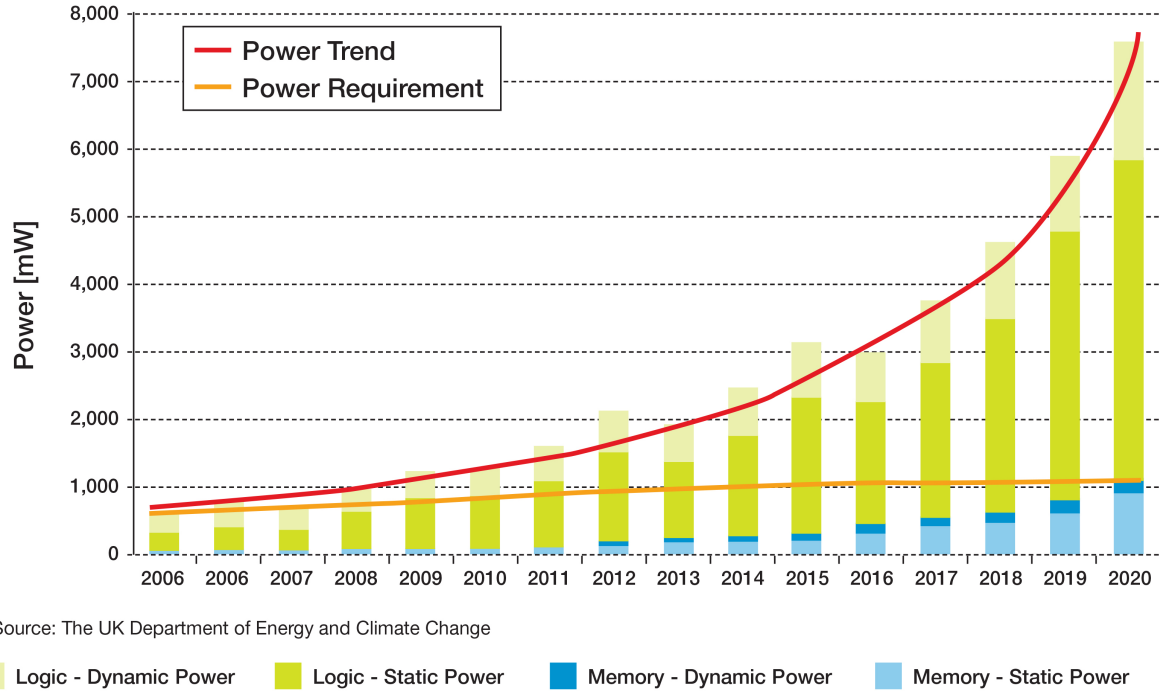
Ничтожное количество?

рассеивание тепла на 1 бит ε_0		
2000-е	2010-е	2020-е
$\approx 10^6$	$\approx 10^3$	≈ 1 ??

В пересчете на процессор в целом суммарная энергия вырастает уже до величин порядка одного 1 Вт.

Закон Мура $\Rightarrow$ экспоненциальное увеличение выделяемой энергии из-за потери информации.

Выделение тепла и технологии



22-нм технологии $\Rightarrow$ выделение тепла $=$ 5...10 МВт на см^2 процессора. Солнце выделяет (всего!) $6{,}5$ кВт/см^2.

«Тепловое проклятием» существующей парадигмы ВТ.

Intel: 2015 г. — 14-нм, 2020 г. — 7-нм, к 2022 г. — 4 нм

λ(зелёный) $= 500...565$ нм, размер вируса $= 20...300$ нм.

Выход: обратимость

Обычные компьютеры — обрабатываемая информация может появляться из ниоткуда, и потом исчезать в никуда.

Обратимость возникает на всех уровнях ВТ:

1. физической реализации вычислений;
2. архитектуры компьютера (процессоры, память, ...);
3. языков и алгоритмов.

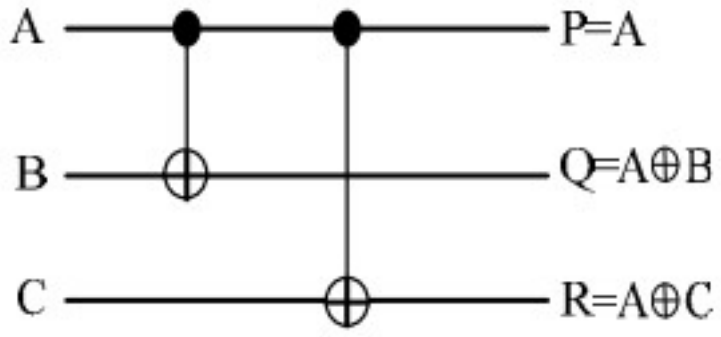
и на всех необходимо поддерживать обратимость.

План выступления

Понятие обратимости

Вычисления *логически обратимы*, если по выходным величинам можно восстановить входные.

Их реализовывают на комбинационных обратимых элементах, осуществляющих некоторую биекцию (перестановку) на множестве всех входных данных
$\Rightarrow$ это $n \times n$ элементы.



Ч. Беннетт (начало 1970-х): любое вычисление, которое можно сделать на обычном необратимом компьютере, можно выполнить так, что оно будет обратимо.

При этом экспоненциально увеличивается число операций и объем используемой памяти, но с этим можно бороться...

Мусорные (стоковые) биты

На выходе обратимого элемента образуются биты

- *информационные* — значений вычисляемой функции;
- *мусорные* — с их помощью можно восстановить значения входных сигналов; если их стереть, произойдет рассеяние энергии, чего требуется избежать.

			«МУСОР»	
A	B	ВЫХОД	C	D
1	1	1	1	0
1	0	0	1	0
0	1	0	0	1
0	0	0	0	0

Память делятся на области:

1. рабочую, содержащую информацию для дальнейших вычислений,
2. вспомогательную, хранящую информация для поддержания обратимости.

Когда в рабочей зоне меняется бит, во вспомогательной происходит противоположный переход.

Простейшие вентили обратимой логики

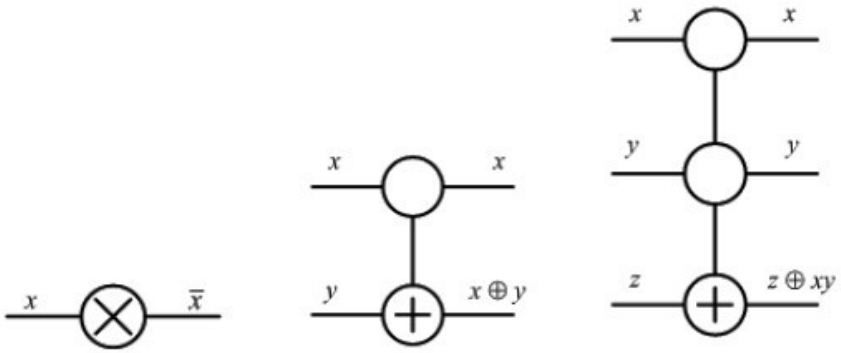


Рис. 1. Элементы NOT, CNOT (Controlled NOT, Фейнмана) и CCNOT (Controlled Controlled NOT, Тоффоли, TG)

Универсальный элемент должен реализовывать любую функционально полную систему БФ и обеспечивать *каскадирование* — ветвление сигналов $\Rightarrow$ возможность организации соединения элементов.

Элементы NOT и CNOT *не универсальны*.

Универсальность. Вентиль Фредкина (FRG)

Элемент Тоффоли универсален: на них можно

- реализовать любую логическую функцию;
- осуществить каскадирование: при входе $(x, 1, 0)$ на выходе будет вектор $(x, 1, x)$.

Вентиль Фредкина (CSWAP, Controlled SWAP, управляемый обмен, FRG) — универсальный обратимый вентиль.

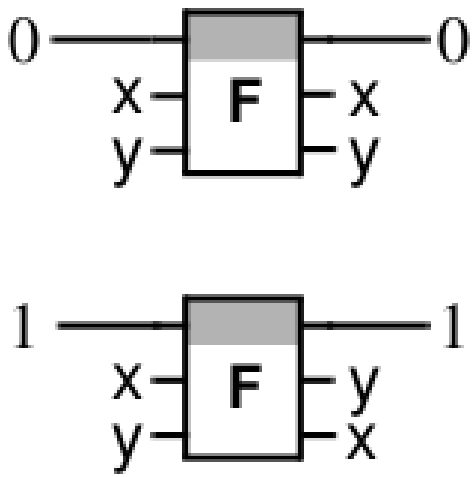
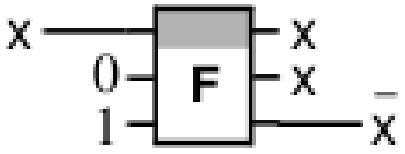
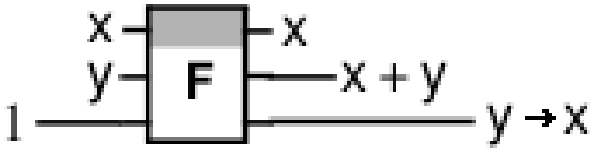


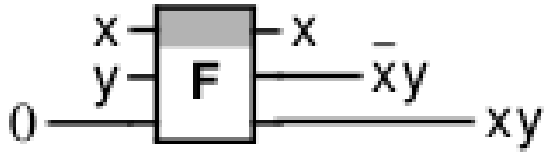
Рис. 2. Изображение элемента Фредкина, 1-й вход — управляющий

Универсальность вентиля Фредкина

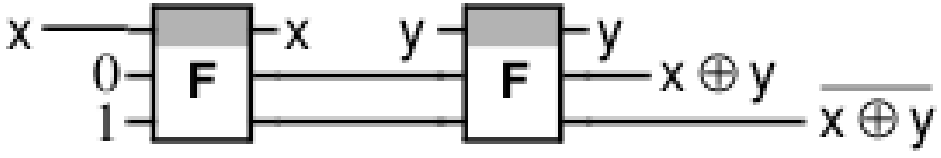


Fredkin gate implementation of NOT/FAN-OUT





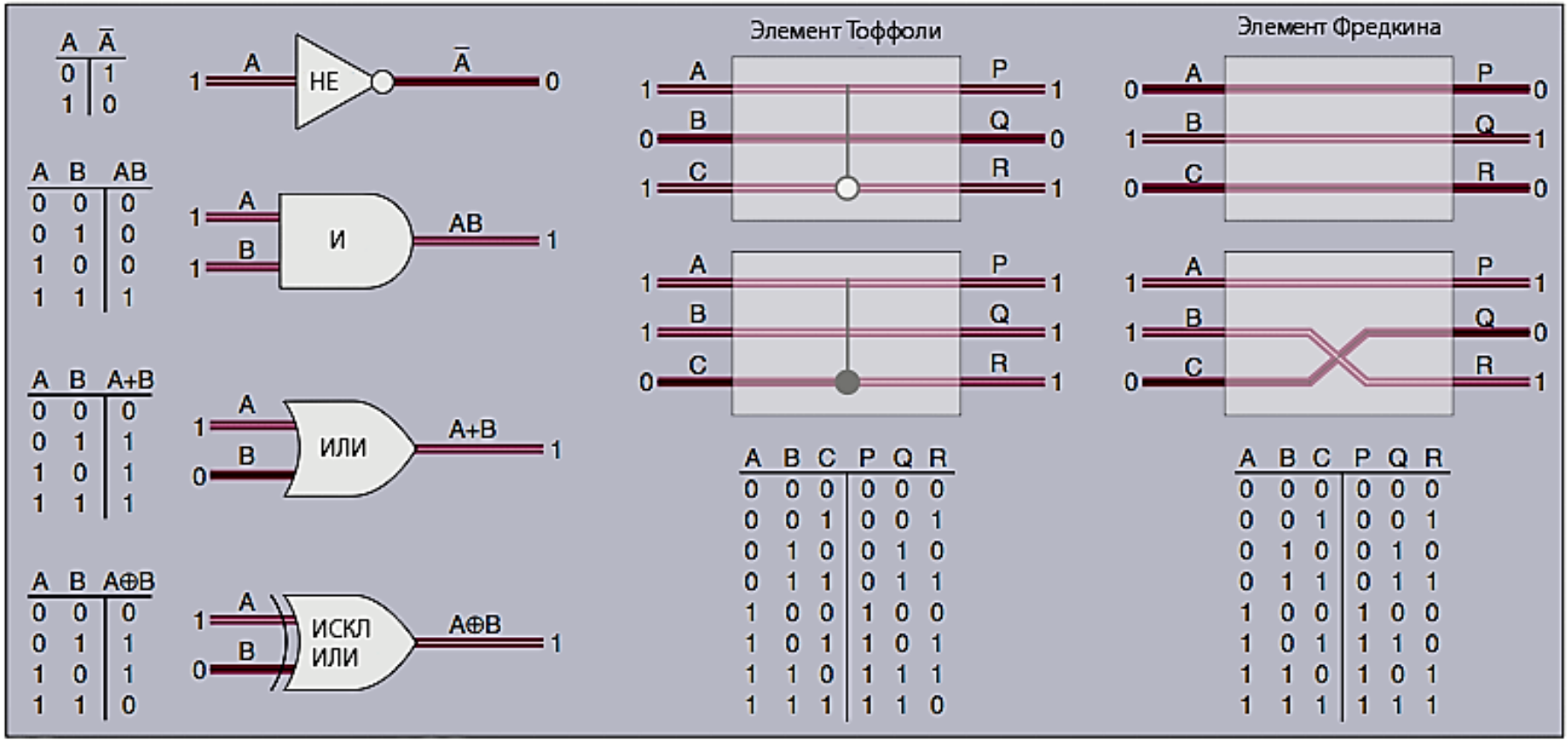
Fredkin gate implementation of OR2 and AND2



Fredkin gate implementation of XOR2

Сам факт того, что только управляемым перемешиванием проводков между входом и выходом можно получить вычисление любой булевой функции далеко не очевиден.

Самообратимость



Элементы NOT, CNOT, TG и FRG — обратны самим себе.

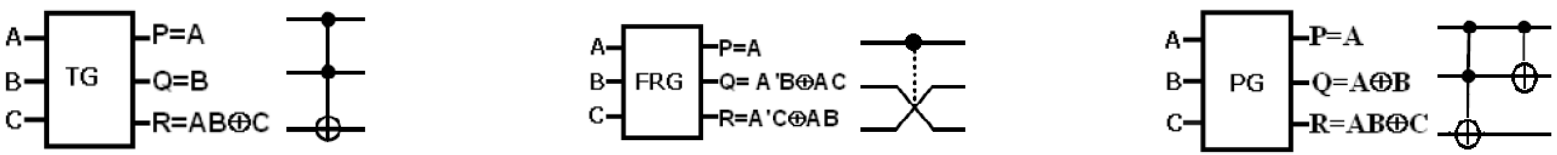
Вентиль Переса (PG)

— реализует обратимый *полусумматор* не является универсальным. Он имеет три входа A, B, C и три выхода P, Q, R.

Формулы выходов:

$$P = A, \quad Q = A \oplus B, \quad R = C \oplus AB.$$

Здесь A и B — суммируемые разряды, C — перенос из предыдущего разряда, Q — сумма, R — перенос в следующий разряд и P — мусорный выход.



а) б) в)

Рис. 3. Изображение элементов а) TG, б) FRG, в) PG.

PG=TG+CNOT

Вентиль Переса может быть реализован на гейтах TG и CNOT, соединённых последовательно, как показано на рис. 4.

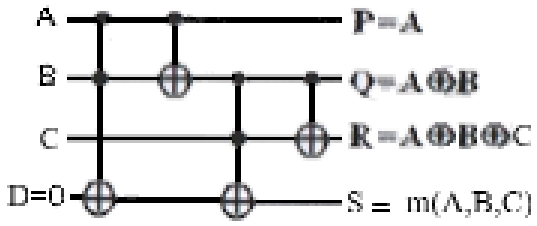


Рис. 4. Реализация элемента Переса на гейтах TG и CNOT.

Обобщённый (множественно управляемый) гейт Фредкина (обобщённый CSWAP, C^nSWAP, MCF, Multiple-Control Fredkin Gate) имеет n входов $A_1, \ldots, A_n$ из которых первые $n-2$ управляющие, и n выходов $P_1, \ldots, P_n$, последние два из которых — сигнальные, повторяющие A_{n-1} и A_n, если все остальные входы — единичные, или не делая этого, иначе.

Утилизация мусора

Мусорные биты просто отбросить нельзя, их нужно образом утилизировать.

Метод грубой силы: сохранение каждого мусорного значения на выходе $\Rightarrow$ такой метод съест всю доступную память.

Идея (Фредкин, Тоффоли и их студенты):

1) произвести вычисления, получив большое количество мусора;
2) записать результат, получив ещё немного мусора;
3) выполнить вычисления в обратном направлении, уничтожив мусор, полученный на шаге 1).

Требуется дополнительная память, но если код программы разработаны весьма аккуратно, то количество накапливающегося мусора можно держать в разумных рамках.

Методы синтеза обратимых схем

Единых методов синтеза не существует.

Для целей синтеза схем обратимые элементы объединяют в библиотеки.

- **NCT** = NOT+CNOT+TG — наименьший набором элементов, из которых можно составить любую обратимую схему.
- **NCTSF**= **NCT**+FG;
- **NCTSFP** = **NCTSF**+PG;
- **GT/MCT** — парметрическая библиотека, состоящая из n-битных обобщённых элементов Тоффоли; включает в себя все гейты из **NCT**.
- **GT+PG**.

Методы синтеза обратимых схем

Пример: полусумматор (необратимый)

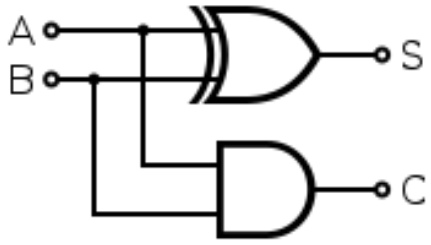
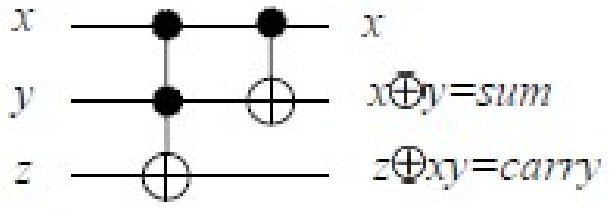


Рис. 5. Схема полусумматора на функциональных элементах XOR2 и AND2

Обратимый полусумматор образуют TG и CNOT, соединённые последовательно —



Реализованная обратимая логика

Исследования в области синтеза средств на базе обратимой логике проводятся в MIT, University of Florida, Universiteit Gent...

Разрабатывается все: от физики до систем программирования.

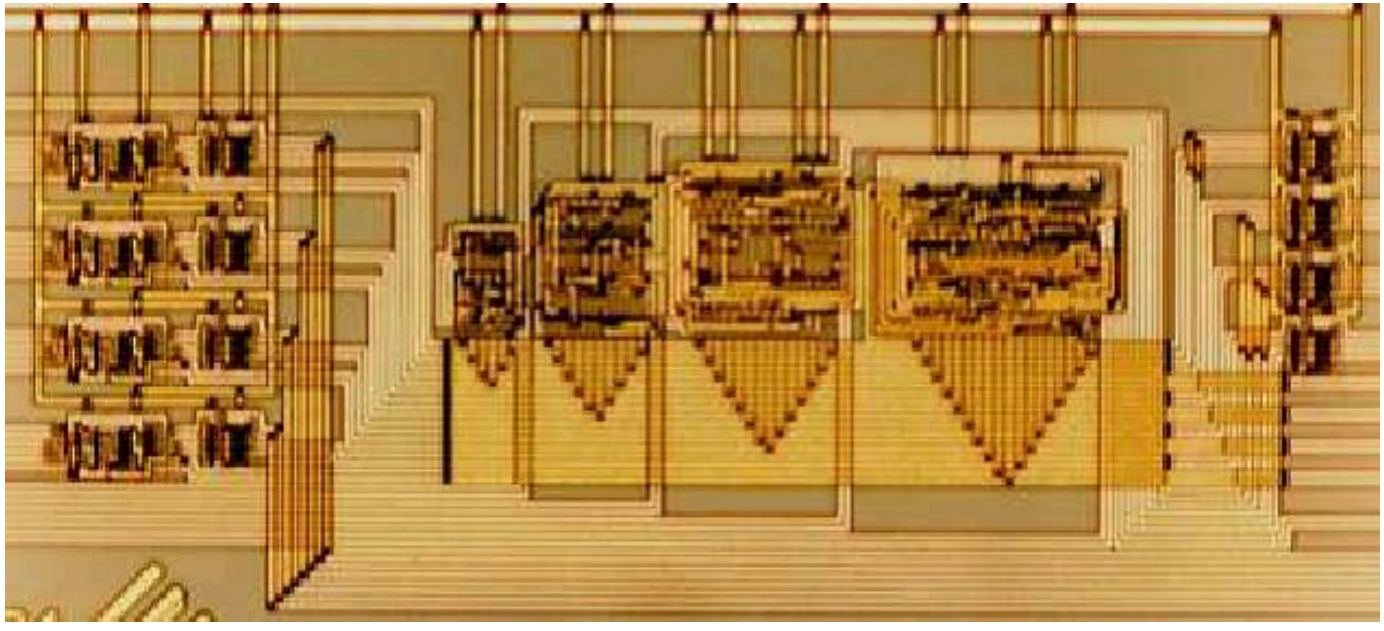
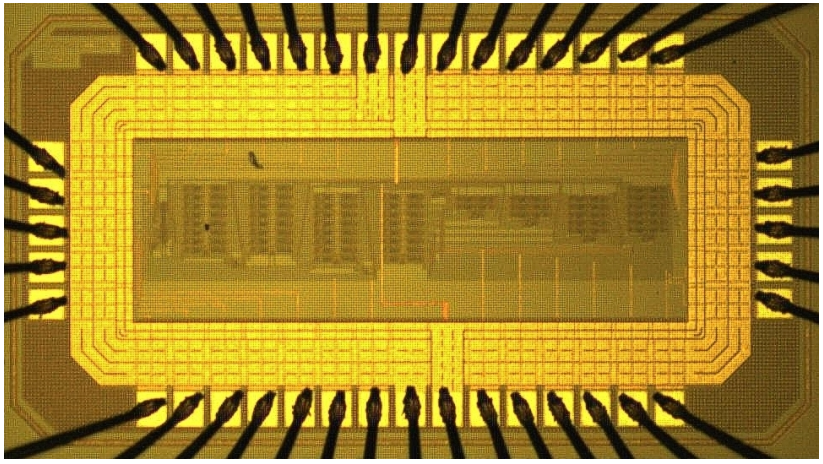


Рис. 6. Микрофотография обратимого 4-битного быстрого сумматора на FG, 320 транзистров.

Реализованная обратимая логика



Обратимый 8-битный умножитель на $\sqrt{2}/2$, 2 504 транзистора

Обратимый 8-битный сумматор, 140 мк$\times$230 мк

План выступления

Сбой — кратковременный случайный самоустраняемый отказ

Причины:

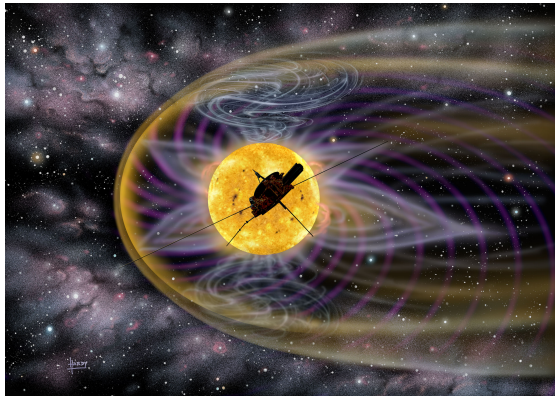
- радиационные помехи (до 50%);
- скачки напряжений питания;
- деградация сигналов во времени и др.

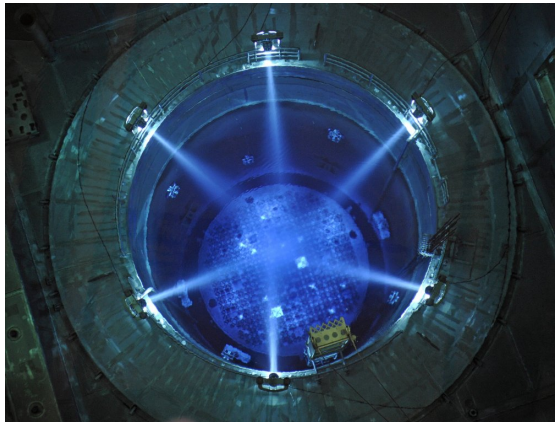
Ионизация высокой интенсивности: α- и γ-излучение + протоны и нейтроны $\rightarrow$ импульсы переходного тока $\rightarrow$ $\rightarrow$ *переключение битов в схемах функциональной логики*

Период работоспособности аппаратуры спутников связи — не менее 15 лет

Рад — внесистемная единица измерения поглощённой дозы ионизирующего излучения

Место	*Доза*
Поверхность Земли	**1 кРад/год**
Низкие орбиты Земли (МКС)	2 кРад/год
Радиационные пояса Земли	**100 кРад/год**
Геостационарные орбиты Земли	10 кРад/год
Поверхность Марса	1 кРад/год
Путь до Марса	5 кРад/год
Объекты ядерной энергетики	**100 кРад/день**





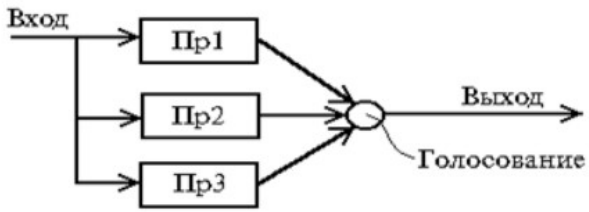
Критически важные (mission-critical) системы, приложения...

32-битное микропроцессорное ядро LEON3 СнК (летает) для европейских космических проектов — *180* нм-технология

Подходы к решению задачи

Рассматриваем пока комбинационные схемы

1. *троирование* (тройное модульное резервирование, TMR, Triple Modular Redundancy)
   Наличие *воутера*



2. *частичное дублирование* (наиболее сбоеопасных подсхем);
3. защита *схемотехническими методами* т. н. *радиационное закаливание (radiation hardening)* — размеры элемента, технология, ...;
4. *обнаружение + пересчёт* (CED, Concurrent Error Detection) — временна́я избыточность;
5. ...

Недостатки: увеличение площади (TMR: >200%), энергопотребления, снижение скорости...

6. применение *избыточного кодирования* (error-correcting code).

Обратимые логические элементы, сохраняющие чётность

1. Элемент Фредкина FRG (рассмотрен ранее).
2. *Двойной элемент Фейнмана* (F2G) — 3×3-гейт:

$$P = A, \quad Q = A \oplus B, \quad R = A \oplus C.$$

3. *Новый сбоеустойчивый элемент* (NFT) — 3×3-гейт не имеющий адресных входов:

$$P = A \oplus B, \quad Q = B'C \oplus AC', \quad R = BC \oplus AC'.$$

4. *Элемент Ислама* (IG, предложен Сайфулом Исламом в 2009 г.) — 4×4-гейт:

$$P = A, \quad Q = A \oplus B, \quad R = AB \oplus C, \quad S = BD \oplus B'(A \oplus D).$$

5. *Модифицированный элемент Ислама* (MIG) — улучшенная версия 4×4-гейта IG, реализующая для первых трёх выходов те же формулы, что и IG и $S = AB' \oplus D$.

Схема называется *сбоеустойчивой*, если она обеспечивает обнаружение или обнаружение и исправление ошибок на своём выходе. Переходя от элементов к схемам, замечаем, что существует два вида сбоеустойчивых схем: схемы, сохраняющие чётность, состоящие из сохраняющих чётность обратимых элементов, и схемы, устойчивые к ошибкам.

Примером схемы первого вида может служить сумматор, предложенный С. Исламом:

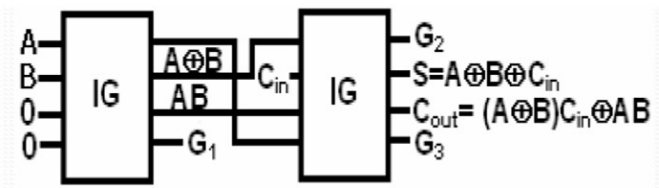


Рис. 7. Однобитный обратимый сумматор на элементах IG, сохраняющий чётность

Обнаружение ошибок

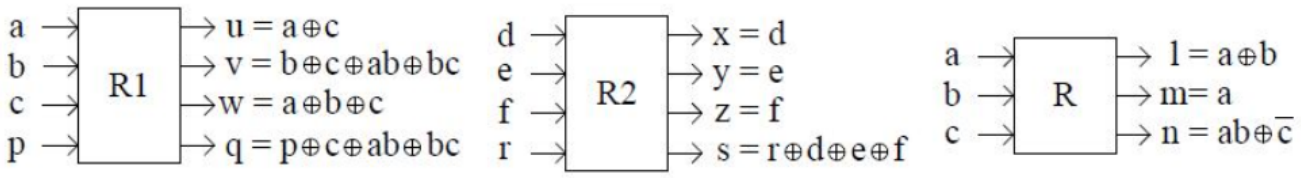
В простейшем случае сбоеустойчивая схема обеспечивает лишь обнаружение возникшей ошибки.

Исправление — многократный пересчёта выхода — *CED, Concurrent Error Detection*.

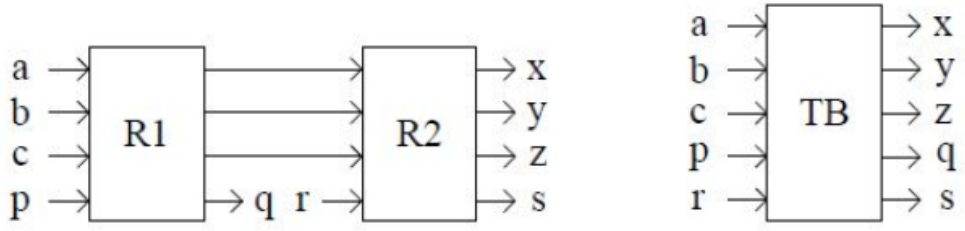
На сегодняшний день исследования методов синтеза сбоеустойчивых обратимых схем ограничиваются практически исключительно указанным простейшим вариантом.

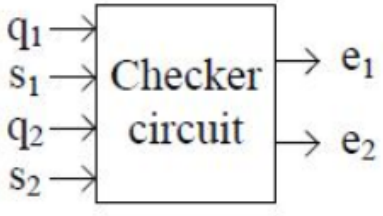
Ясна ограниченность таких методов в возможности обнаружения ошибок.

Онлайн-тестируемых схемы: на элементах R1, R2, R



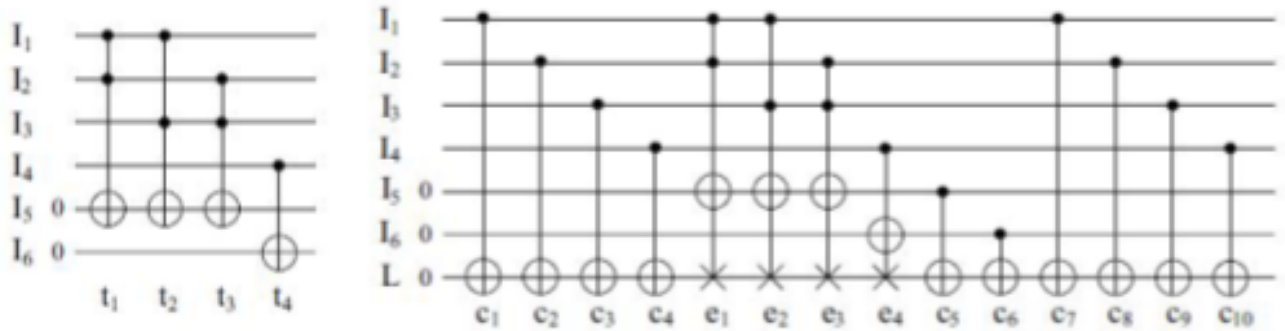
Тестируемые блоки (ТВ) составляются из R1, R2, имеет два выхода чётности q и s, которые подаются на проверяющую схему (Checker circuit), составленную из 8 элементов R.



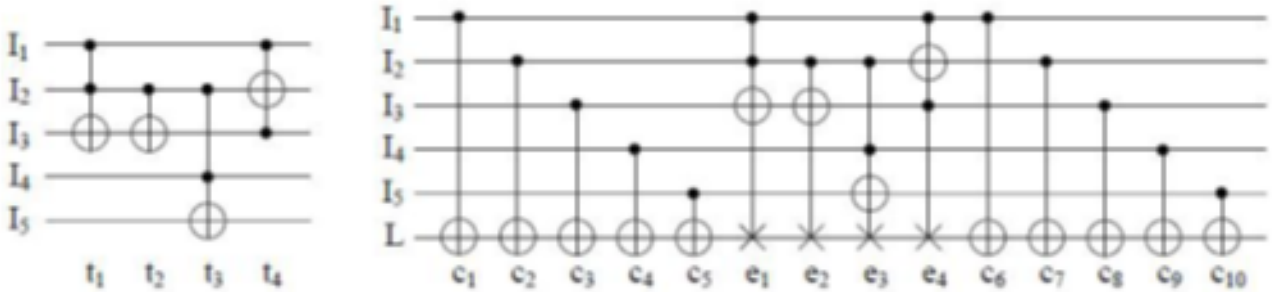


Виды онлайн тестируемых схем

- Схемы на элементах с пара-, многофазными входами: Ошибка фиксируется проверкой выходного вектора
- Схемы, основаные на ESOP:



- Схемы, основаные на Toffoli:



Проверяющий элемент составлен из 4 элементов Тоффоли и 3 элементов Фредкина с 3 битами

Что синтезировано?

В последнее время предложены следующие схемы сбоеустойчивых вычислительных устройств использующих контроль чётности:

- полный сумматор из сохраняющих чётность блоков (2010);
- АЛУ (2013);
- компрессор (устройство для сжатия динамического диапазона звукового сигнала);
- полный сумматор (2015).

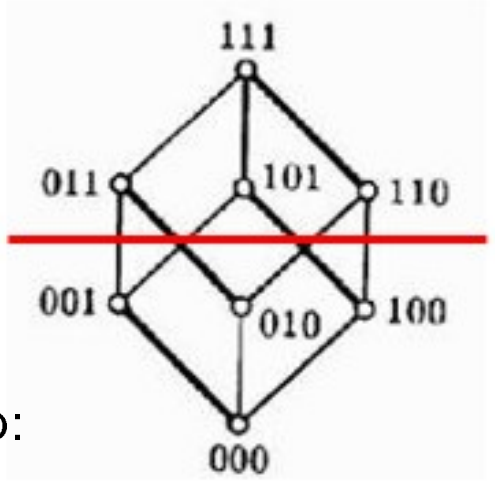
План выступления

Пространство Хэмминга

n-мерный единичный куб — *пространство Хэмминга*

Будем работать в 3-мерном таком пространстве, кодировать

$$0 \longmapsto \underbrace{000}_{\text{Полюс\_0}} \qquad 1 \longmapsto \underbrace{111}_{\text{Полюс\_1}}$$



и использовать 3 проводника вместо одного:

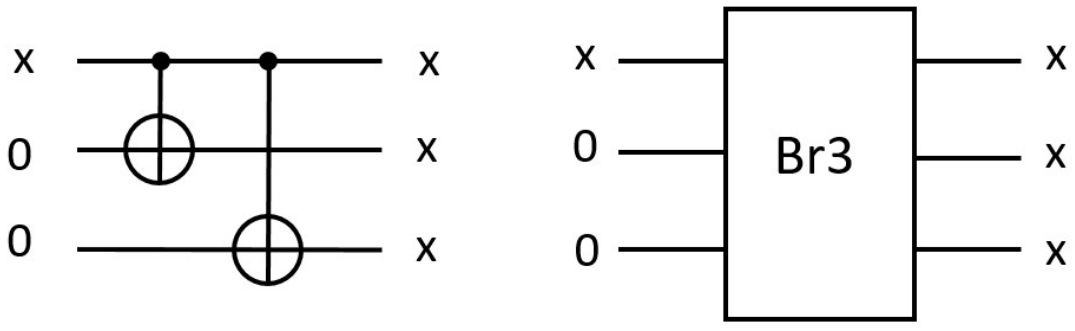
$$A \longmapsto A_1,\ A_2,\ A_3.$$

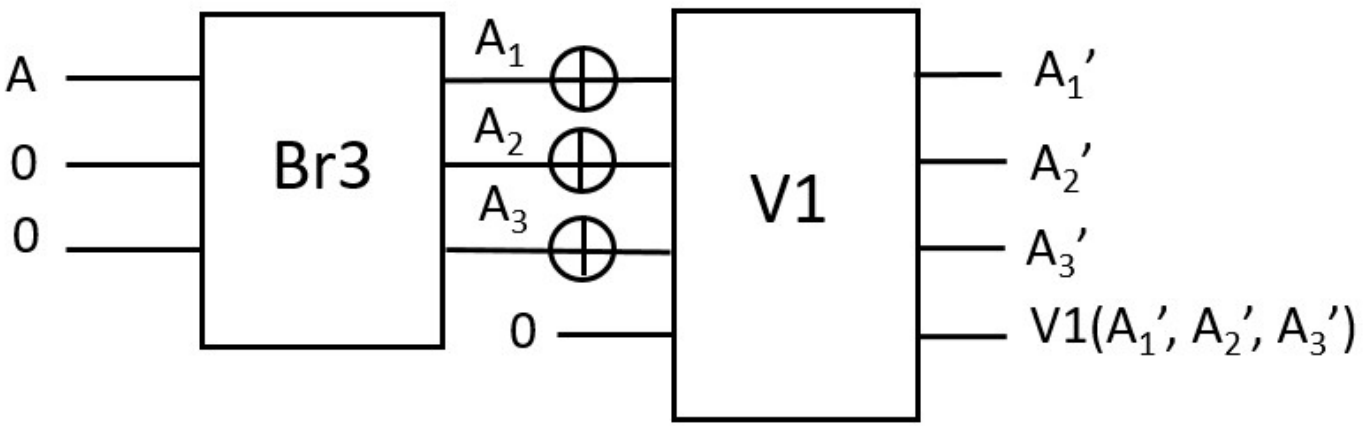
Коррекция одиночной ошибки происходит автоматически в неявном виде.

Для обратимой схемотехники это существенно: происходит исправление ошибки, а не просто фиксация факта, что ошибка произошла.

Отличие от TMR на уровне элементов —

— здесь имеется воутер, который не защищен от ошибок





Сбой в воутере V1, приводящий к искажению сигнального выхода, данной схемой исправляться не будет.

Элемент CNOT в пространстве Хемминга

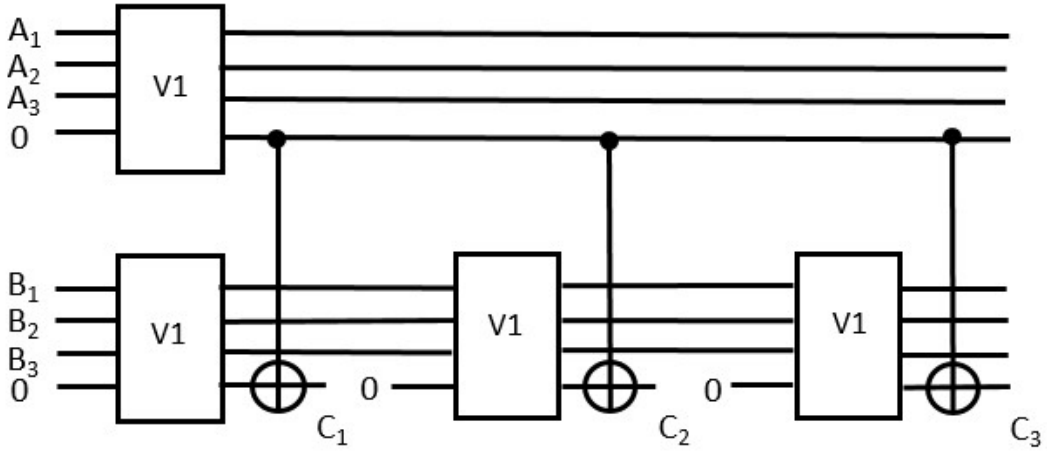


Рис. 8. Гейт HCNOT управляемого инвертирования.

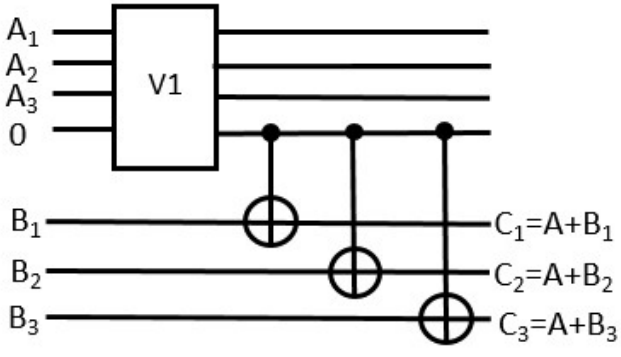
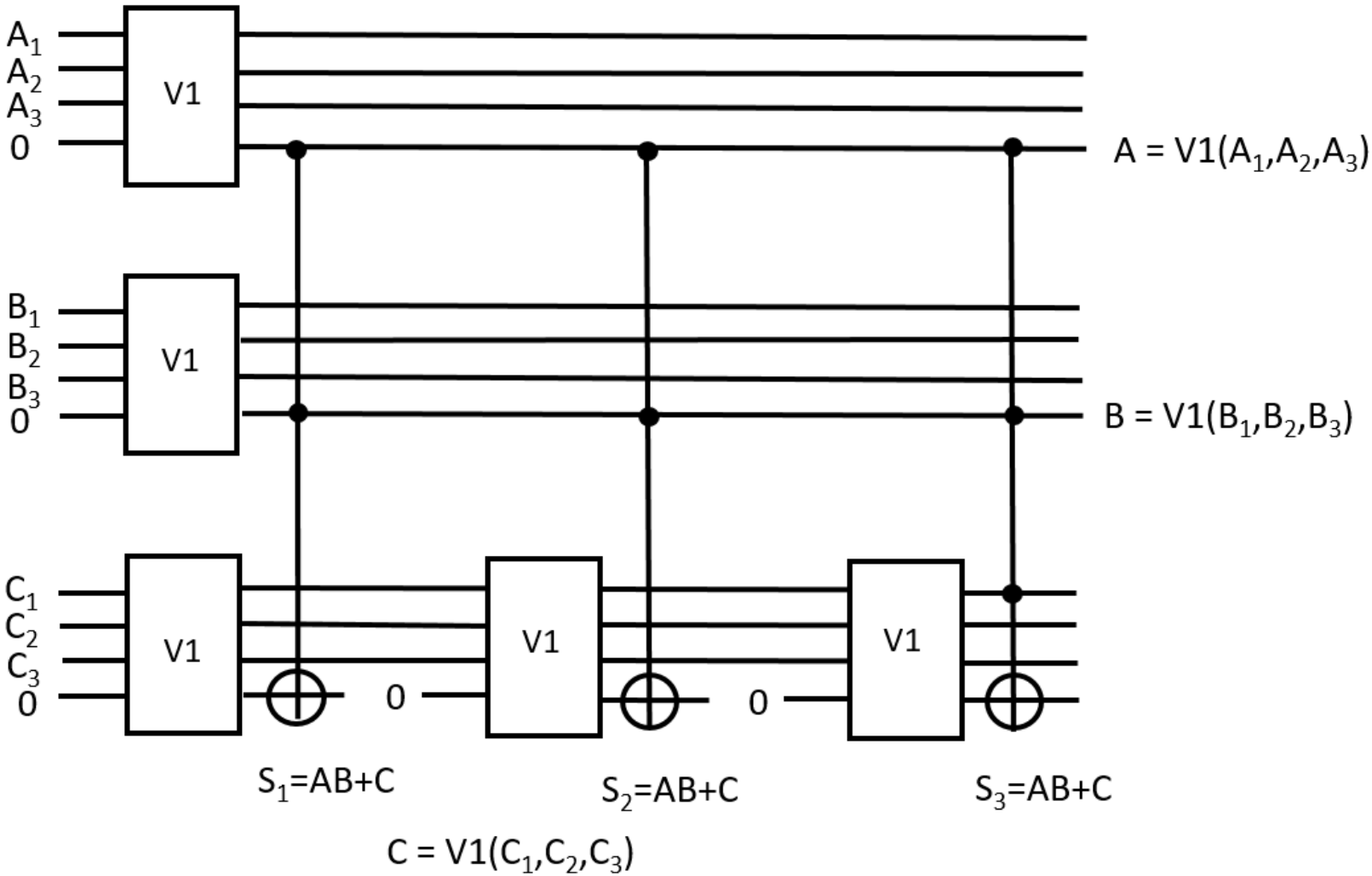
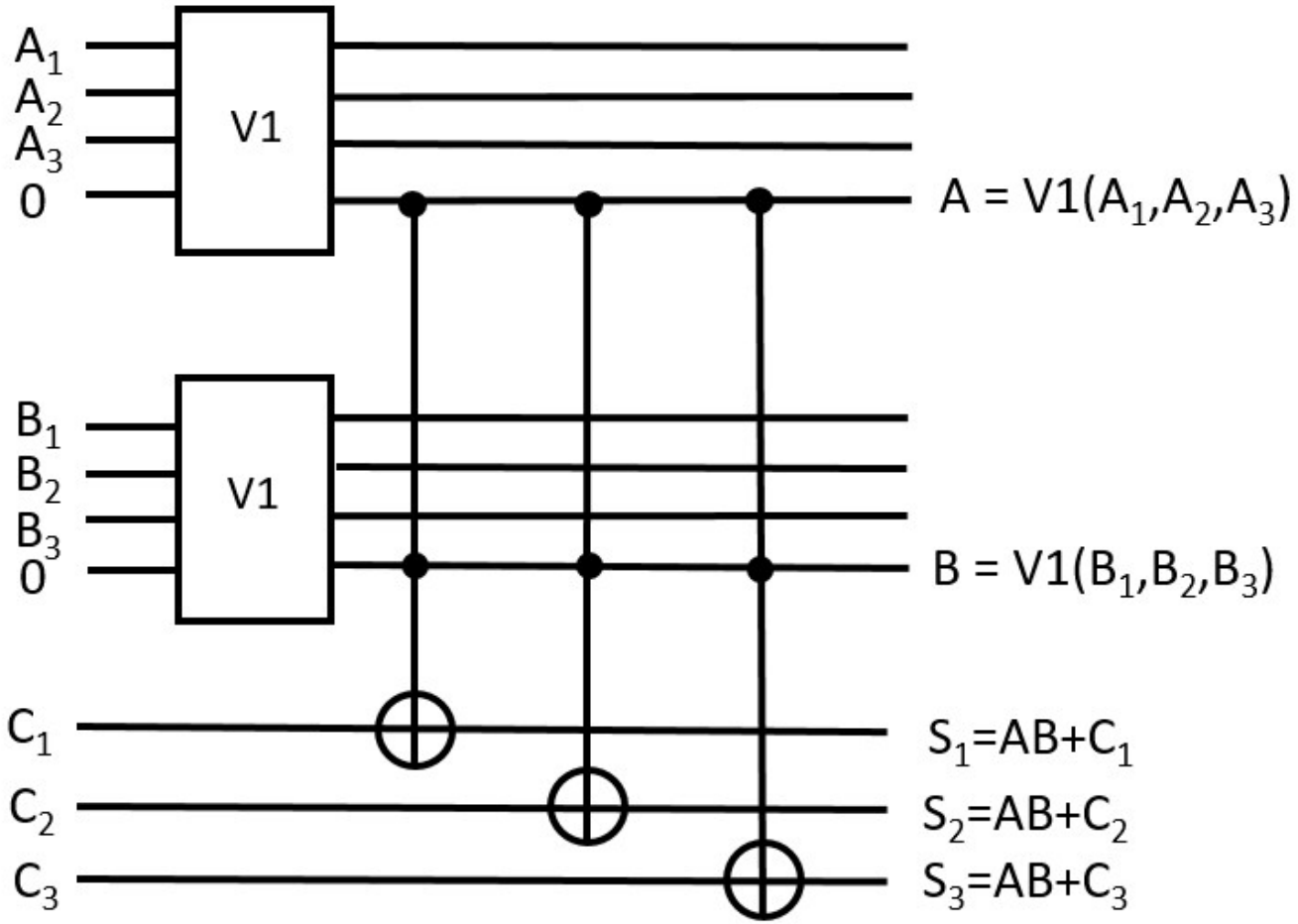


Рис. 9. Облегчённый гейт HLCNOT управляемого инвертирования.

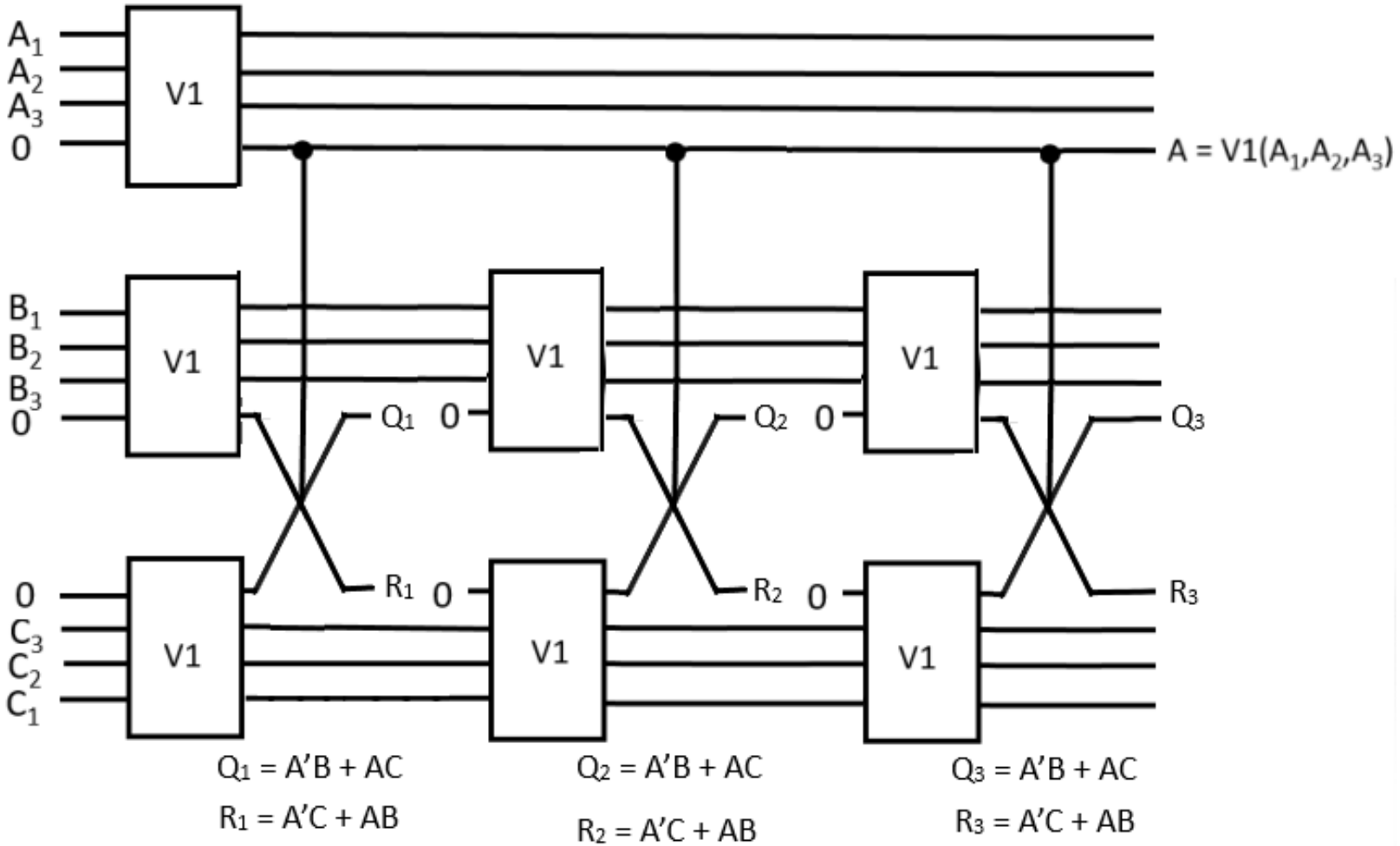
Гейт Тоффоли HTG в пространстве Хэмминга



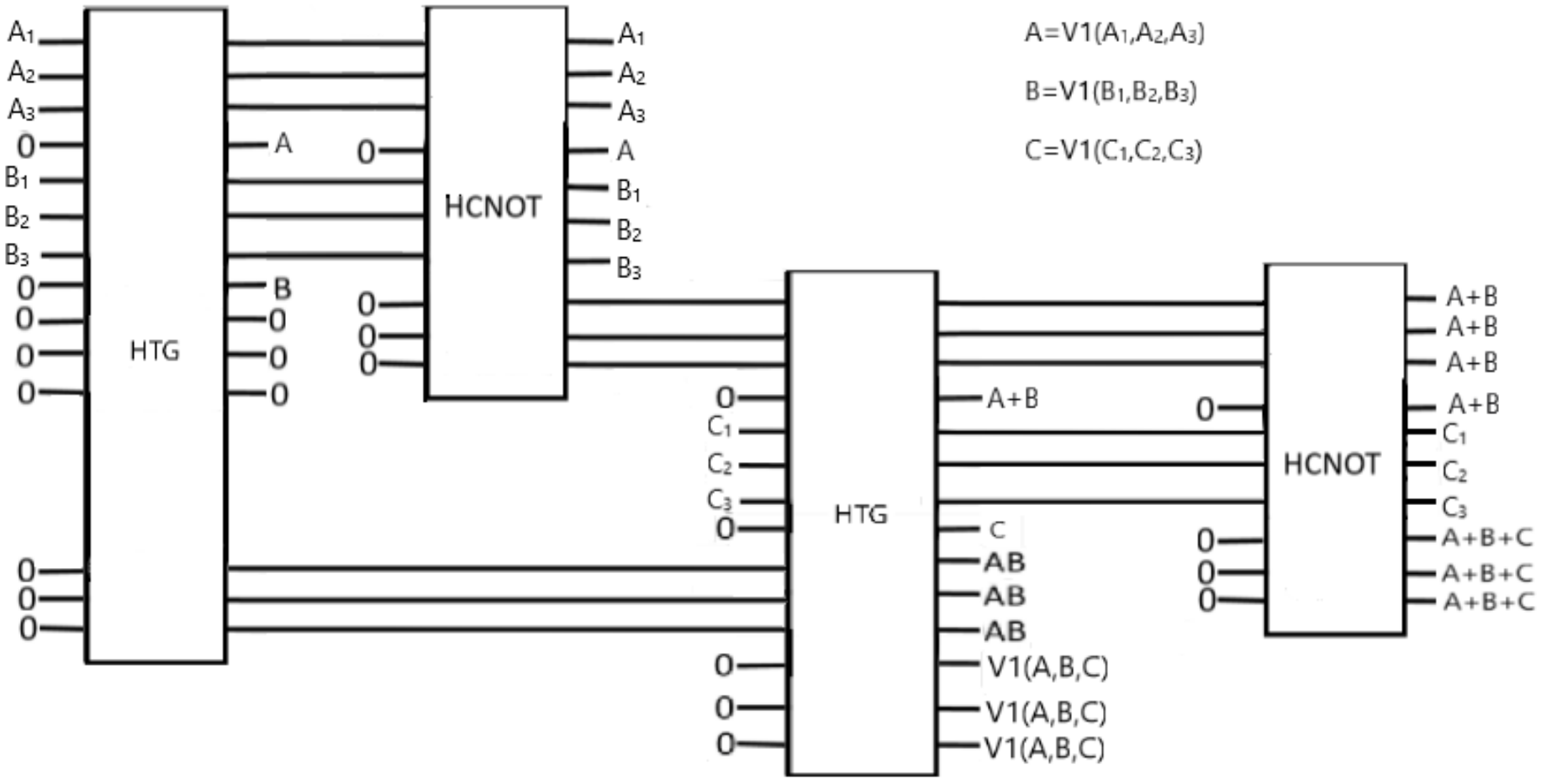
Гейт Тоффоли HTG в пространстве Хэмминга: облегчённая схема



Элемент Фредкина в Хэмминговом пространстве



Обратимый сумматор в Хэмминговом пространстве



Синтез в пространстве Хэмминга: выводы

Сравнение сложности обычных обратимых схем с построенными в хэммингoвом пространстве:

- число вентилей в TMR возрастает примерно в 4 раза, как при кодировании в пространстве Хэмминга оно возрастает на порядок;
- сбоеустойчивость предлагаемого метода превосходит TMR.

Вывод: метод вряд ли применимым для всей схемы, оставаясь целесообразным для наиболее уязвимых элементов и подсхем, сбоеустойчивость которых критически важна для функционирования всей вычислительной системы.

Поляризованные пространства Хэмминга

В целях более сбалансированного представления сигналов 0 и 1, и, соответственно, нагрузки на транзисторные вентили при реализации схем на кристалле, более эффективным является выбор при кодировании в пространстве Хэмминга:

$$\text{Полюс\_0} = 010, \qquad \text{Полюс\_1} = 101$$

Вектор поляризации $= 010 = 2_{10}$.

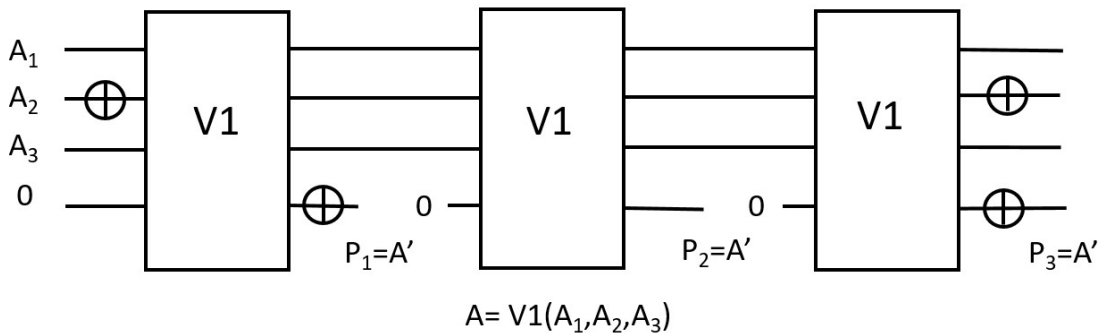
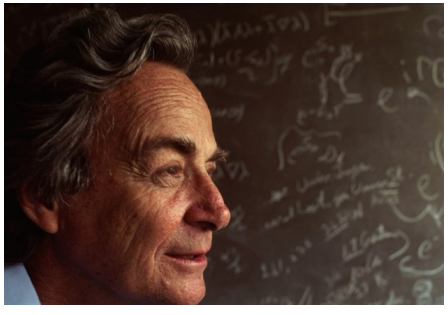


Рис. 10. Элемент H_2NOT

Спасибо за внимание!

Вопросы?

Рольф У. Ландауэр (1927–1999), IBM Fellow



Ричард Ф. Фейнман (1918–1988),
California Institute of Technology

Томмазо Тоффоли (1943), Boston University

Эдвард Фредкин (1934), Carnegie Mellon University