

УДК 621.382

А.И. Солдатов, О.Х. Ким, А.А. Солдатов, М.А. Костина, Г.Н. Нариманова

Бесконфликтный, отказоустойчивый и компактный программируемый коммутатор

Приводится техническое решение построения коммутатора. Исследуются его бесконфликтность, отказоустойчивость и компактность. На примере коммутатора на 16 выводов показано, что предложенная структура коммутатора позволяет выполнять электрические соединения на всем множестве внешних выводов со 100% использованием ресурсов по количеству реализуемых цепей. Приведены два варианта реализации 8 цепей на 16-выводном коммутаторе. Показана его отказоустойчивость. Предложенный подход построения коммутатора позволяет реализовать коммутатор любого размера с уменьшением количества ключевых элементов.

Ключевые слова: кросс-бар, кросс-маршрутизатор, программируемая коммутационная среда, бесконфликтность, отказоустойчивость, компактность.

doi: 10.21293/1818-0442-2021-24-3-12-17

В настоящее время кластерные многопроцессорные вычислительные системы (МВС) и суперкомпьютеры являются основной вычислительной платформой для большинства современных технических комплексов. Известные архитектуры кластерных МВС и суперкомпьютеров (векторные, конвейерные и др.) функционально ориентированы и направлены на выполнение определенного класса вычислительных алгоритмов. В разных прикладных применениях используются наиболее оптимальные архитектуры МВС и суперкомпьютеров.

Общей проблемой для всех типов архитектур кластерных МВС и суперкомпьютеров является разработка бесконфликтного, отказоустойчивого и компактного кросс-бара на большое количество абонентов, способного парировать в режиме on-line большое количество отказов. Кросс-бар обеспечивает связь между абонентами по принципу «каждый с каждым». Бесконфликтный, отказоустойчивый и компактный кросс-бар необходим для разработки высокопроизводительных и отказоустойчивых мультиядерных процессорных узлов в виде СБИС и/или сложного модуля 3D-интеграции.

Для кластерных МВС и суперкомпьютеров, использующих интерфейсы и протоколы SpaceWire-GigaSpaceWire-SpaceFibre, получивших широкое признание в аэрокосмической отрасли [1–3], актуальной проблемой является разработка бесконфликтной, отказоустойчивой и компактной системы коммутации (NxN), обеспечивающей связь между двумя группами абонентов по принципу «каждый с каждым». Бесконфликтная, отказоустойчивая и компактная система коммутации (NxN) необходима для разработки в виде СБИС или/и сложного модуля 3D-интеграции электронных компонентов кросс-маршрутизаторов, обеспечивающей связь по принципу «каждый с каждым» между процессорами, входящими в разные процессорные узлы.

В системе PERCS (Productive Easy-to-use Reliable Computing System) фирмы IBM [4] процессорный узел содержит 4 процессора Power PC, связанных между собой кросс-баром по принципу

«каждый с каждым». В системе YARC фирмы CRAY [5] процессорный узел представлен 4 спаренными процессорными чипами Xeon, также связанными между собой системой коммутации типа кросс-бар. Попытка увеличения количества процессоров в процессорном узле до 8 в системе YARC привела к ухудшению технических характеристик и свойств используемого кросс-бара.

В рассматриваемых аналогах кластерных суперкомпьютеров (системе PERCS и системе YARC) кросс-бар для процессорного узла выбран благодаря его свойству бесконфликтности. Этот выбор можно считать вынужденным и обоснованным с технической точки зрения. Другие топологии сетей (сеть Клоза, мульти (кольца), гиперкубы, торы и др.) в сравнении с кросс-баром имеют меньшую аппаратную избыточность, но по этой причине всем этим системам коммутации неизбежно присущи сетевые конфликты и тупики, за которыми встают такие проблемы, как размещение больших буферов по сети, управление маршрутизацией, введение виртуальных каналов и, по сути, разработка специфичной сетевой операционной системы (ОС). Все известные топологии сетей, в том числе и кросс-бар, не отказоустойчивы и по этой причине резко снижают производительность процессорных узлов при появлении хотя бы одного дефекта, а при появлении нескольких дефектов процессорный узел часто «зависает» и/или работает практически «вхолостую».

Актуальность проблемы состоит в том, что электрической схемы бесконфликтного, отказоустойчивого и компактного кросс-бара, способного парировать в режиме on-line большое количество отказов, в настоящее время не существует. Размерность существующих аналогов ограничена 4–6 абонентами по причине квадратичного роста объема аппаратуры при увеличении числа абонентов и проблемами обеспечения бесконфликтной маршрутизации информации между абонентами по мере увеличения числа абонентов.

Лучшим отечественным решением поставленной проблемы считается 24-портовый кросс-бар

ОАО НИЦЭВТ (2018) [6–8]. Кросс-бар выполнен в виде кристалла с архитектурой полного графа (8×8), затем четыре кристалла соединили в один маршрутизатор. Для этого по два порта от каждого кристалла пришлось использовать на соединение кристаллов между собой. Оставшиеся 24 порта используются для внешних соединений. В этом кросс-баре часто возникают трудно прогнозируемые конфликты, потери в отказоустойчивости и другие проблемы. Этот кросс-бар можно назвать кросс-баром условно.

Существующие аналоги систем коммутации сигналов ($N \times N$) также не обладают отказоустойчивостью, так как между любыми двумя выводами из ($N \times N$) существует только один путь и при отказе хотя бы одного ключевого элемента маршрутизация информации по этому пути становится невозможной. Бесконфликтная, отказоустойчивая система коммутации ($N \times N$) на большое число абонентов в настоящее время также еще не создана.

Постановка задачи

В связи с этим в настоящее время единственным выходом для повышения надежности процессорных узлов и кросс-маршрутизаторов является резервирование всех кросс-баров и систем коммутации ($N \times N$) путем дублирования и/или мажорирования. При этом имеет место соответствующий кратный рост объема аппаратуры при незначительном увеличении надежности системы в целом.

Из существующих технических идей решения этой проблемы можно выделить исследования, проведенные отечественными учеными М.Ф. Караваем, В.С. Подлазовым [9]. Исследователями предложена новая концепция построения кросс-маршрутизаторов с логическими функциями квазиполного графа на базе разработанной математической модели структуры сетей SpaceWire-GigaSpaceWire-SpaceFibre – теоретико-групповые блок-схемы (block-designs) в комбинаторике: $B(N, s, \sigma)$, где N – число вершин (абонентов); s – степени всех узлов графа; σ – степень резервирования связей. Эквивалентная ей графовая модель – двудольный граф (рис. 1), где одна группа вершин интерпретируется как абоненты (прямоугольники), а вторая – как локальные коммутаторы (окружности).

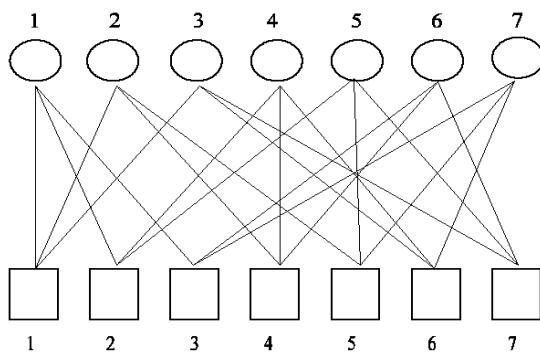


Рис. 1. Двудольный граф 7×7

Эта конструкция была названа авторами *квазиполным графом*, чтобы подчеркнуть совпадение логических характеристик с характеристиками полно-

го графа для того же числа абонентов. В то же время число портов и линков у квазиполного графа уменьшается в \sqrt{N} раз. Для 100 абонентов число портов у абонентов сокращается в 10 раз. Доказано, что это решение минимальное. Любая попытка редукции этого решения приводит к потере достигнутых свойств полного графа. В общем виде эта задача принадлежит к классу *NP-полных задач*, что означает доступность точного решения универсальной задачи только через полный перебор всех возможностей.

Общее количество абонентов и локальных коммутаторов равно N . Степени всех узлов графа равны s . Степень резервирования связей – σ . Между параметрами графа существует зависимость

$$N = s \frac{s-1}{\sigma} + 1. \quad (1)$$

Таким образом, современный уровень решения рассматриваемой проблемы – это кросс-бар на (4–8) абонентов, не обладающий отказоустойчивостью. Причина – квадратичный рост объема аппаратуры и проблемы обеспечения бесконфликтной маршрутизации информации между абонентами по мере увеличения числа абонентов и отсутствие электрической схемы кросс-бара, обладающего отказоустойчивостью.

По мнению авторов статьи, основным источником проблем, возникающих при разработке высокопроизводительных и отказоустойчивых кластерных МВС и суперкомпьютеров, является отсутствие электрической схемы системы коммутации, позволяющей разработать:

1) бесконфликтный, отказоустойчивый и компактный кросс-бар на большое число абонентов, например до 64 абонентов;

2) бесконфликтную, отказоустойчивую и компактную систему коммутации ($N \times N$) на большое число абонентов, например до (64×64) , для разработки бесконфликтных и отказоустойчивых кросс-маршрутизаторов для сетей SpaceWire-GigaSpaceWire-SpaceFibre.

Предлагаемое решение

Обозначенные выше проблемы создания кластерных МВС и суперкомпьютеров предлагается решить на базе разработанного бесконфликтного и компактного программируемого коммутатора [10–12]. Бесконфликтность подразумевает полную независимость каждого электрического соединения от других. На рис. 2 представлена схема коммутатора на 16 внешних выводов (абонентов). Внешние выводы коммутатора для подключения абонентов представлены группами вертикальных шин (A1–A4, B1–B4, C1–C4, D1–D4) верхней группы матричных коммутаторов K1.1...K1.4. Предлагаемая концепция построения коммутатора основана на использовании двух групп матричных коммутаторов уменьшенной размерности, в которых i -е горизонтальные шины каждого матричного коммутатора верхней группы соединяются с i -ми вертикальными шинами i -го матричного коммутатора нижней группы, что позволяет обеспечить выполнение любых электрических соединений по принципу «точка–точка» на всем множестве внешних выводов.

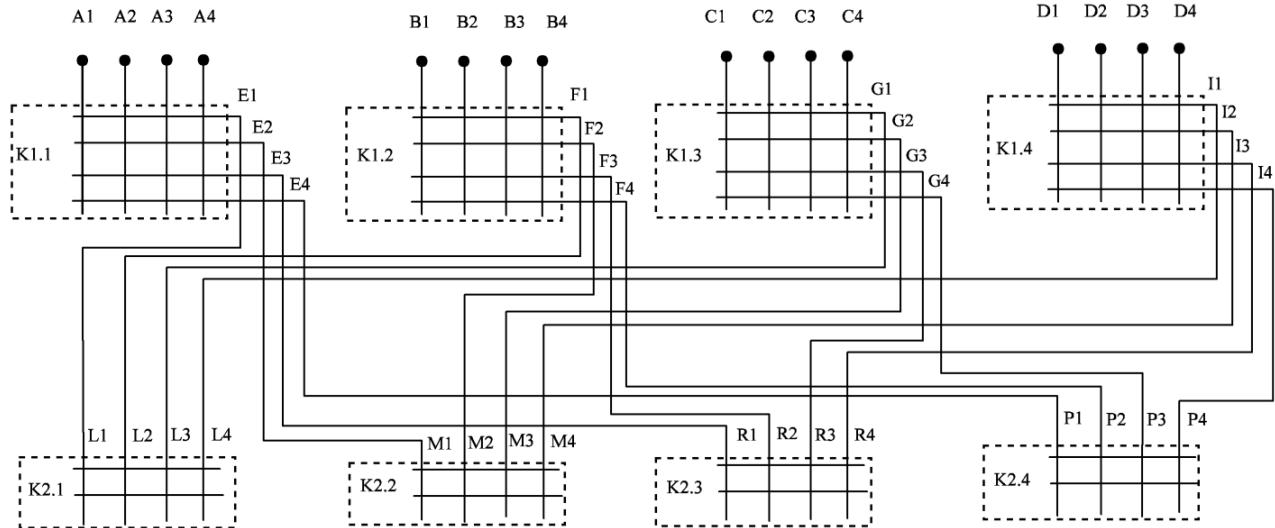


Рис. 2. Электрическая схема бесконфликтной, отказоустойчивой и компактной ПКС на 16 внешних выводов

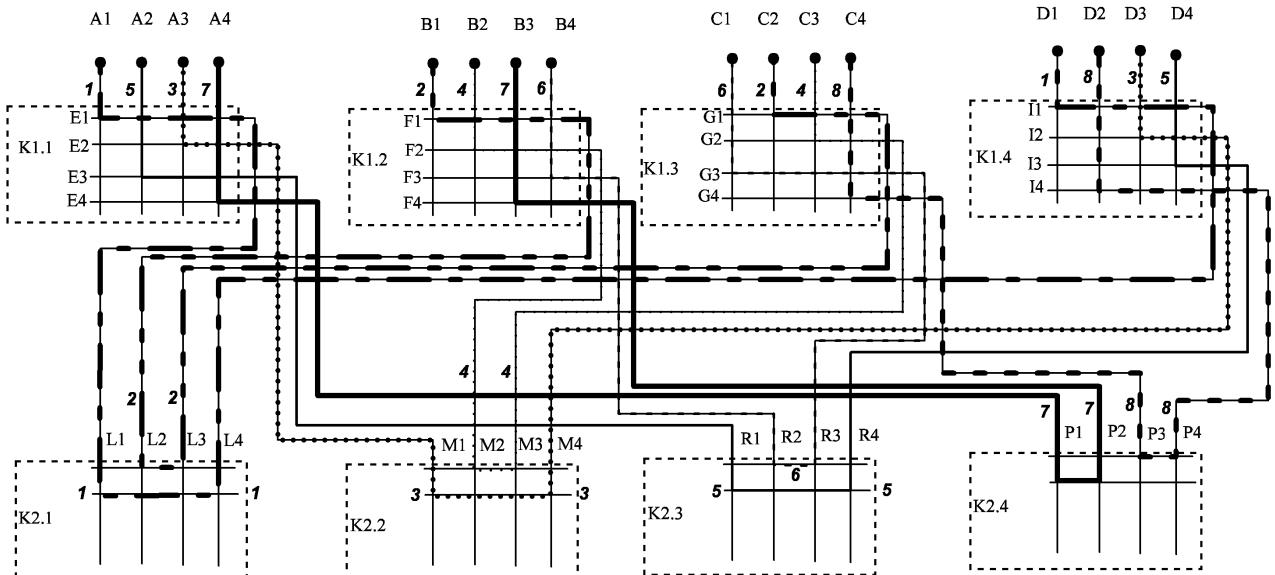


Рис. 3. Пример выполнения 8 электрических цепей, содержащих по два вывода

На рис. 3 показан один из возможных вариантов реализации 8 электрических цепей по принципу «точка–точка» на множестве из 16 внешних выводов, т.е. рассматривается вариант со 100% использованием ресурсов коммутатора по количеству реализуемых цепей. Для более наглядного представления все цепи показаны разными линиями и помечены цифрами. Цепь A1-D1 показана штрих-пунктирной линией и отмечена цифрой – 1; цепь B1-C2 – штрих-пунктирной линией с двумя точками – 2; цепь A3-D3 – точечной линией – 3; цепь B2-C3 – точечной разреженной линией – 4; цепь A2-D4 – сплошной тонкой линией – 5; цепь B4-C1 – штриховой линией – 6; цепь A4-B3 – толстой сплошной линией – 7 и цепь C4-D2 – штриховой линией (короткие штрихи) – 8.

Нетрудно заметить, что возможна реализация нескольких различных вариантов линий связи для соединения указанных выше внешних выводов. На основании теоремы Холла о паросочетаниях [13] можно утверждать, что данная электрическая схема коммутатора является бесконфликтной, т.е. на нем

всегда могут быть реализованы любые 8 электрических цепей, содержащих по 2 вывода. Следует отметить, что при реализации меньшего количества соединений проблема нахождения электрических связей оказывается проще.

На рис. 4 представлен еще один вариант выполнения тех же самых электрических соединений, которые показаны на рис. 3. Отличие состоит в изменении соединительных линий в первом коммутаторе второй группы K2.1. Цепь A1-D1 (1) выполнена на первой горизонтальной шине этого коммутатора, а цепь B1-C2 (4) – на второй горизонтальной шине этого же коммутатора.

Такой вариант выполнения соединений может потребоваться, если, например, ключевой элемент первой горизонтальной и второй вертикальной линии первого коммутатора второй группы K2.1 выйдет из строя. Вариантов выполнения указанных соединений может быть $N/2$. Этим обеспечивается отказоустойчивость схемы.

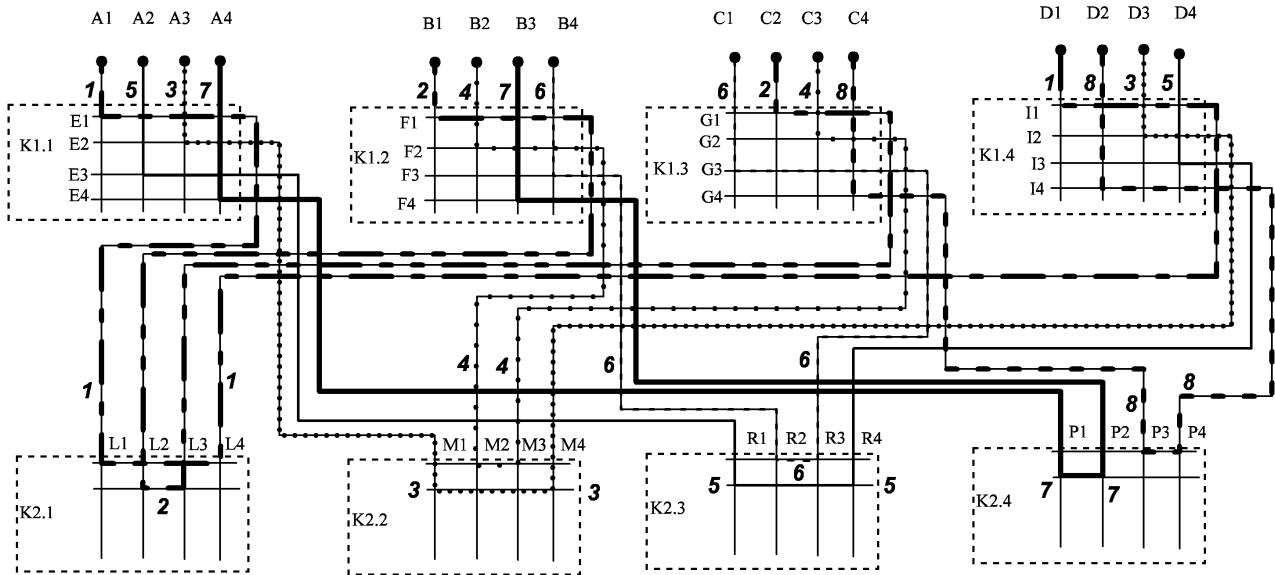


Рис. 4. Второй вариант выполнения электрических соединений, показанных на рис. 3

Количество ключей в предлагаемом коммутаторе равно

$$C_1 = n \cdot (n \cdot n) + n \cdot \left(\frac{n \cdot n}{2} \right) = 1,5n^3, \quad (2)$$

где n – количество внешних выводов уменьшенных матричных коммутаторов, N – количество внешних выводов обычного матричного коммутатора.

С учетом того, что

$$n = \sqrt{N}, \quad (3)$$

получаем:

$$C_1 = \frac{3}{2} \sqrt{N^3}. \quad (4)$$

Количество ключей в матричном коммутаторе аналогичной размерности, содержащем N внешних выводов, равно

$$C_2 = N \cdot \left(\frac{N}{2} \right) = \frac{N^2}{2}. \quad (5)$$

Используя выражения (4) и (5), найдем коэффициент уменьшения избыточности

$$C = \frac{C_2}{C_1} = \frac{\frac{N^2}{2}}{\frac{3}{2} \sqrt{N^3}} = \frac{\sqrt{N}}{3}. \quad (6)$$

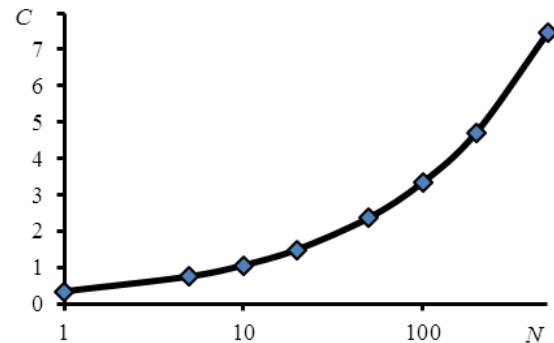
Графическая зависимость параметра C от количества внешних выводов N , рассчитанная по формуле (6) представлена на рис. 5.

Из графика следует, что при увеличении количества внешних выводов N относительное сокращение объема аппаратуры увеличивается. Это позволяет уменьшить размеры кристалла, увеличить надежность и снизить потребляемую мощность предложенного коммутатора.

Заключение

Предлагаемая электрическая схема коммутатора является бесконфликтной и отказоустойчивой, так как между любыми двумя выводами существует $N/2$

различных линий связи. Во всех аналогах между любыми двумя выводами существует только один путь. Предложенный подход может быть реализован для любого другого количества внешних выводов (абонентов), при этом в электрической схеме коммутатора общее количество ключей и схем управления ключами по сравнению с матричным коммутатором такой же размерности уменьшается в $\sqrt{N}/3$.

Рис. 5. График зависимости параметра C от количества внешних выводов N

На базе предлагаемого коммутатора становится возможной разработка бесконфликтной, отказоустойчивой и компактной системы коммутации размерностью (NxN) и кросс-бара на N внешних выводов (абонентов).

Литература

1. Журавлев В. SpaceWire: Взгляд со стороны. – Ч. 1 / В. Журавлев, А. Немытов, А. Першин // Современная электроника. – 2017. – № 8. – С. 36–42.
2. Журавлев В. SpaceWire: Взгляд со стороны. – Ч. 2 / В. Журавлев, А. Немытов, А. Першин // Современная электроника. – 2017. – № 9. – С. 26–32.
3. Горбунов С.Ф. Сетевые интерфейсы космических аппаратов: перспективы развития и проблемы внедрения /

С.Ф. Горбунов, В.Ю. Гришин, П.М. Еремеев // Наноиндустрия: спецвып. – М.: Техносфера, 2019 (89). – С. 128–130.

4. Rajamony R. PERCS: The IBM POWER7-IH high-performance computing system / R. Rajamony, L.B. Arimilli, K. Gildea // IBM Journal of Research and Development. – 2011. – Vol. 55, No. 3. – P. 3:1–3:12.

5. Resch M. High performance computing on vector system 2008 / M. Resch, R. Sabine, K. Benkert, M. Galle, W. Bez, H. Kobayashi, T. Hirayama. – Springer, 2009. – 241 p.

6. Новая разработка Ростеха позволяет создавать суперкомпьютеры в домашних условиях [Электронный ресурс]. – Режим доступа: <https://rostec.ru/news/novaya-razrabotka-rostekha-pozvolit-sozdavat-superkompyutery-v-domashnikh-usloviyakh/>, свободный (дата обращения: 16.10.2020).

7. Каравай М.Ф. Расширение коммутационных возможностей сети «Ангара» / М.Ф. Каравай, В.С. Подлазов // Матер. 12-й мультиконф. по проблемам управления (МКПУ-2019, Дивноморское, Геленджик). – Ростов н/Д; – Таганрог: Южный фед. ун-т (Таганрог), 2019. – № 3. – С. 54–56.

8. Подлазов В.С. Расширение возможностей системной сети «Ангара» / В.С. Подлазов, М.Ф. Каравай // Труды 13-го Всерос. совещания по проблемам управления (ВСПУ XIII, Москва, 2019). – М.: ИПУ РАН, 2019. – С. 2440–2444.

9. Системная сеть с малым диаметром из малопортовых маршрутизаторов / М.Ф. Каравай, В.С. Подлазов // Управление большими системами: сборник трудов (ИФ РИНЦ 0,661). – 2015. – № 56. – С. 201–210.

10. Пат. 2402061 РФ, МПК G 06 F 7/00. Пространственная коммутационная среда (варианты) / О.Х. Ким, Я.С. Пеккер, А.И. Солдатов, Д.М. Ким (РФ). – № 2402061; заявл. 16.03.09; опубл. 20.10.10, Бюл. № 29. – 4 с.

11. Солдатов А.И. Технические и алгоритмические проблемы коммутации современной электроники / А.И. Солдатов, О.Х. Ким // Изв. высш. учеб. заведений. Физика. – 2010. – Т. 53, № 9/3. – С. 308–311.

12. Солдатов А.И. Программируемая коммутационная среда / А.И. Солдатов, А.Ю. Матросова, О.Х. Ким, А.А. Солдатов, М.А. Костина // Вестник Том. гос. ун-та. Управление, вычислительная техника и информатика. – 2020. – № 50. – С. 114–122.

13. Romanowicz E. The Hall marriage theorem / E. Romanowicz, A. Grabowski // Formalized Mathematics. – 2004. – Vol. 12, Iss. 3. – P. 315–320.

Солдатов Андрей Алексеевич

Канд. техн. наук, доцент ОЭИ НИ ТПУ
Ленина пр-т, 30, г. Томск, Россия, 634050,
Канд. техн. наук, доцент каф. УИ ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050,
ORCID: 0000-0003-0696-716X
Тел.: +7 (382-2) 60-62-97
Эл. почта: soldatov.88@bk.ru

Костина Мария Алексеевна

Канд. техн. наук, доцент ОЭИ НИ ТПУ
Ленина пр-т, 30, г. Томск, Россия, 634050,
Канд. техн. наук, доцент каф. УИ ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050,
ORCID: 0000-0000-2626-6002
Тел.: +7 (382-2) 70-17-38
Эл. почта: mashenkasoldatova@mail.ru

Нариманова Гуфана Нурабековна

Канд. техн. наук, доцент, зав. каф. УИ ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050,
ORCID: 0000-0002-0885-9480
Тел.: +7 (382-2) 70-17-37
Эл. почта: guftana@mail.ru

Soldatov A.I., Kim O.H., Soldatov A.A., Kostina M.A., Narimanova G.N.

Conflict-free, fault-tolerant, and compact programmable switching unit

The article provides a technical solution for building a switching unit. Its conflict-free nature, fault tolerance, and compactness are investigated. Using the example of a 16-pin switching unit, it is shown that the proposed structure allows performing electrical connections with 100% use of external pins. Two options for implementing 8 circuits on a 16-pin switching unit are given. The fault tolerance is shown. The proposed approach to building a switching unit allows to implement a switching unit of any size with a reduced number of switch elements.

Keywords: cross-bar, cross-router, programmable switching unit (PSU), conflict-free, fault tolerance, compact.

doi: 10.21293/1818-0442-2021-24-3-12-17

References

- Zhuravlev V., Nemytov A., A. Pershin [SpaceWire: a View from the side. Part 1]. *Modern Electronics*, 2017, no. 8, pp. 36–42 (in Russ.).
- Zhuravlev V., Nemytov A., A. Pershin [SpaceWire: a View from the side. Part 2]. *Modern Electronics*, 2017, no. 9, pp. 26–32 (in Russ.).
- Gorbunov S.F., Grishin V.Yu., Eremeev P.M. [Network interfaces of spacecraft: development prospects and implementation problems] *Nanoindustry*. Special issue 2019 (89). Moscow Technosfera, 2019, pp. 128–130 (in Russ.).
- Rajamony R., Arimilli L.B., Gildea K. PERCS: The IBM POWER7-IH high-performance computing system. *IBM Journal of Research and Development*, 2011, vol. 55, no. 3, pp. 3:1–3:12.
- Resch M., Sabine R., Benkert K., Galle M., Bez W., Kobayashi H., Hirayama T. High performance computing on vector system 2008. Springer, 2009. 241 p.
- The new development of Rostec allows you to create super-computers «at home». Available at: <https://sdelanounas.ru/blogs/109999/>, free. (Accessed: October 16, 2020) (in Russ.).

Солдатов Алексей Иванович

Д-р техн. наук, проф. отд. электронной инженерии (ОЭИ) Национального исследовательского Томского политехнического ун-та (НИ ТПУ) Ленина пр-т, 30, г. Томск, Россия, 634050, Д-р техн. наук, проф. каф. управления инновациями (УИ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР) Ленина пр-т, 40, г. Томск, Россия, 634050, ORCID: 0000-0003-1892-1644 Тел.: +7 (382-2) 60-62-97 Эл. почта: asoldatof@mail.ru

Ким Олег Хонбинович

Техник каф. УИ ТУСУРа
Ленина пр-т, 40, г. Томск, Россия, 634050,
Тел.: +7 (382-2) 71-17-38
Эл. почта: oh.kim@yandex.ru

7. Karavay M.F., Podlazov V.S. [Expanding the communication capabilities of the Angara network] / Materials of the 12th multi-conference on management problems (icpu–2019, Divnomorskoe, Gelendzhik). Rostov-on-Don – Taganrog: Southern Federal University (Taganrog), 2019, no 3, pp. 54–56 (in Russ.).

8. Podlazov V.S., Karavay M. F. [Expanding the capabilities of the angara system network]. *Proceedings of the 13th all-Russian Meeting on Management Problems* (VSPU XIII, Moscow, 2019). Moscow: IPU RAS, 2019, pp. 2440–2444 (in Russ.).

9. Karavay M. F., Podlazov V. S. [System network with a small diameter of small-port routers]. *Management of Large Systems: Collection of Works*. (If RSCI 0.661). 2015, no 56, pp. 201–210 (in Russ.).

10. Kim O.H., Pekker Ya.S., Soldatov A.I., Kim D.M. *Prostranstvennaya kommutacionnaia sreda (varianty)* [Spatial switching environment (variants)], patent RF, no. 2402061, 2009 (in Russ.).

11. Soldatov A. I., Kim O. H. [Technical and algorithmic problems of switching modern electronics]. *Proceedings of Higher Educational Institutions. Physics*, 2010, vol. 53, no 9/3. pp. 308–311 (in Russ.).

12. [Programmable switching environment]. Soldatov A. I., Matrosova A. Yu., Kim O. H., Soldatov A. A., Kostina M. A. *Bulletin of Tomsk State University. Management, Computing and Computer Science*, 2020, no 50, pp. 114–122 (in Russ.).

13. Romanovich E., Grabovsky A. Hall's marriage theorem. *Formalized Mathematics*. 2004, v.12, iss 3, pp. 315–320.

Oleg H. Kim

Technician, IM Department, TUSUR
40, Lenin pr., Tomsk, Russia, 634050
Phone: +7 (382-2) 70-17-38
Email: oh.kim@yandex.ru

Andrey A. Soldatov

Doctor of Science in Engineering,
Assistant Professor, EED, NR TPU
30, Lenin pr., Tomsk, Russia, 634050
Assistant Professor, IM Department, TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0003-0696-716X
Phone: +7 (382-2) 60-62-97
Email: soldatov.88@bk.ru

Mariy A. Kostina

Doctor of Science in Engineering,
Assistant Professor, EED, NR TPU
30 Lenin pr., Tomsk, Russia, 634050
IM Department, TUSUR,
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0000-2626-6002
Phone: +7 (382-2) 70-17-38
Email: mashenkasoldatova@mail.ru

Gufana N. Narimanova

Doctor of Science in Engineering,
Assistant Professor, IM Department, TUSUR
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0002-0885-9480
Phone: +7 (382-2) 70-17-37
Email: guftana@mail.ru

Alexey I. Soldatov

Doctor of Science in Engineering, Professor,
Electronic Engineering Division (EED),
National Research Tomsk Polytechnic University (NR TPU)
30 Lenin pr., Tomsk, Russia, 634050
Doctor of Science in Engineering, Professor,
Innovation Management (IM) Department, Tomsk State
University of Control Systems and Radioelectronics (TUSUR)
40, Lenin pr., Tomsk, Russia, 634050
ORCID: 0000-0003-1892-1644
Phone: +7 (382-2) 60-62-97
Email: asoldatov@mail.ru