Оптический матричный умножитель NxN с использованием серийных компонентов волоконной оптики

Шамплетов Н.О., аспирант, nshampletov@maltsystem.ru
Елизаров С.Г., к.ф.-м.н., selizar@maltsystem.ru

Удельное энергопотребление электронных вычислителей велико – оптические вычислители могут решить эту проблему

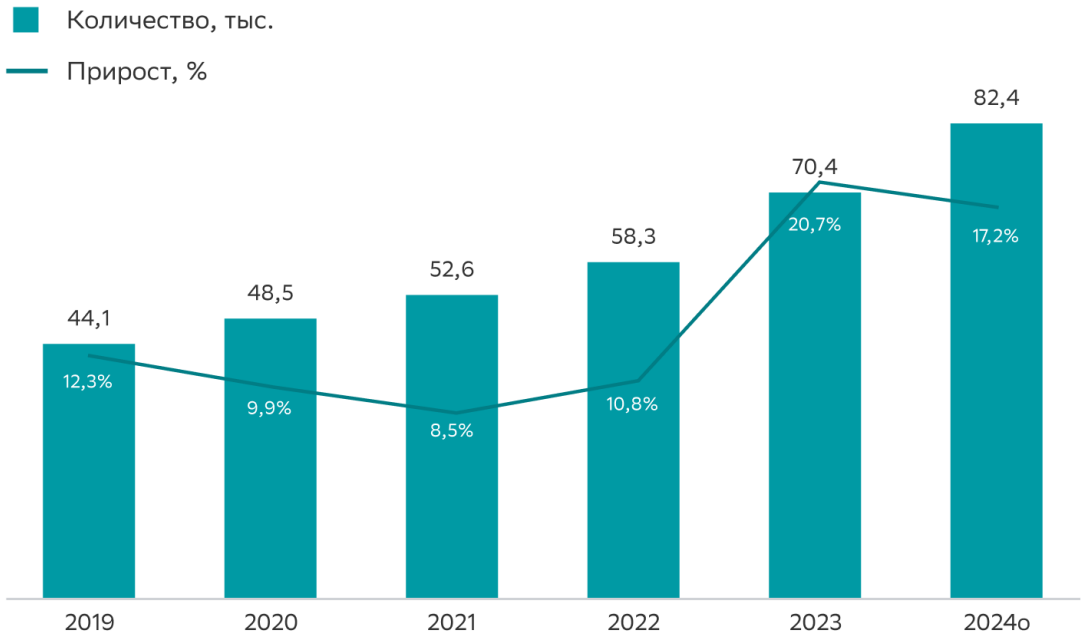
Число операций на 1 Вт потребляемой энергии у оптических вычислителей существенно выше*

Чип	Производительность	Энергоэффективность
Lin Z. et al. [1]	0.12 TOPS	153 TOPS/W
PACE [2]	8.19 TOPS	4.21 TOPS/W
Envise [4]	65.5 TOPS	40.9 TOPS/W (0.82)
TeMPO [5]	368.6 TOPS	23.3 TOPS/W
NVIDIA H100 [3]	~3500 TOPS*	0.71 TOPS/W

✓ Энергоэффективность оптических вычислителей на порядки выше

! По числу операций же электронные вычислители пока лидируют.

Растёт число ЦОД – экономия даже нескольких % энергопотребления даст ощутимый эффект



Динамика роста числа стойко-мест в России за 2019-2024гг

https://sber.pro/publication/rinok-tsod-v-rossii-perspektivi-razvitiya/

* - ссылки на источники см. в Приложении

Наши компетенции

Мальт Систем (MALTsystem) – российский центр проектирования специализированных цифровых и аналоговых СБИС и создания уникального электронного оборудования и модулей на их основе с 2011 года.

Направления деятельности



Интеллектуальная собственность



РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 243031

БЛОК ШИРОКОПОЛОСНОГО ГЕНЕРАТОРА СИГНАЛОВ ПРОИЗВОЛЬНОЙ ФОРМЫ С ОТКРЫТЫМ ИНТЕРФЕЙСОМ

Патентообладатель: *ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "МАЛЬТ СИСТЕМ" (ООО "МАЛЬТ СИСТЕМ") (RU)*

Авторы: *Елизаров Сергей Георгиевич (RU), Галдин Илья Викторович (RU), Монахов Александр Михайлович (RU), Иванов Дмитрий Максимович (RU), Барановский Алексей Олегович (RU), Бурмистров Игорь Александрович (RU)*

Заявка № 2025134449

Приоритет полезной модели 04 декабря 2025 г.

Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации 24 апреля 2026 г.

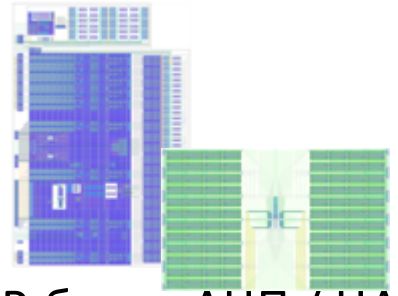
Срок действия исключительного права на полезную модель истекает 04 декабря 2035 г.

Руководитель Федеральной службы по интеллектуальной собственности Ю.С. Зубов

Технология DirectRF



СнК Rune



IP-блоки АЦП / ЦАП



модуль Alpha



MaltAWG / MaltDigi



Все компоненты оптического матричного вычислителя мы так или иначе пробовали у себя прототипировать.

MALTsystem поставила перед собой задачу разработать высокочастотный оптический матричный вычислитель

Для чего нужен?

Оптические ИИ-ускорители требуют матричных вычислителей

$$\vec{Y} = \hat{M} \bullet \vec{X}$$

Операция умножения вектора на матрицу

Наша мотивация

Готовых матричных оптических блоков нет в продаже – хотелось бы это исправить

Разработка всего ИИ-ускорителя требует от команд компетенций в большом количестве сфер - стоит дать им готовый матричный модуль

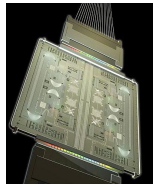
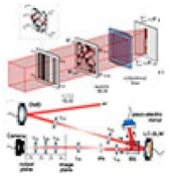
Разработка оптического вычислителя позволит нам расширить компетенции в перспективной сфере оптики

Ограничения

Использовать серийные оптоволоконные компоненты для обеспечения доступности вычислителя

Возможность работы на частотах до ~10 ГГц

Существует несколько платформ* для реализации оптических матричных вычислителей. Наш выбор – волоконная оптика

	Интегральные фотонные схемы (ФИС)	Умеренные потери Компактность	**Дорогостоящее сложное производство для мелких партий**
	Волоконная оптика	**Серийные компоненты** **Широкая доступность** **Дешевизна**	Большие оптические потери Большие габариты
	FreeSpace Optics[6]	Сборка из готовых компонентов Относительно простая реализация параллельных вычислений	**Необходимость дорогостоящего оборудования:** оптические столы, дорогие научные приборы

Для создания вычислителя мы выбрали волоконную оптику

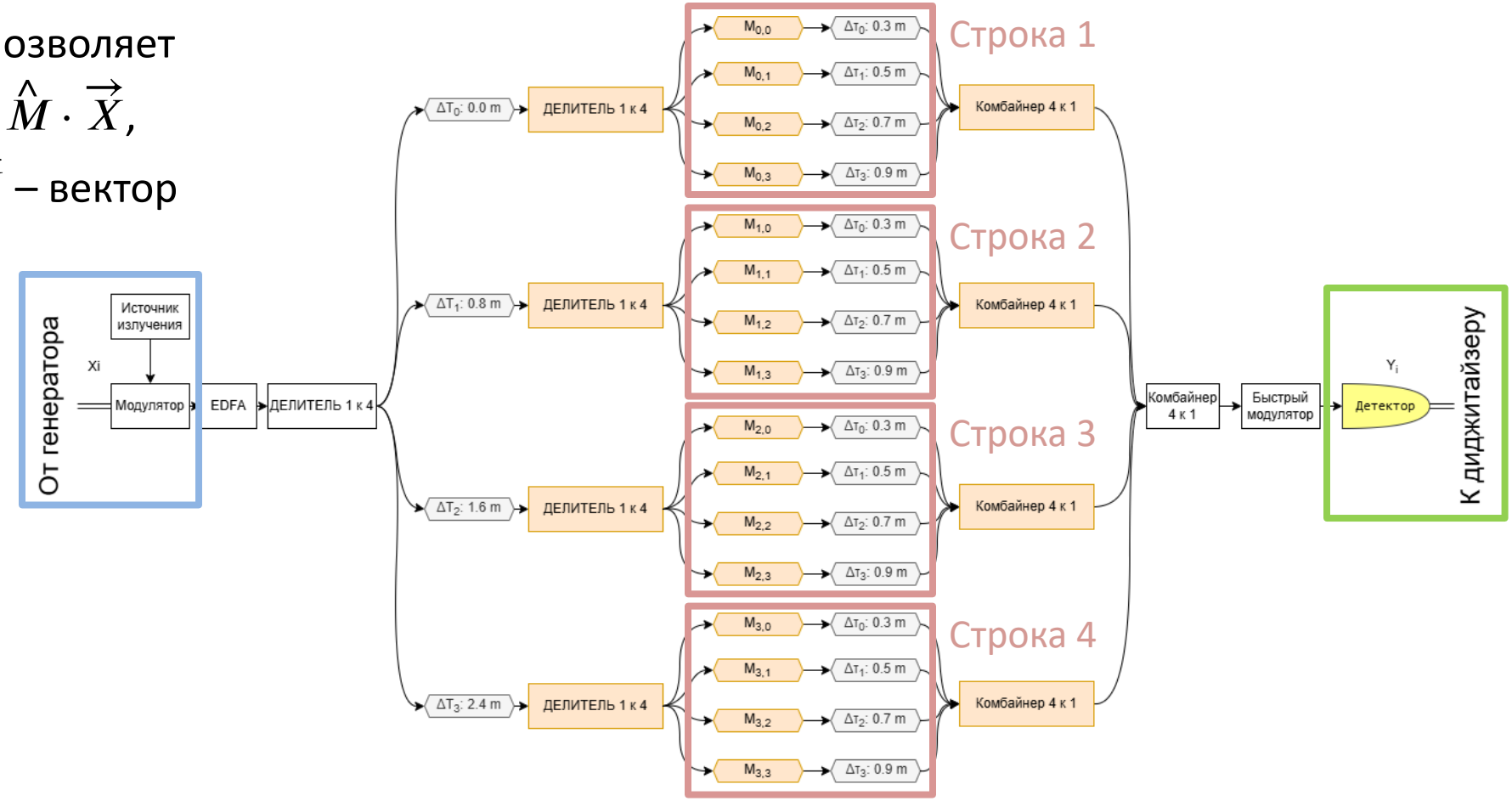
* - ссылки на источники см. в Приложении

** - рассмотрены не все, но наиболее популярные платформы

Архитектура вычислителя $N \times N$. Общее ознакомление

Предлагаемая архитектура позволяет реализовать операцию $\vec{Y} = \hat{M} \cdot \vec{X}$, где $\hat{M}$ – матрица $N \times N$, а $\vec{X}$ – вектор размерности N.

- Блок задания компонент вектора
- Строки матрицы
- Блок приёма результирующего вектора



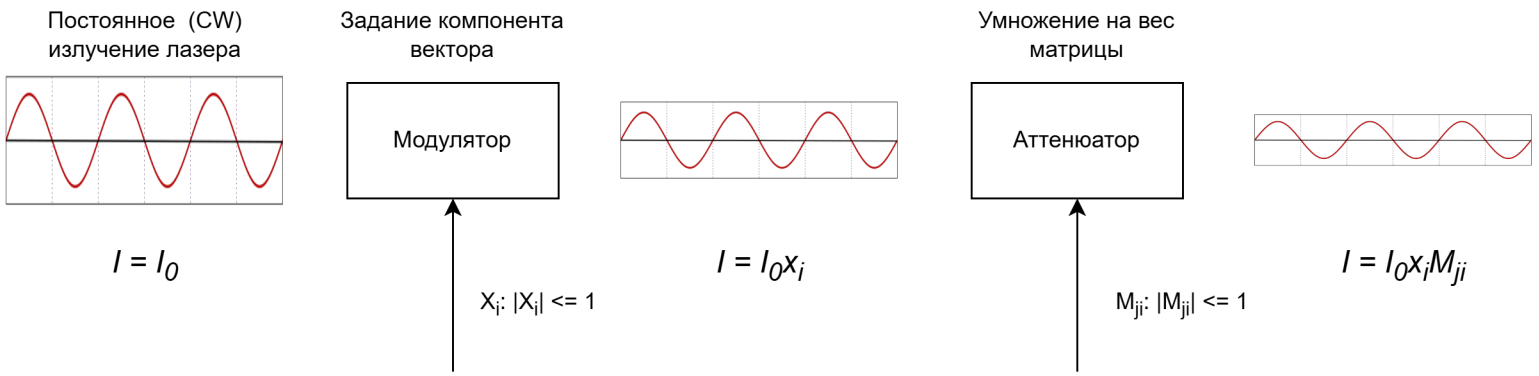
Архитектура MaltPTC для размерности N=4

Базовые принципы работы. «Принцип умножения»

Принятые ограничения

- Значения компонентов вектора кодируем интенсивностью
- Величины компонентов вектора задаём быстрым электрооптичесикм модулятором
- Веса матрицы – положением вала мех.аттенюатора

Схема умножения



Простота измерения: выходное напряжение фотоприёмника пропорционально интенсивности

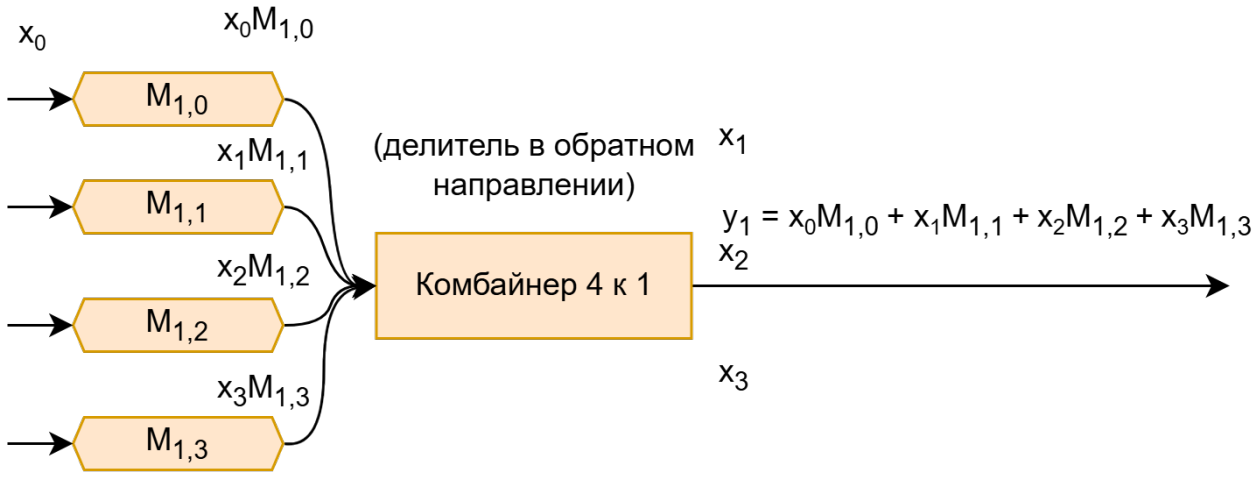
Только положительные значения

Базовые принципы работы. «Принцип сложения»

Принятые ограничения

- Считаем, что интенсивности складываются линейно
- Каждый компонент вектора, X_i, проходит через соответствующий аттенюатор с весом M_{ij}.
- Все параметры принципиально положительные, вычитание не реализуется.

Схема сложения

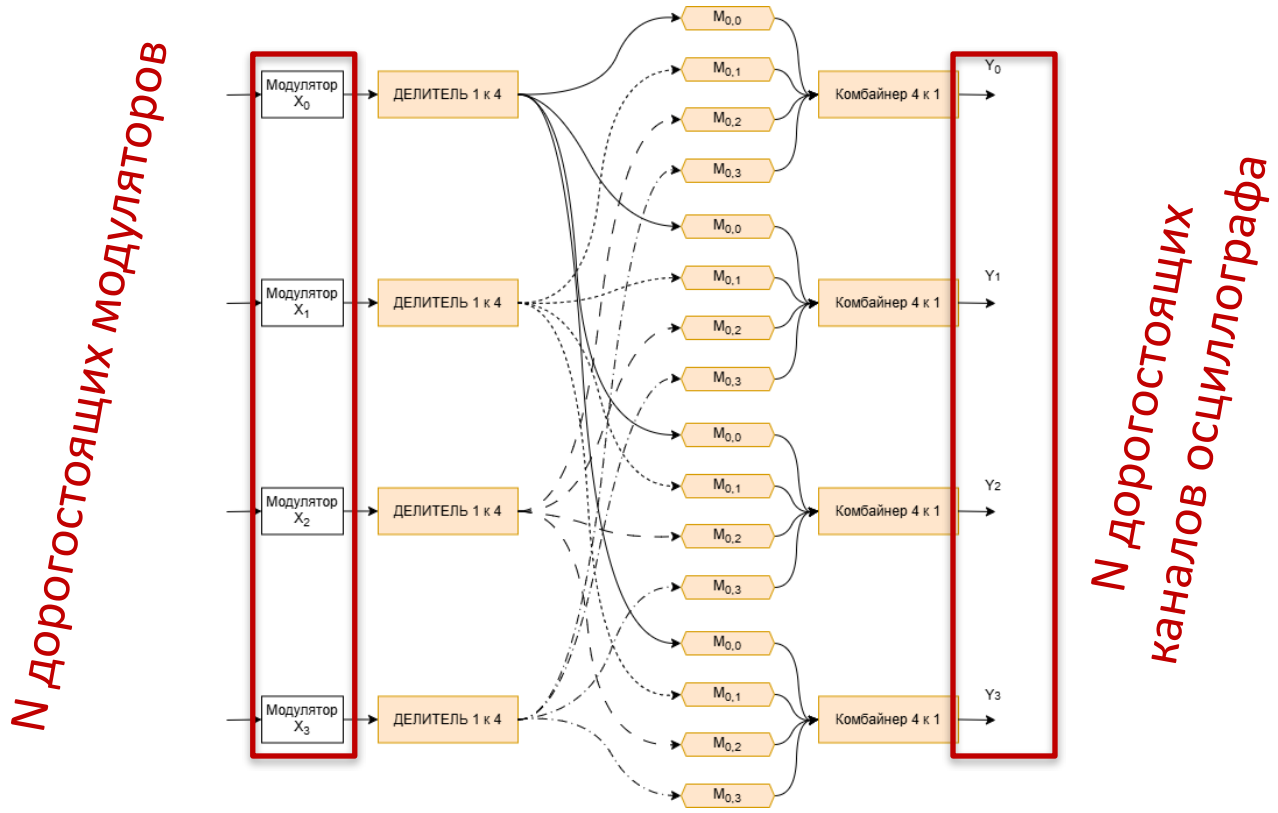


Сборка на широкодоступных и дешёвых делителях

Необходимо использовать некогерентный свет, чтобы избежать интерференции

Базовые принципы работы. «Принцип последовательного вычисления»

Проблема



Упрощённая архитектура для параллельных вычислений

Решение

Переход к последовательному вычислению:

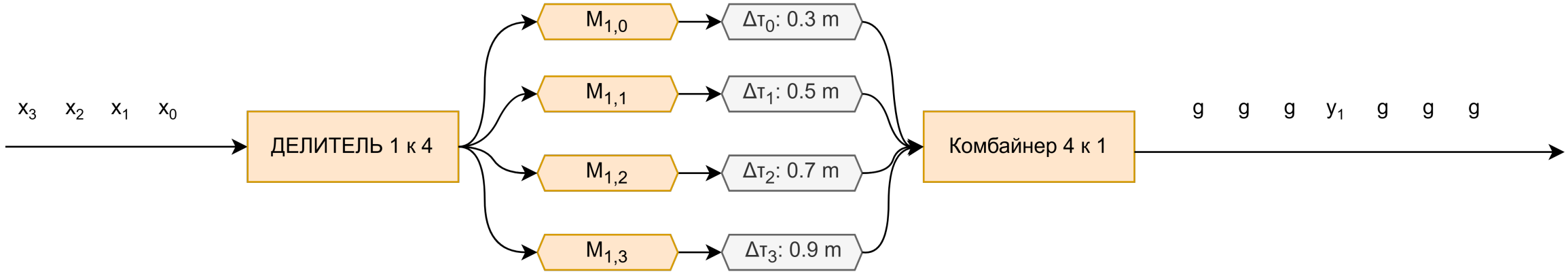
- 1 входной канал (1 быстрый модулятор)
- 1 выходной канал (1 быстрый канал осциллографа)

Вычисления существенно замедляются

В N раз уменьшается число источников и быстрых дорогостоящих модуляторов

Базовые принципы работы. «Принцип последовательного вычисления»

Схема последовательного вычисления

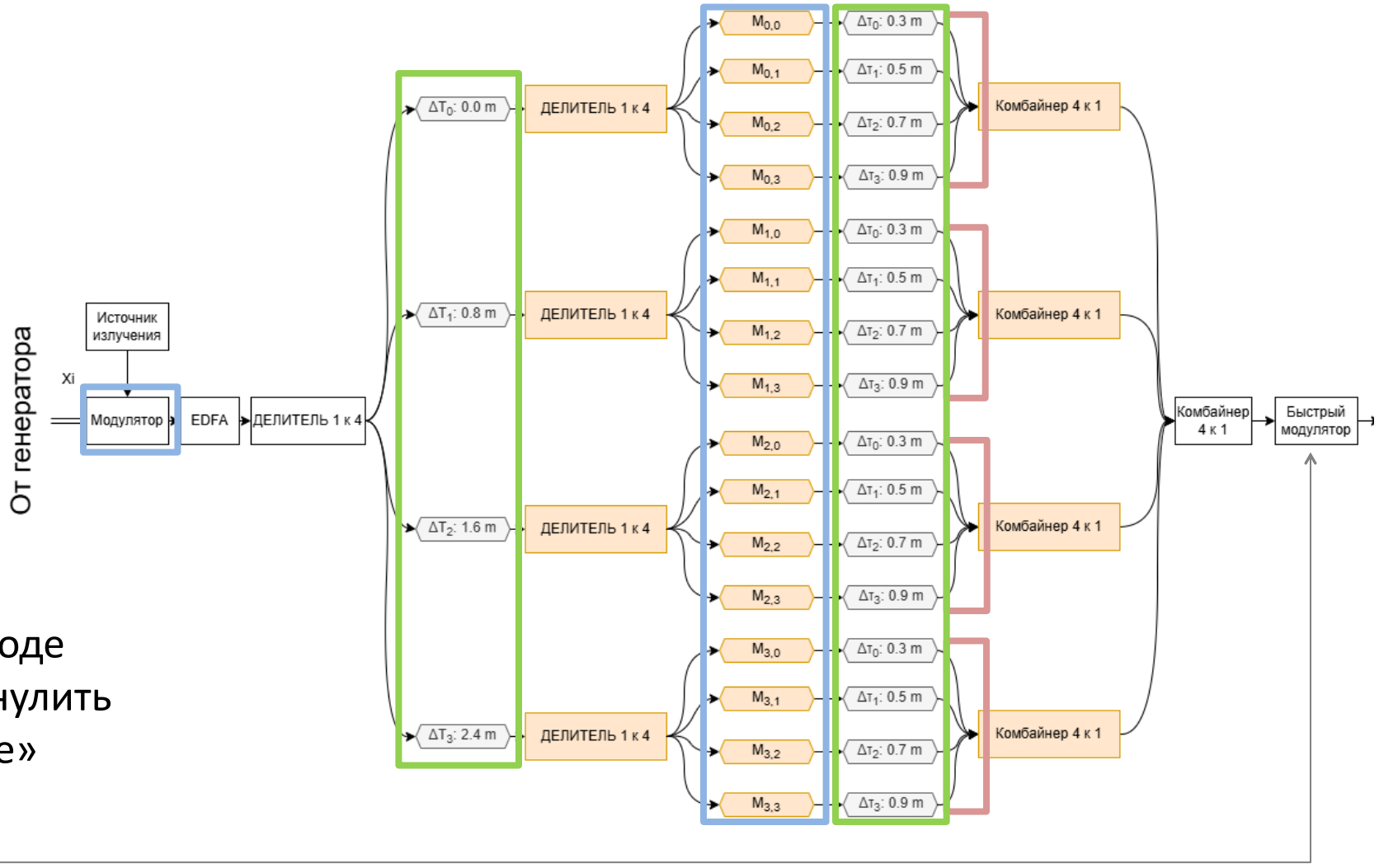


Канал	Пространственное распределение импульсов						
M1,0				x3 M1,0	x2 M1,0	x1 M1,0	x0 M1,0
M1,1			x3 M1,1	x2 M1,1	x1 M1,1	x0 M1,1	
M1,2		x3 M1,2	x2 M1,2	x1 M1,2	x0 M1,2		
M1,3	x3 M1,3	x2 M1,3	x1 M1,3	x0 M1,3			
На выходе комбайнера:	g	g	g	y1	g	g	g

Взгляд на архитектура вычислителя $N \times N$ после осознания принципов его работы

Участки применения принципов:

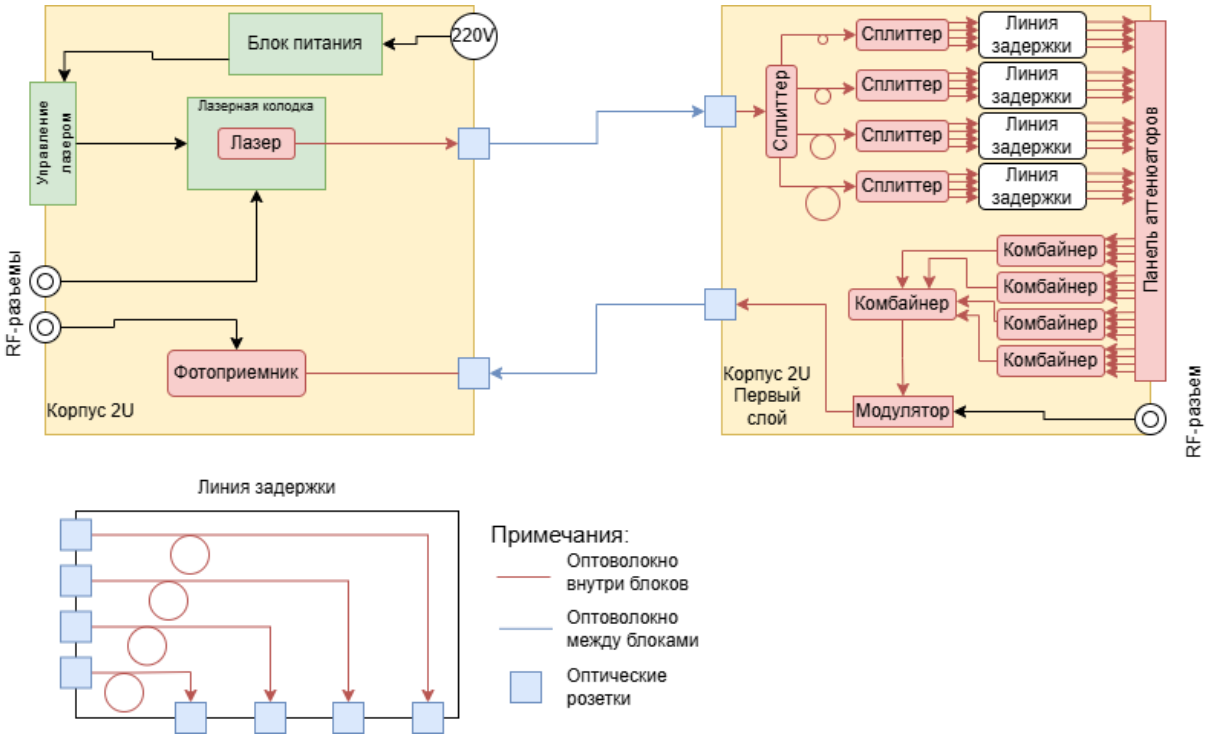
- Принцип умножения:
- Принцип сложения:
- Принцип последовательного вычисления:



Быстрый модулятор на выходе вычислителя позволяет обнулить промежуточные «мусорные» значения

Функциональная схема вычислителя 4x4

Схема MaltPTC 4x4



Характеристики MaltPTC 4x4

Оценка оптических потерь:

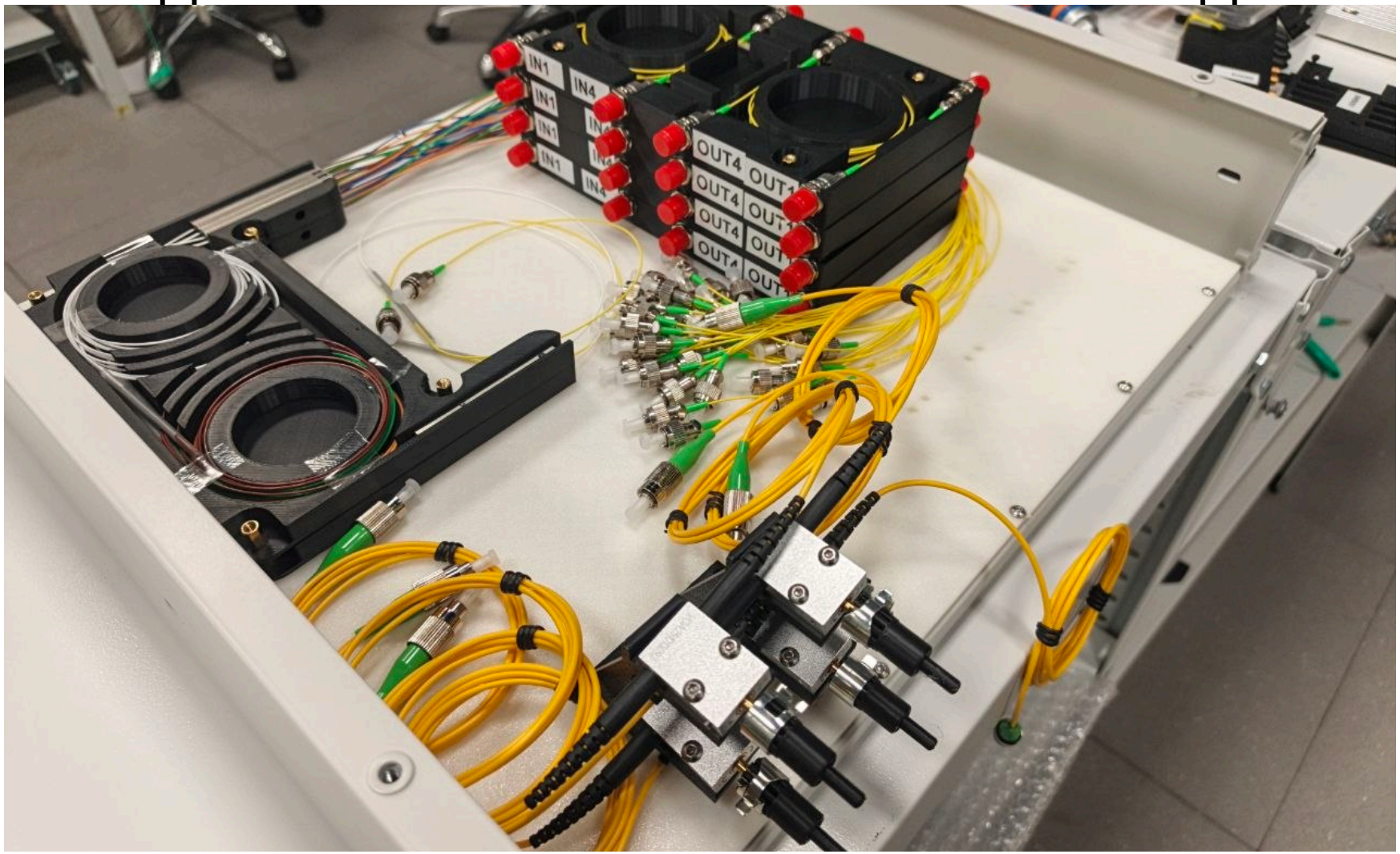
- Блока излучателя и приёмника – -7 дБ
- Блока матрицы M - -16 дБ

Итого: ~ -23 дБ

Для компенсации потерь можно использовать оптические усилители, G ~20 дБ

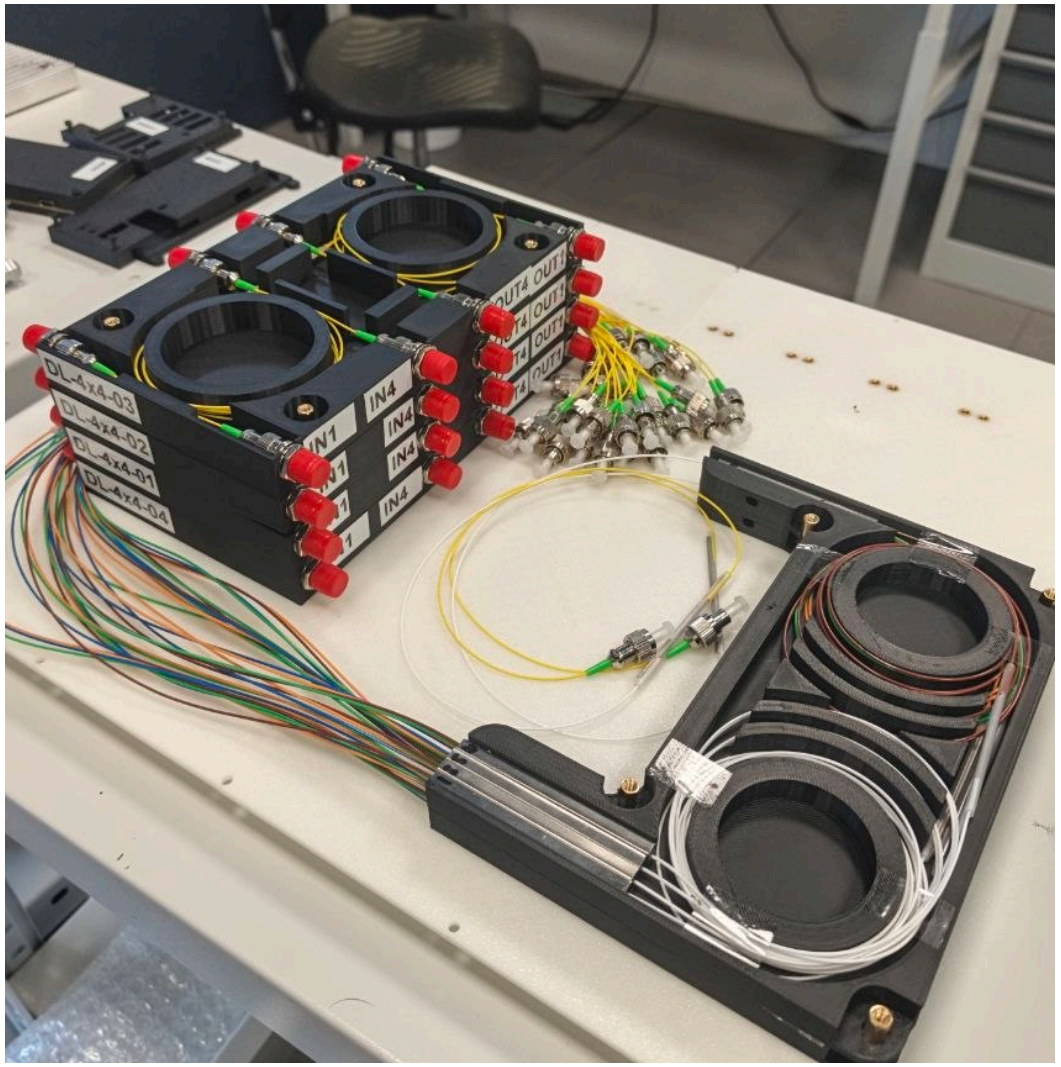
Скорость подачи воздействия – до 10 ГГц

Воплощение идеи вычислителя – оптический стенд.

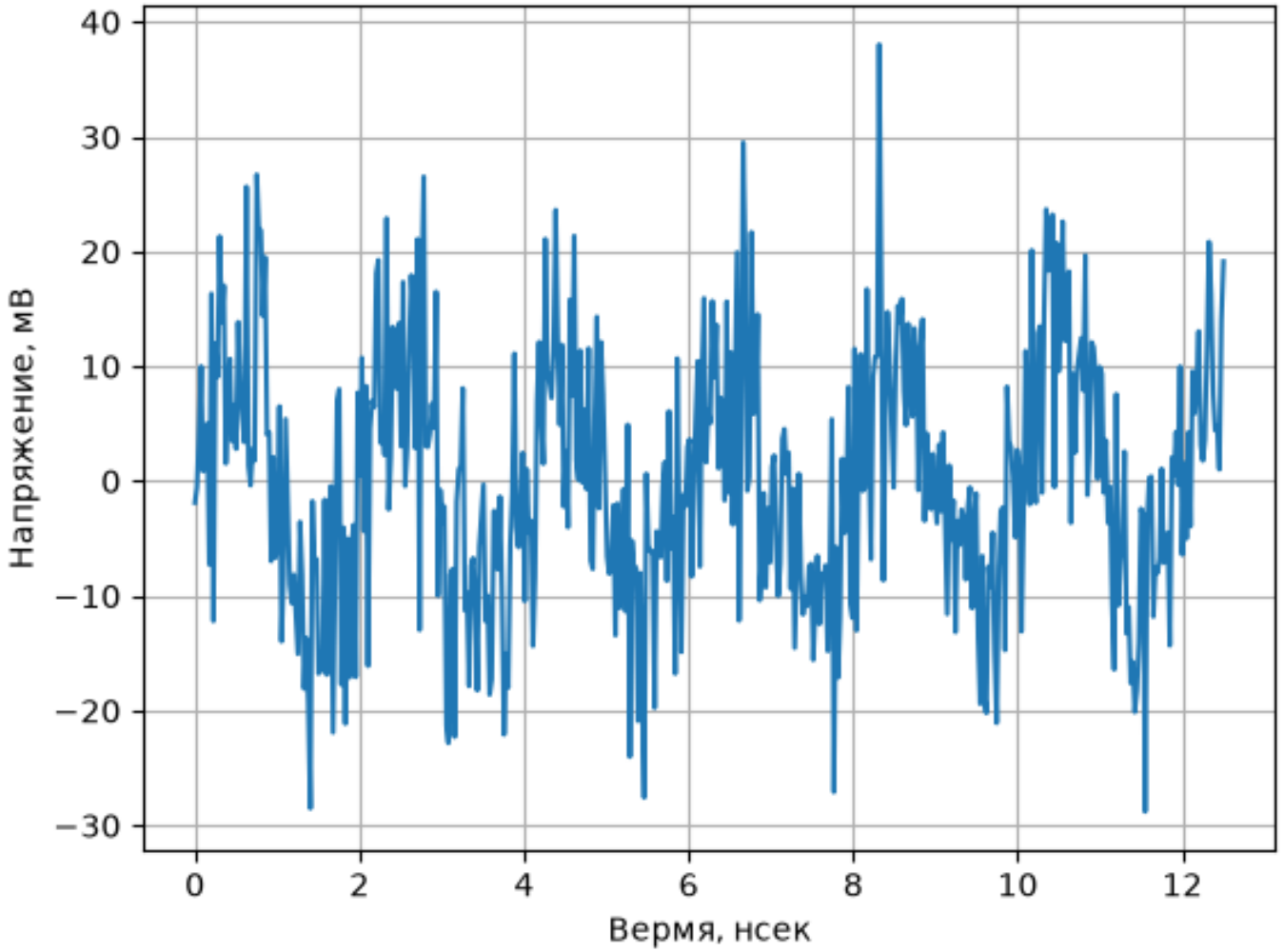


Воплощение идеи вычислителя – оптический стенд.





Без EDFA сигнал сильно затухает и затупляется

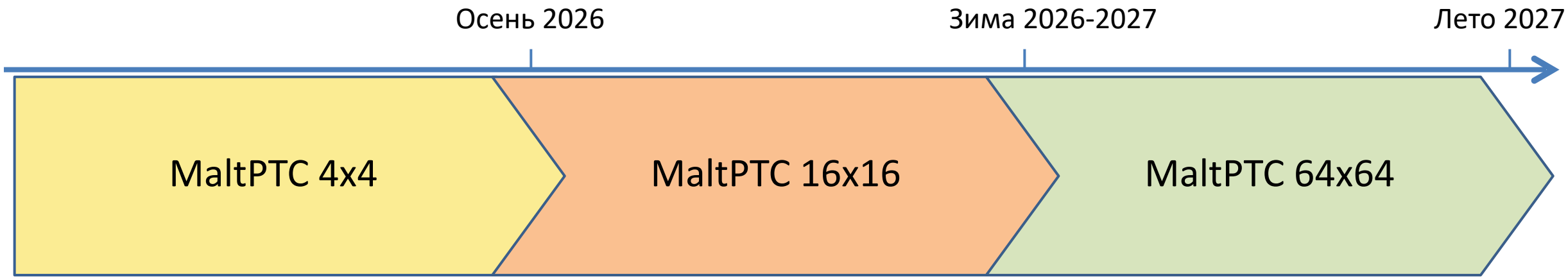


Использованное оборудование:

1. ROSA SD-R-FC-13-1.25G-4P-1-M
2. Генератор MaltAWG
3. Осциллограф Tektronix
4. Лазерный источник FLD5F20NP

Эксперимент показал, что оптический сигнал без использования усилителей (EDFA) становится слишком малым.

Наши планы по развитию направления – дойти до размерности за ближайшие несколько лет

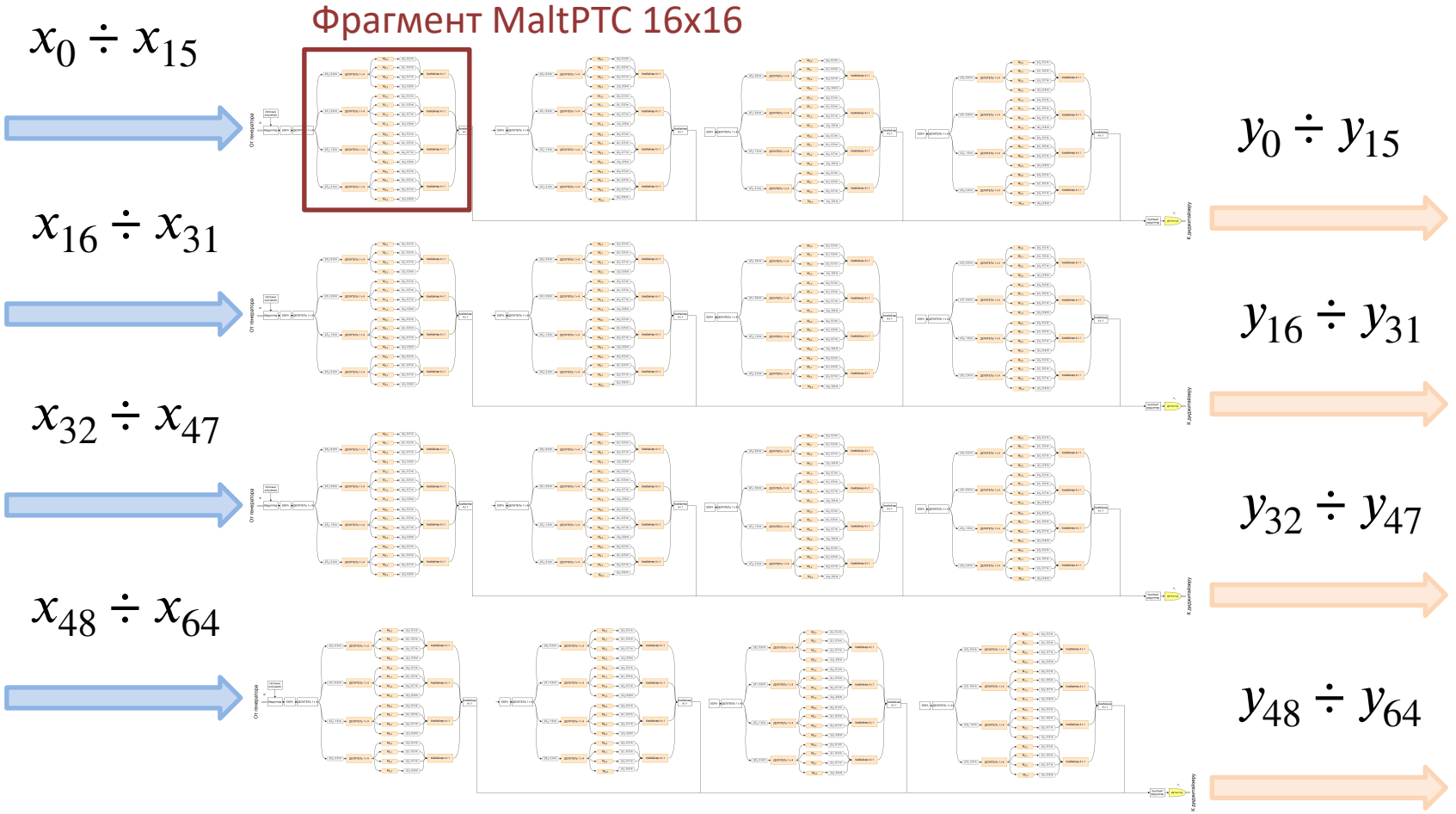


- Последовательные вычисления
- НЕ разреженная матрица

- Последовательные вычисления
- Разреженная матрица

- Последовательно-параллельные вычисления
- Разреженная матрица

О «последовательно-параллельной» сборке MaltPTC 64x64



Архитектура MaltPTC для размерности N=64

Вычислитель большей размерности можно строить на базе вычислителей меньшей размерности

Выводы

1. Показана принципиальная реализуемость вычислений последовательным образом
2. Обоснован переход к последовательным вычислениям в связи с сокращением цены производства
3. Потери в системе велики, но компенсируются серийными волоконными усилителями
4. Сколько такой вычислитель будет стоить? Бесплатно - после обкатки мы выложим чертежи всех деталей. Если будет сложно - поможем!

Наши контакты



ООО «Мальт Систем»

maltsystem.ru
info@maltsystem.ru

тел.: +7(495)133-6248

Научный руководитель
к.ф.-м.н., Елизаров
Сергей Георгиевич

selizar@maltsystem.ru

Список используемой литературы

1. Lin Z. et al. 120 GOPS Photonic tensor core in thin-film lithium niobate for inference and in situ training //Nature Communications. – 2024. – Т. 15. – №. 1. – С. 9081.
2. Hua S. et al. An integrated large-scale photonic accelerator with ultralow latency //Nature. – 2025. – Т. 640. – №. 8058. – С. 361-367.
3. https://www.nvidia.com/en-eu/data-center/h100/
4. Ahmed S. R. et al. Universal photonic artificial intelligence acceleration //Nature. – 2025. – Т. 640. – №. 8058. – С. 368-374.
5. Zhang M. et al. Tempo: Efficient time-multiplexed dynamic photonic tensor core for edge ai with compact slow-light electro-optic modulator //Journal of Applied Physics. – 2024. – Т. 135. – №. 22.
6. https://doi.org/10.1364/OL.401675