

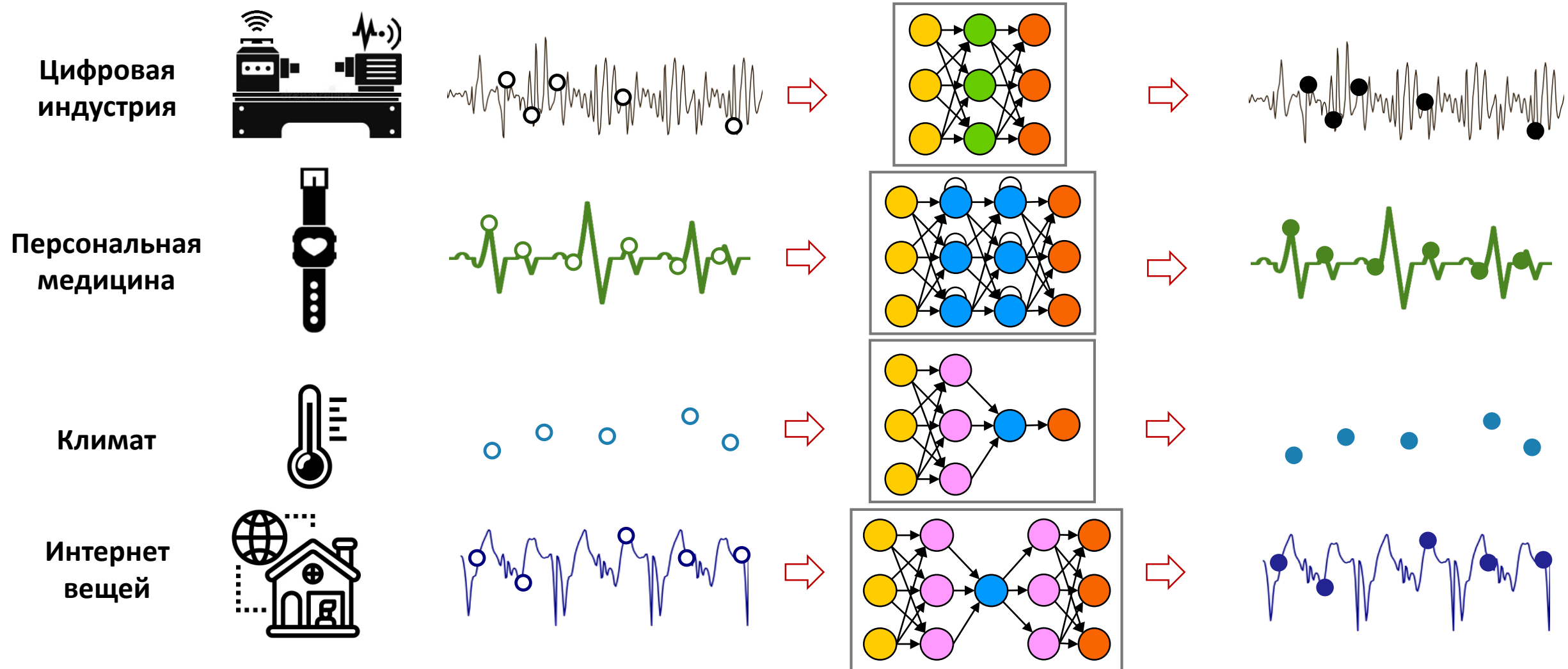
Международная научная конференция
Параллельные вычислительные технологии 2023
Санкт-Петербург, 28-30 марта 2023

Параллельное вычисление функции потерь на графическом процессоре для нейросетевых моделей восстановления временных рядов

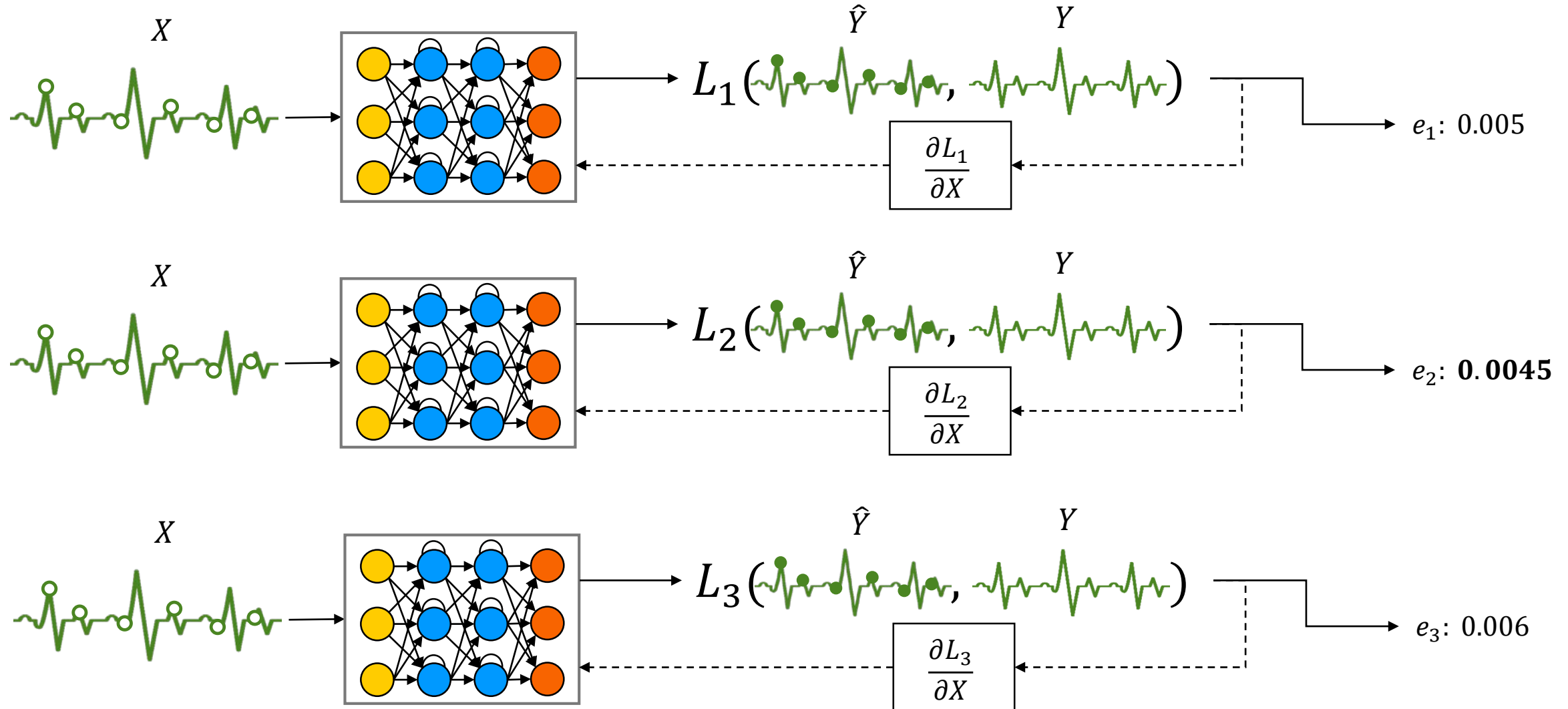
А.А. Юртин, М.Л. Цымблер

Южно-Уральский государственный университет, Челябинск

Восстановление временных рядов нейронными сетями



Обучение нейронной сети

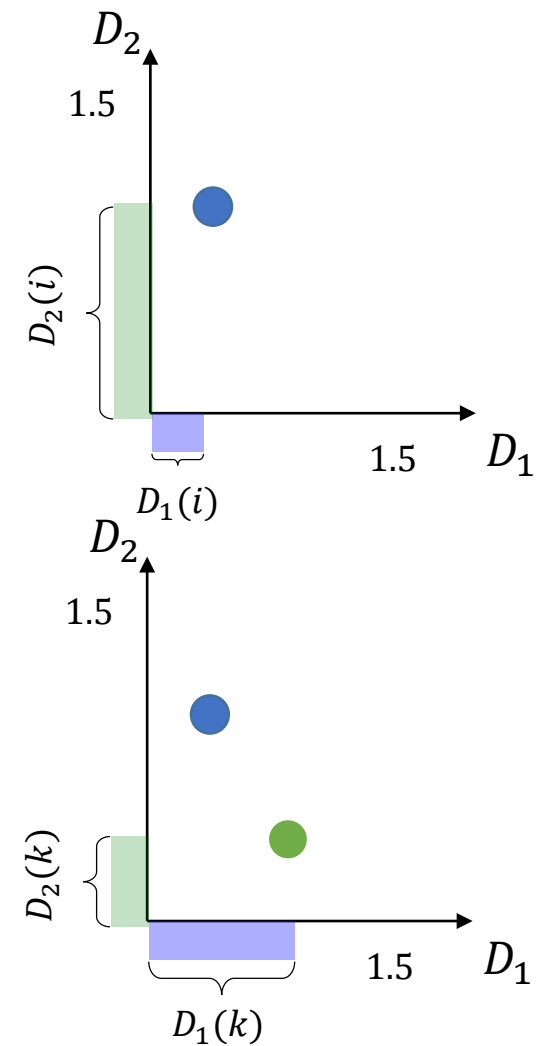
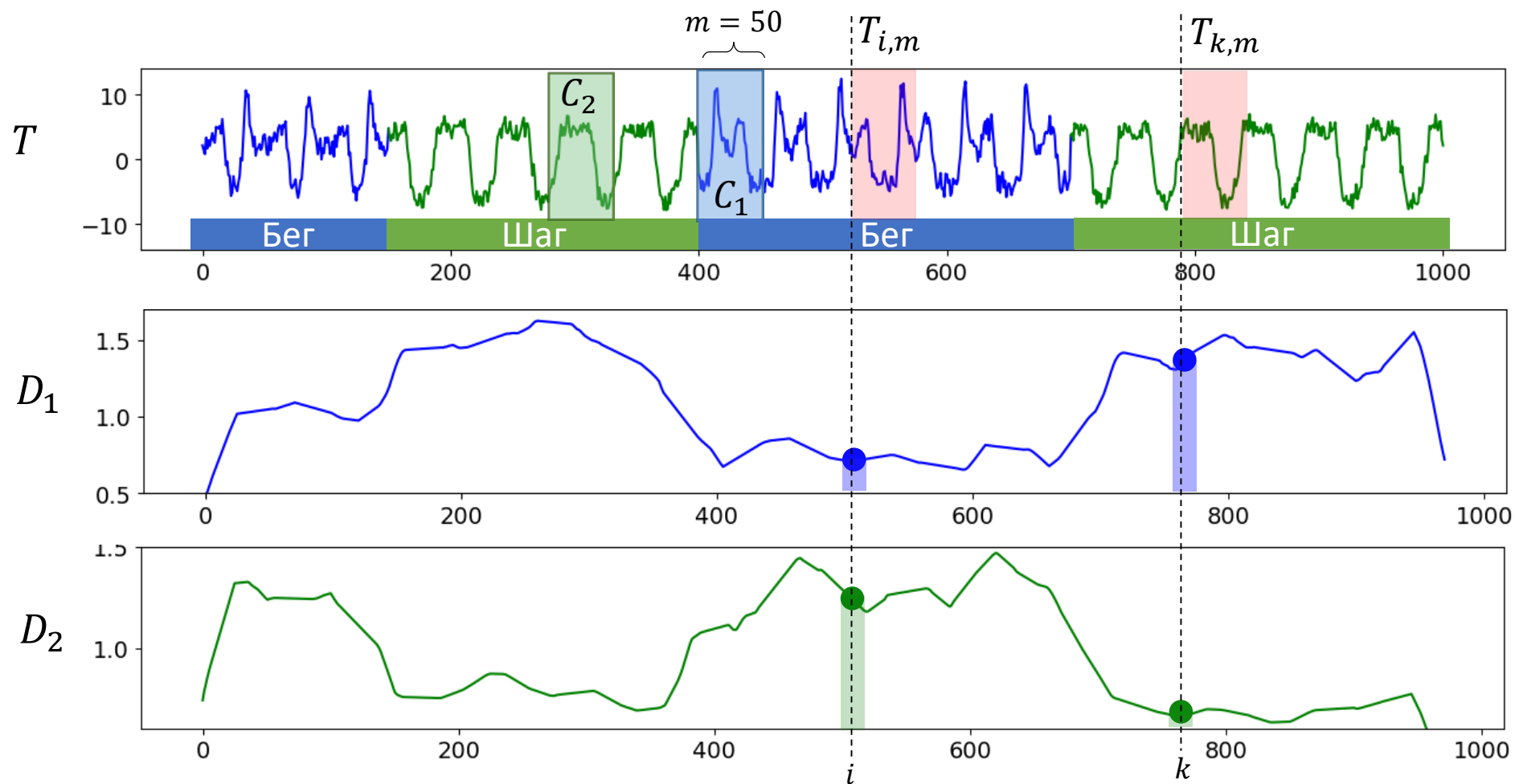


Обзор функций потерь

Функция	Формула	Сложность	Цель оптимизации
Mean square error	$\text{MSE} = \frac{1}{n} \sum_{i=0}^n (\hat{y}_i - y_i)^2$	$O(n)$	Уменьшение среднего значения квадрата ошибки между прогнозируемыми и истинными точками
Mean absolute error	$\text{MAE} = \frac{1}{n} \sum_{i=0}^n \hat{y}_i - y_i $	$O(n)$	Уменьшение модуля ошибки между прогнозируемыми и истинными точками
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{\sum_{i=0}^n (\bar{y}_i - y_i)^2}$	$O(n)$	Уменьшение отношения среднего отклонения прогнозируемых точек к среднему значению истинных точек
Root Mean Square error	$\text{RMSE} = \sum_{i=0}^n \sqrt{\frac{1}{n} (\hat{y}_i - y_i)^2}$	$O(n)$	Уменьшение значения корня среднего значения квадрата ошибки между прогнозируемыми и истинными точками
Dynamic time warping*	$\text{DTW}(\hat{Y}, Y) = d(m, m)$ $d(i, j) = (\hat{y}_i - y_i)^2 + \min \begin{cases} d(i-1, j) \\ d(i, j-1) \\ d(i-1, j-1) \end{cases}$	$O(n^2)$	Уменьшение пути временной трансформации подпоследовательности прогнозируемых точек в подпоследовательность истинных точек

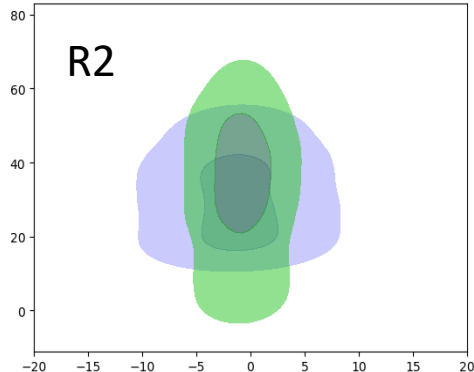
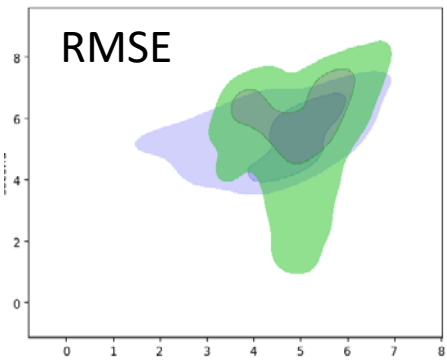
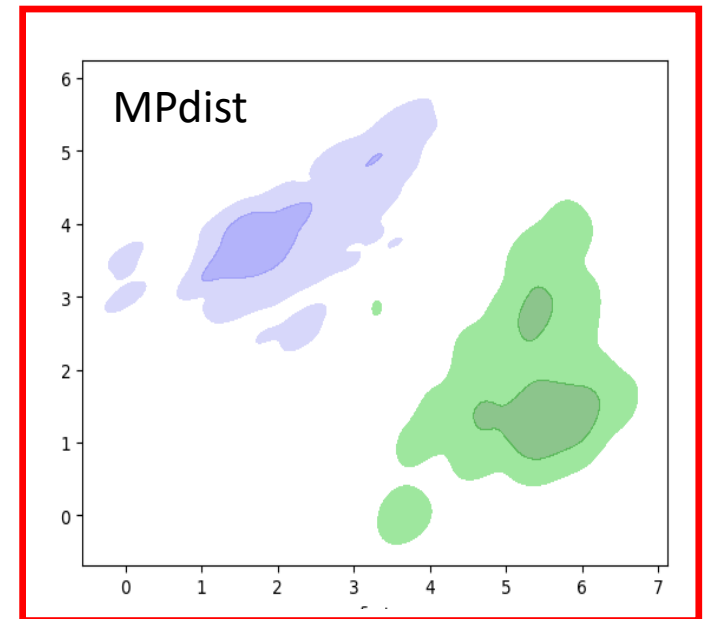
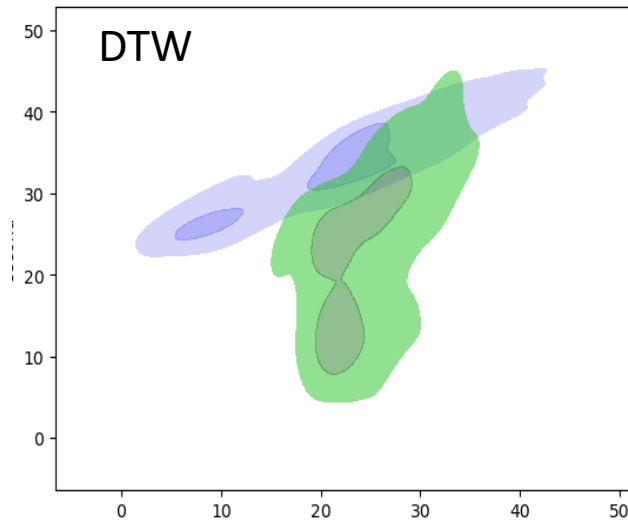
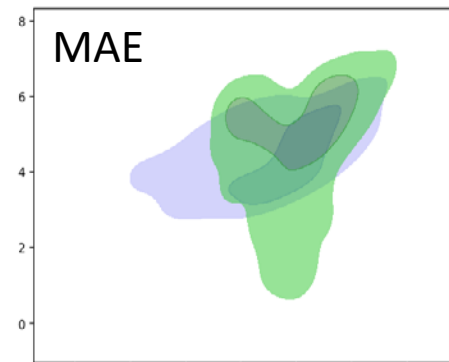
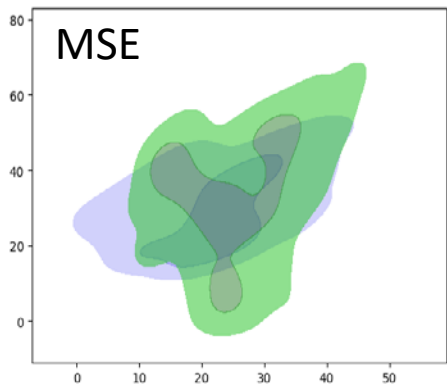
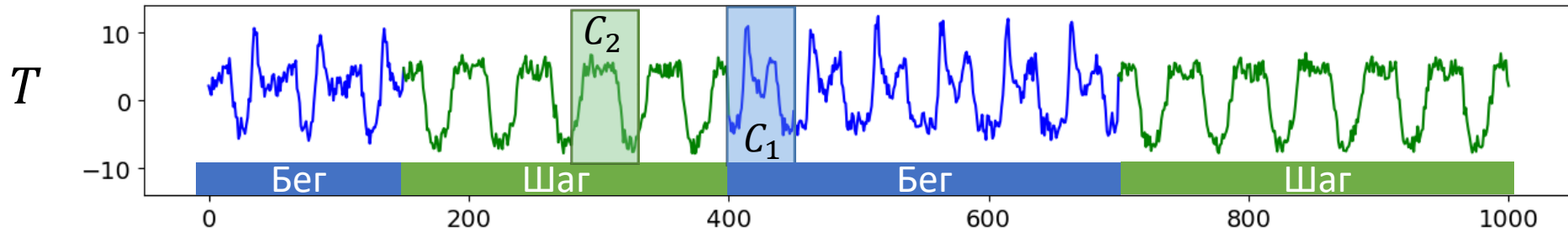
*Maghomi M., et al. DeepNAG: Deep Non-Adversarial Gesture Generation. 26th ICIUI. 2020. P. 213—223. <https://doi.org/10.1145/3397481.3450675>

Не все функции потерь одинаковы...



$$D_j(i) = \text{Dist}(T_{i,m}, C_j)$$

Не все функции потерь одинаковы...

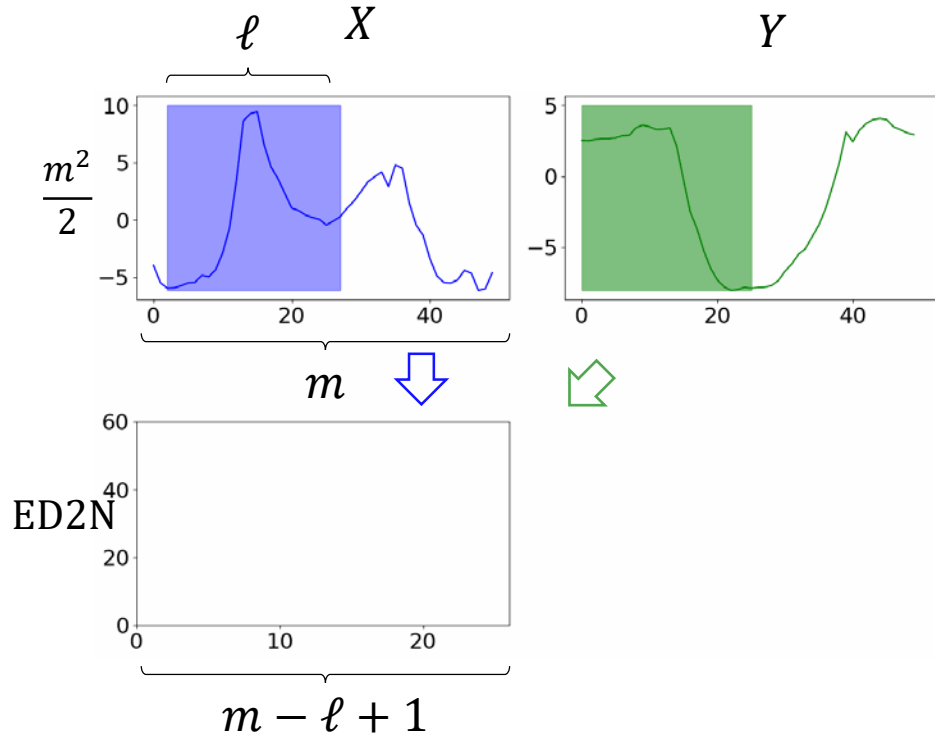


*Gharghabi S., Imani S., Bagnall A.J., Darvishzadeh A., Keogh E.J.: Matrix Profile XII: MPdist: A Novel Time Series Distance Measure to Allow Data Mining in More Challenging Scenarios. ICDM 2018: 965-970. DOI: 10.1109/ICDM.2018.00119. <https://doi.org/10.1145/3397481.3450675>

Новая функция потерь: $MP\ell_{\text{oss}}$

Функция	Формула	Сложность	Цель оптимизации
Mean square error	$\text{MSE} = \frac{1}{n} \sum_{i=0}^n (\hat{y}_i - y_i)^2$	$O(n)$	Уменьшение среднего значения квадрата ошибки между прогнозируемыми и истинными точками
Mean absolute error	$\text{MAE} = \frac{1}{n} \sum_{i=0}^n \hat{y}_i - y_i $	$O(n)$	Уменьшение модуля ошибки между прогнозируемыми и истинными точками
Coefficient of determination	$R^2 = 1 - \frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{\sum_{i=0}^n (\bar{y}_i - y_i)^2}$	$O(n)$	Уменьшение отношения среднего отклонения прогнозируемых точек к среднему значению истинных точек
Root Mean Square error	$\text{RMSE} = \sum_{i=0}^n \sqrt{\frac{1}{n} (\hat{y}_i - y_i)^2}$	$O(n)$	Уменьшение значения корня среднего значения квадрата ошибки между прогнозируемыми и истинными точками
Dynamic time warping	$\text{DTW}(\hat{Y}, Y) = d(m, m)$ $d(i, j) = (\hat{y}_i - y_j)^2 + \min \begin{cases} d(i-1, j) \\ d(i, j-1) \\ d(i-1, j-1) \end{cases}$	$O(n^2)$	Уменьшение пути временной трансформации подпоследовательности прогнозируемых точек в подпоследовательность истинных точек
Matrix profile loss	$\text{MP}\ell_{\text{oss}}(\hat{Y}, Y) = ?$	$O(n^3)$	Уменьшение отличия между подпоследовательностью прогнозируемых точек от подпоследовательность истинных точек по мере MPdist

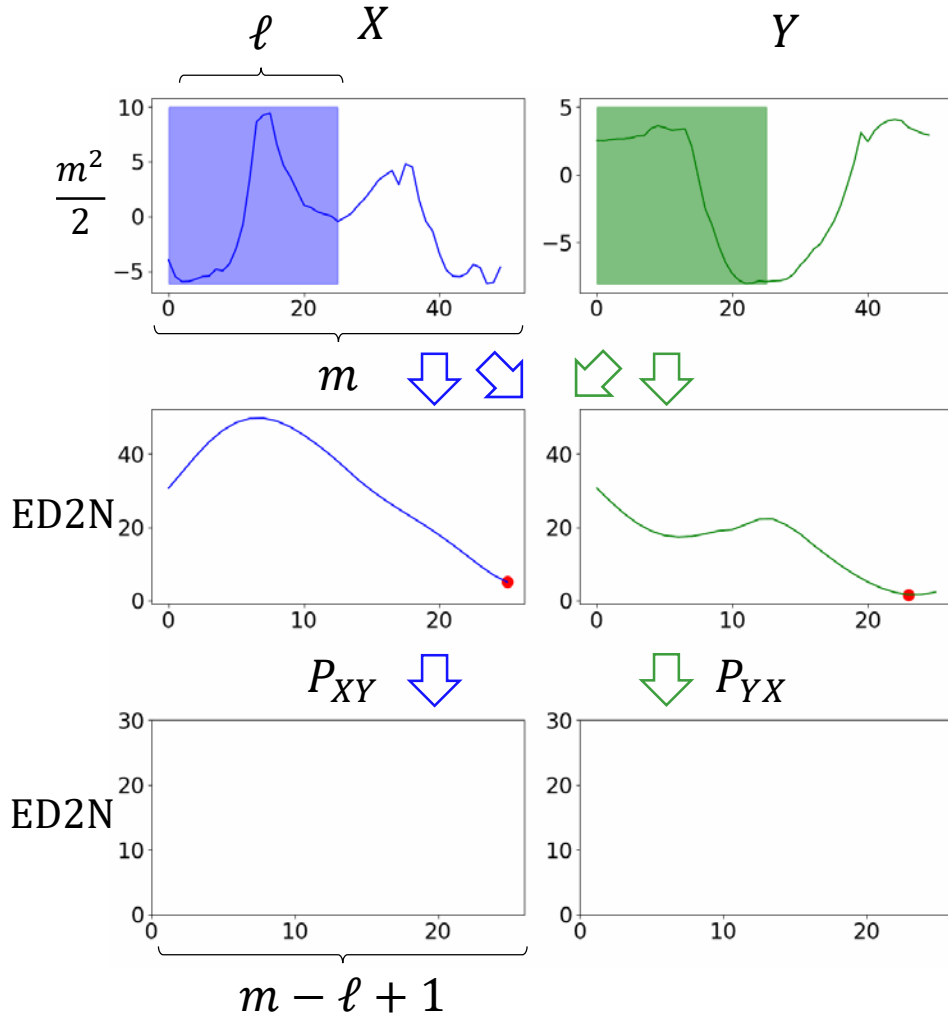
MP loss: формальные обозначения



$$\text{ED2N}(X, Y) = \sum_{i=1}^{\ell} (\hat{x}_i - \hat{y}_i)^2$$

$$\hat{x}_i = \frac{x_i - \mu_x}{\sigma_x} \quad \mu_x = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i \quad \sigma_x = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i^2 - \mu_x^2$$

MP loss: формальные обозначения



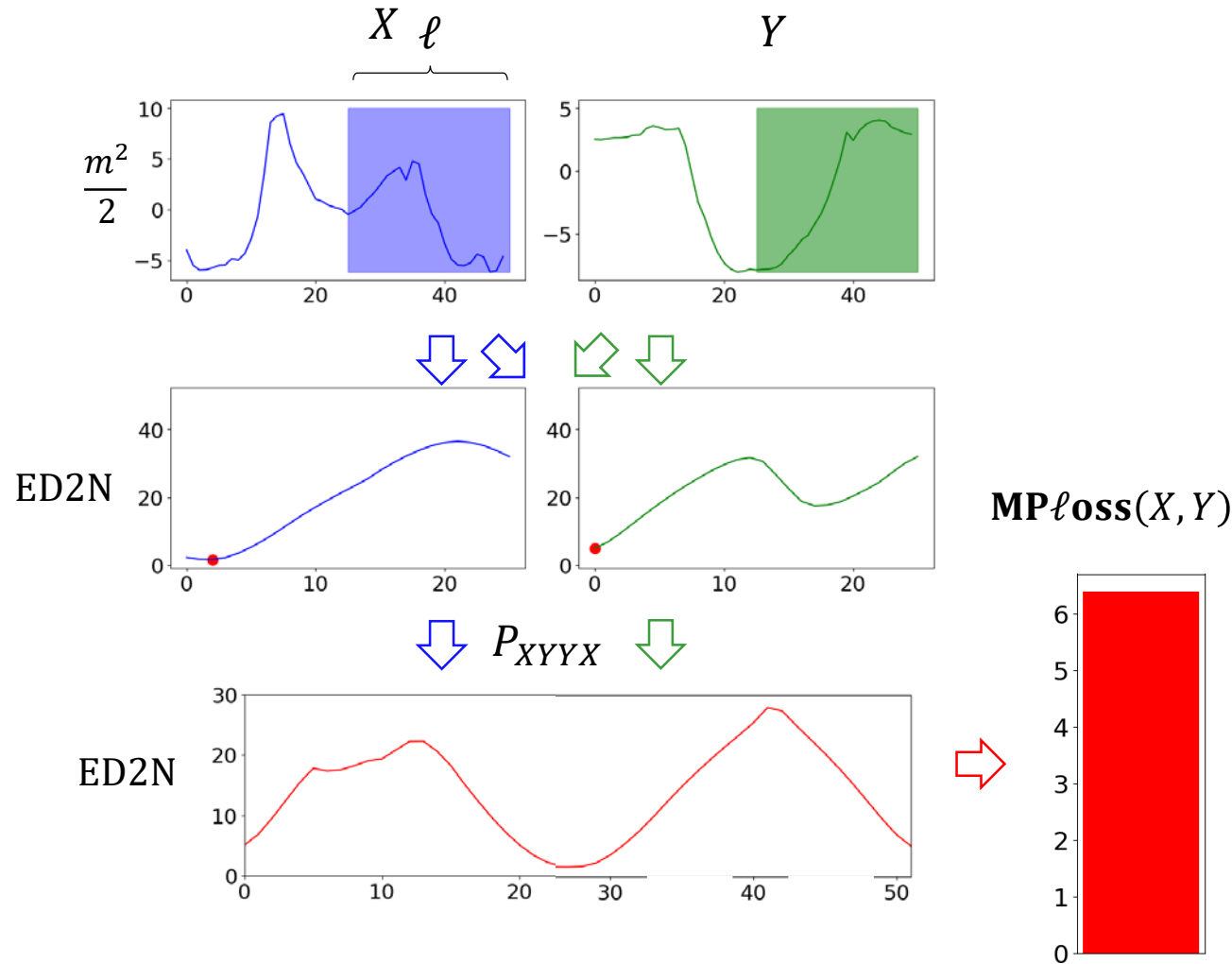
$$\text{ED2N}(X, Y) = \sum_{i=1}^{\ell} (\hat{x}_i - \hat{y}_i)^2$$

$$\hat{x}_i = \frac{x_i - \mu_x}{\sigma_x} \quad \mu_x = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i \quad \sigma_x = \sqrt{\frac{1}{\ell} \sum_{i=1}^{\ell} x_i^2 - \mu_x^2}$$

$$\{P_{XY}(i) = \text{ED2N}(X_{i,\ell}, Y_{j,\ell})\}_{i=1}^{m-\ell+1},$$

$$B_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} \text{ED2N}(X_{i,\ell}, Y_{j,\ell})$$

MPloss: формальные обозначения



$$ED2N(X, Y) = \sum_{i=1}^{\ell} (\hat{x}_i - \hat{y}_i)^2$$

$$\hat{x}_i = \frac{x_i - \mu_x}{\sigma_x} \quad \mu_x = \frac{1}{\ell} \sum_{i=1}^{\ell} x_i \quad \sigma_x = \sqrt{\frac{1}{\ell} \sum_{i=1}^{\ell} x_i^2 - \mu_x^2}$$

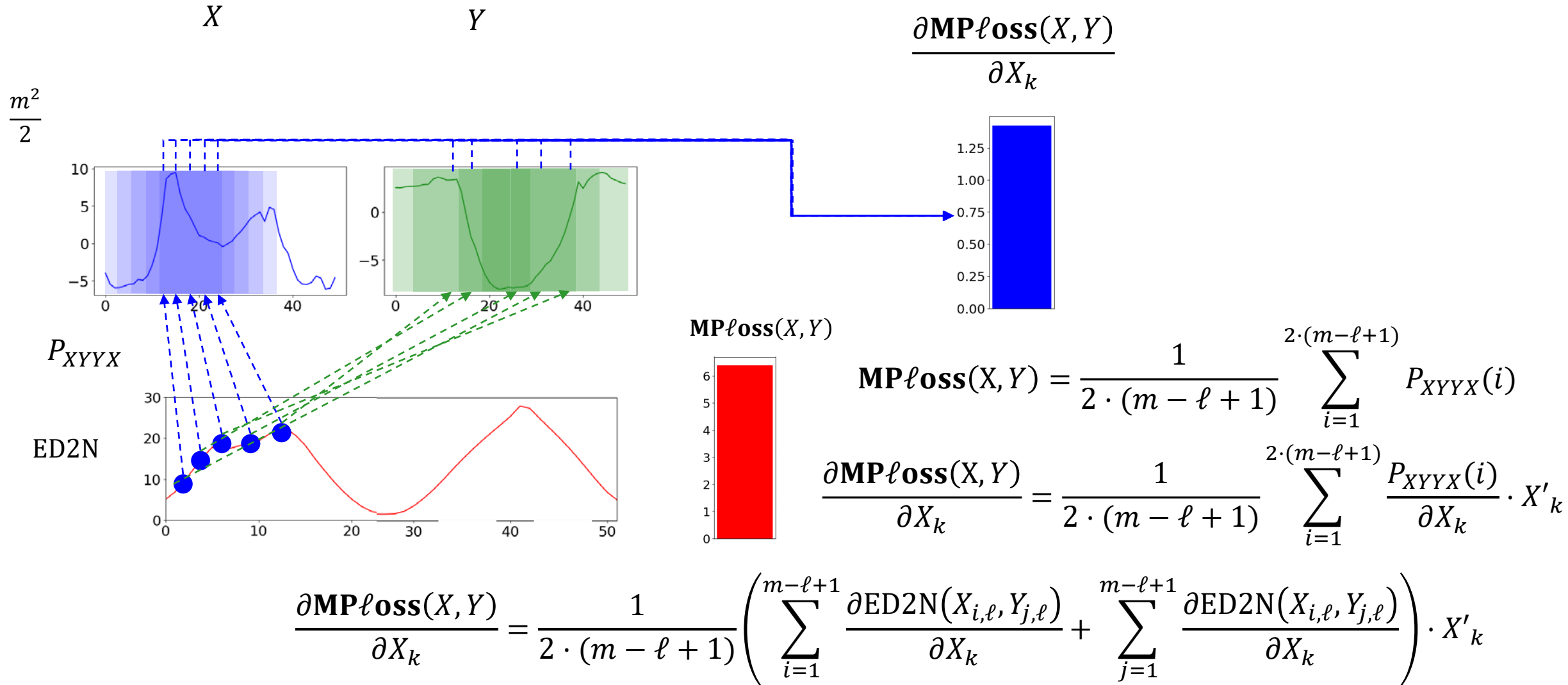
$$\{P_{XY}(i) = ED2N(X_{i,\ell}, Y_{j,\ell})\}_{i=1}^{m-\ell+1},$$

$$B_{j,\ell} = \arg \min_{1 \leq q \leq m-\ell+1} ED2N(X_{i,\ell}, Y_{j,\ell})$$

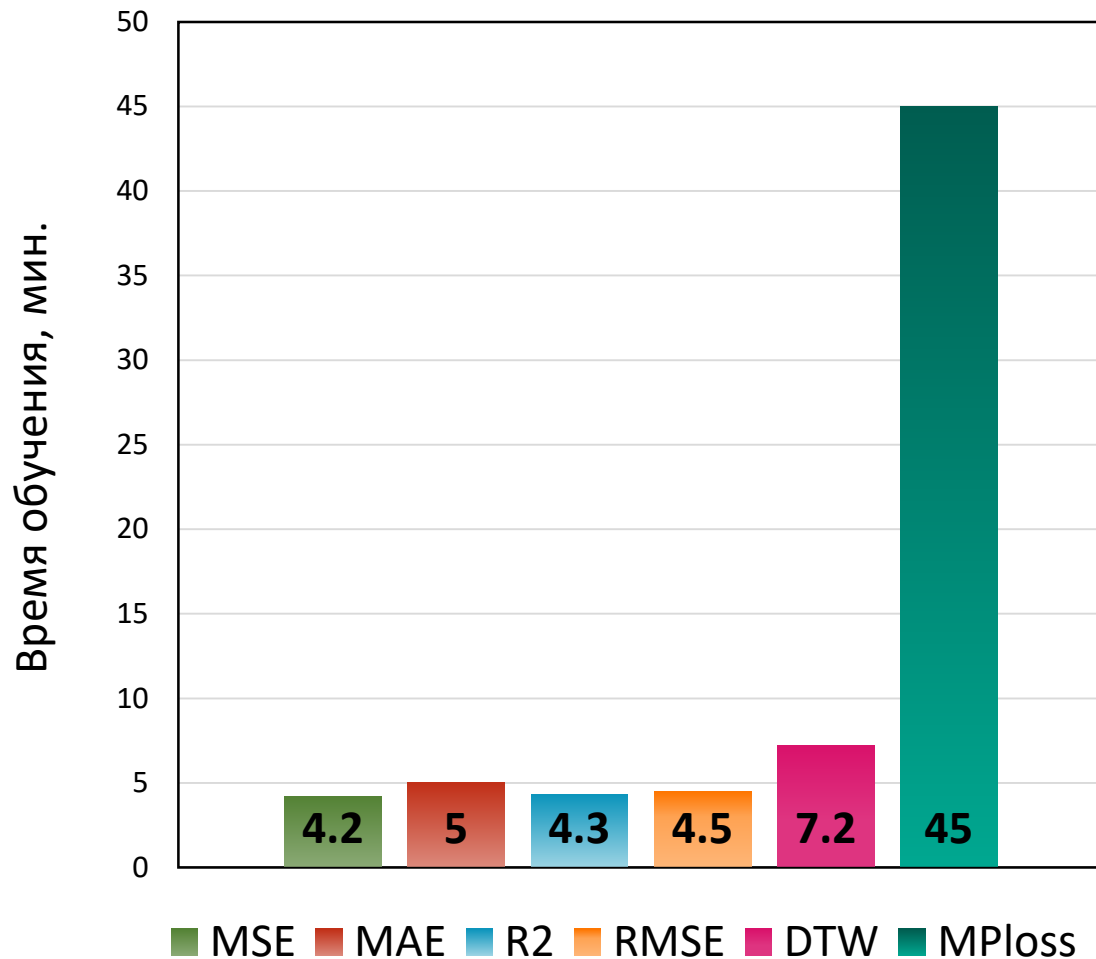
$$P_{XYYX} = P_{XY} \cdot P_{YX}$$

$$MPloss(X, Y) = \frac{1}{2 \cdot (m - \ell + 1)} \sum_{i=1}^{2 \cdot (m - \ell + 1)} P_{XYYX}(i)$$

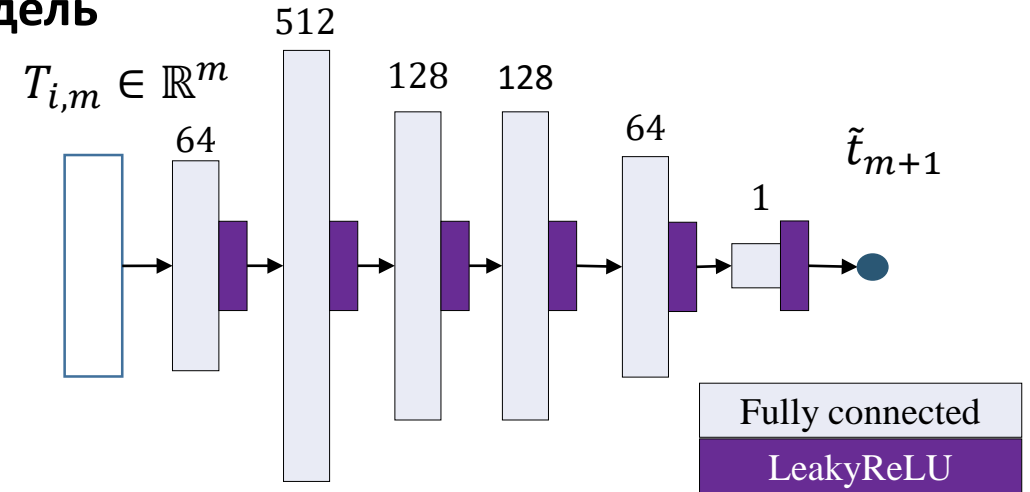
MPloss: Формальное определение



MpLoss: проблема последовательной реализации



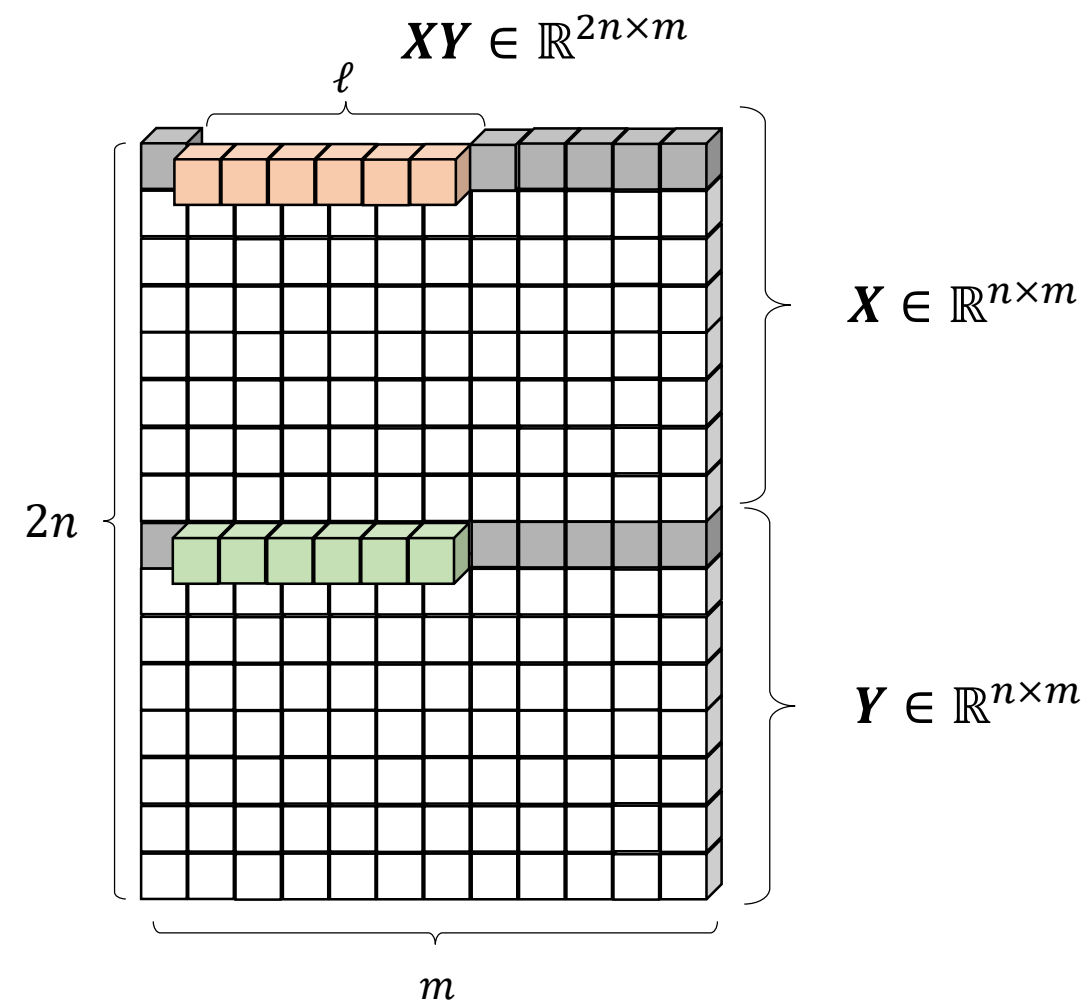
Модель



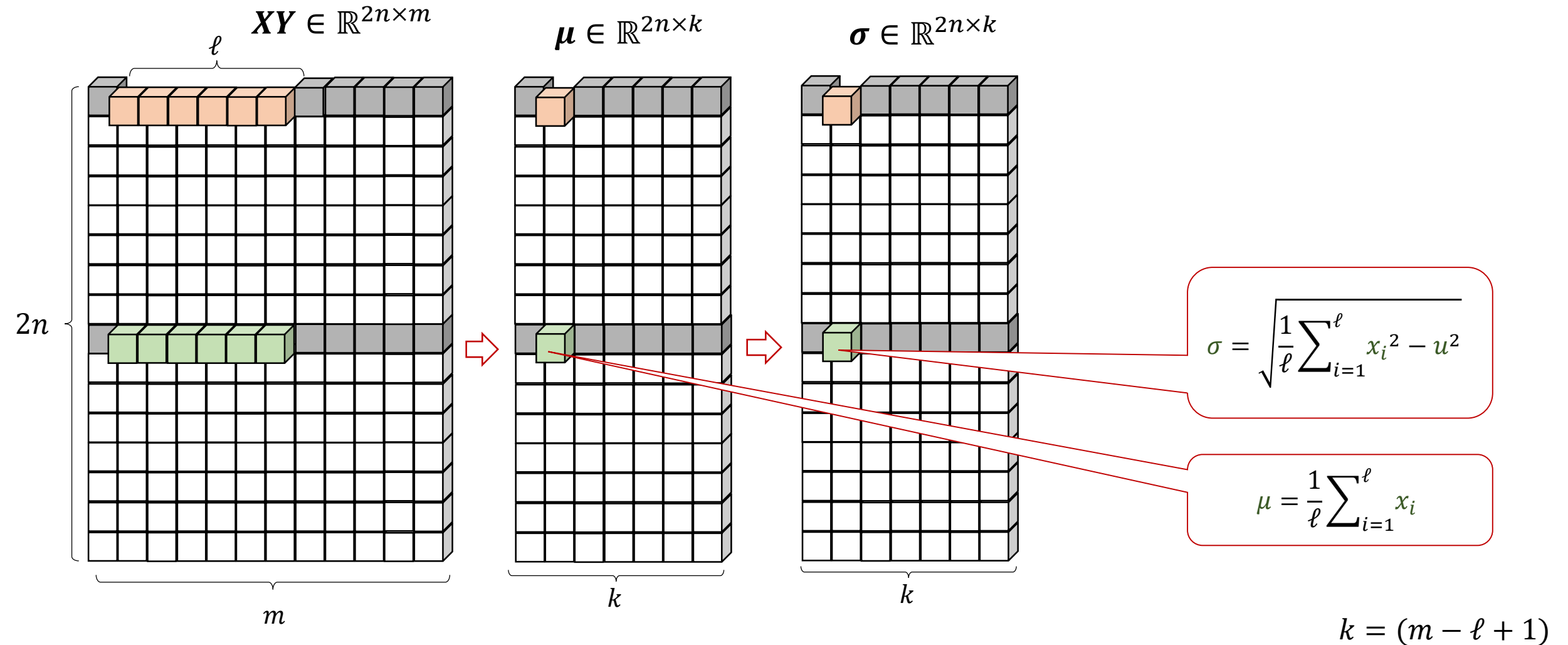
Параметры обучения

Параметр	Значение
Длина подпоследовательности, m	64
количество эпох	100
Размер партии (batch size)	64
Длина ряда	10 000

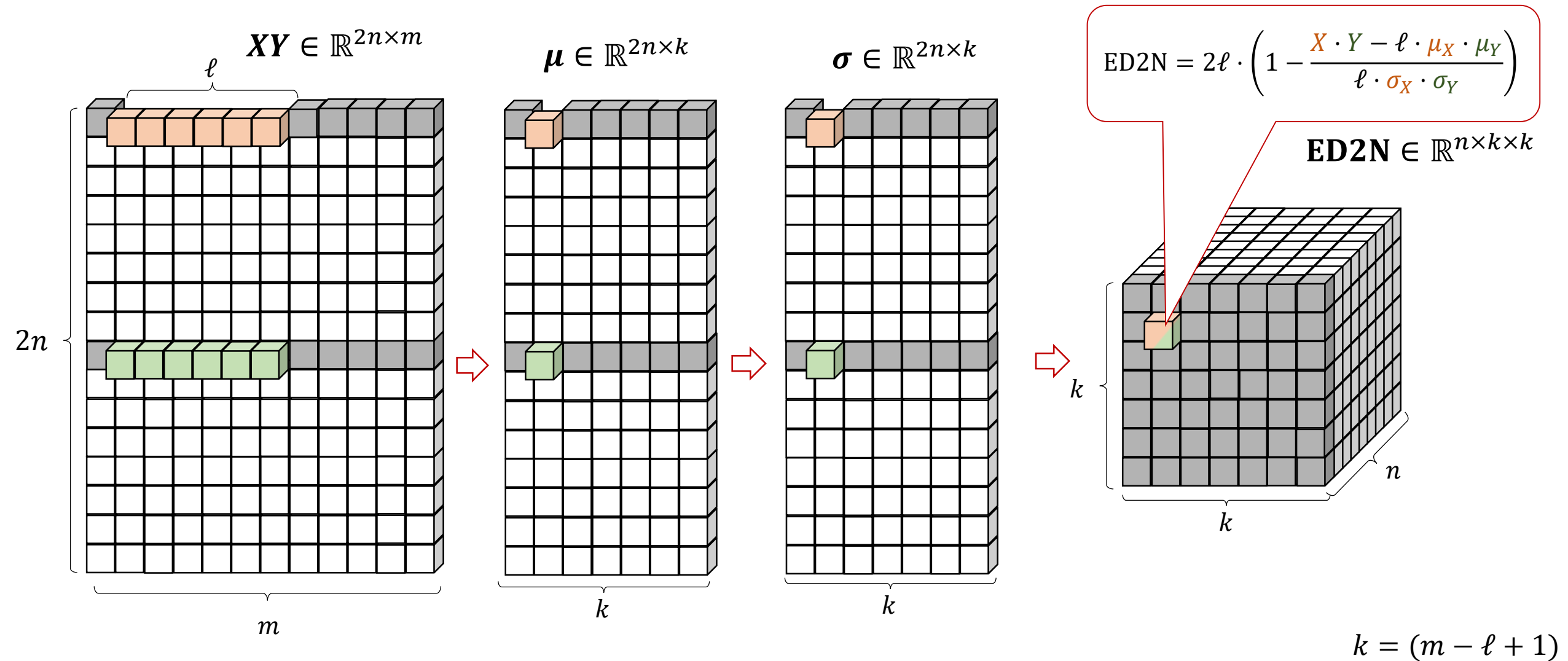
MPoss: параллельное вычисление



MP ℓ oss: параллельная нормализация

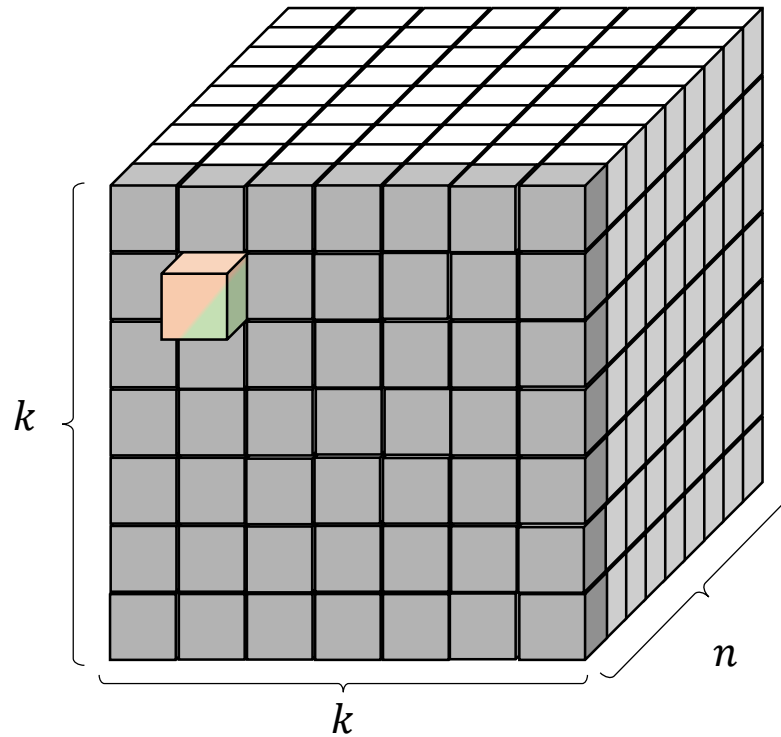


MP ℓ oss: вычисление ED2N



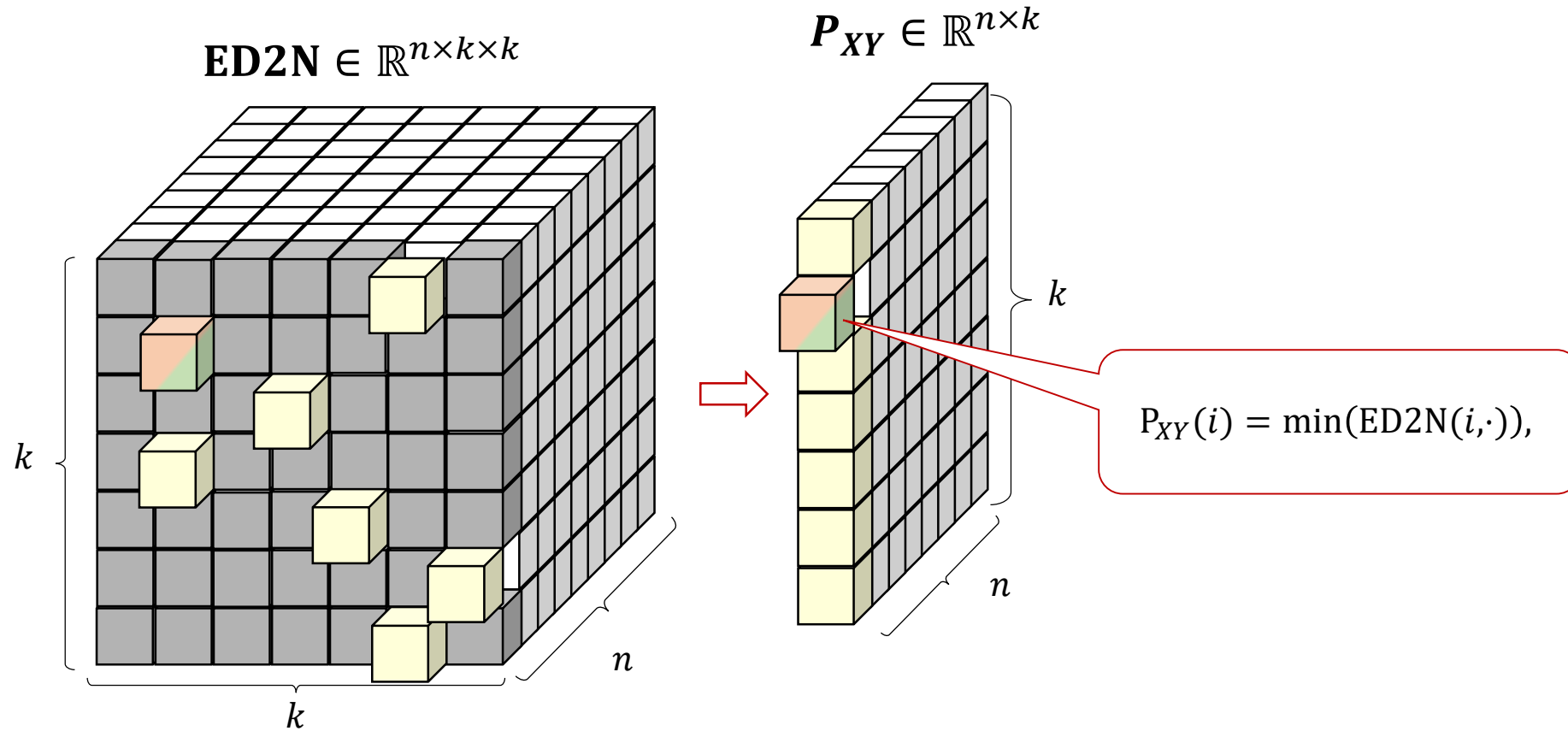
MPloss: вычисление матричных профилей

$$\mathbf{ED2N} \in \mathbb{R}^{n \times k \times k}$$



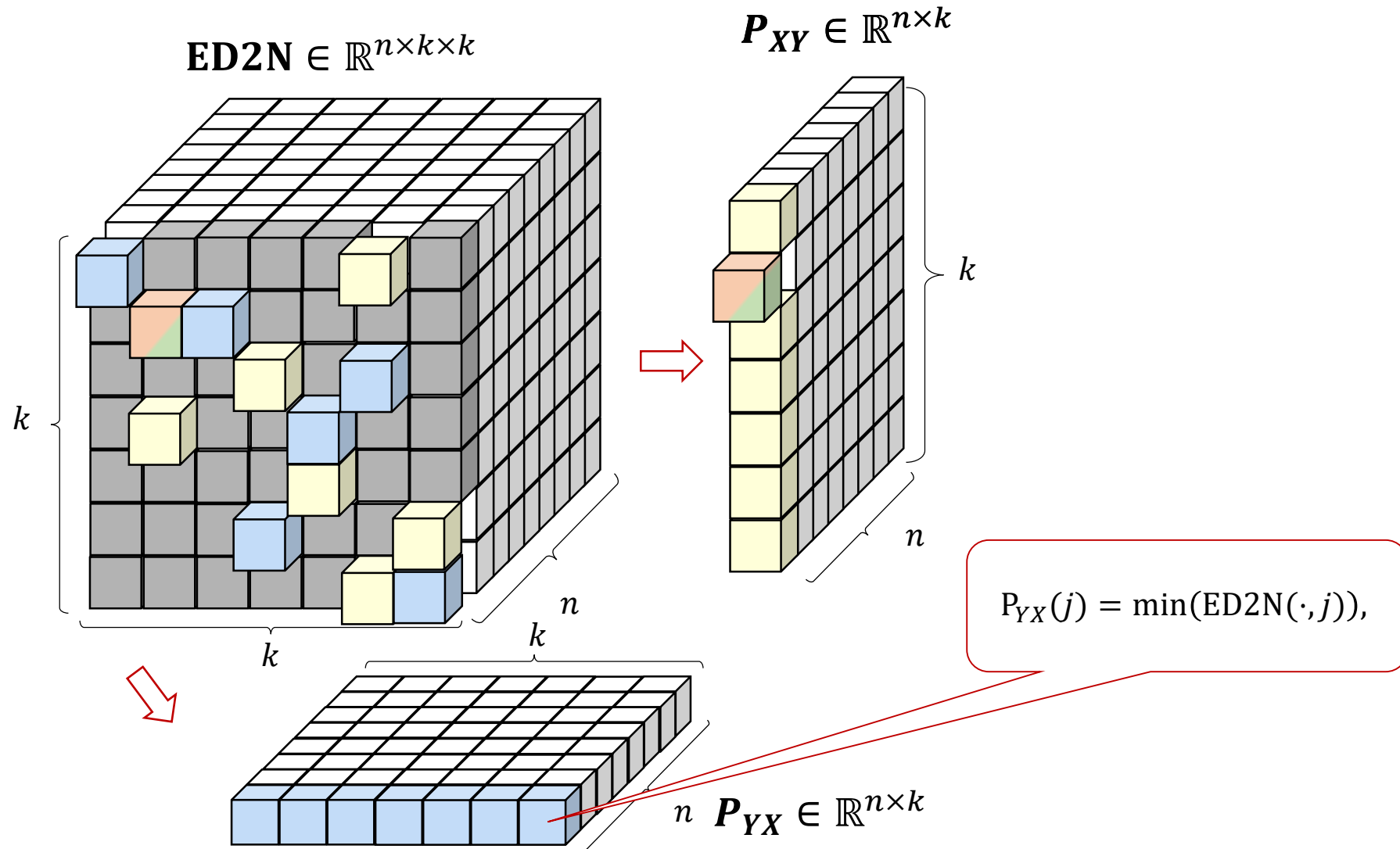
$$k = (m - \ell + 1)$$

MPloss вычисление матричных профилей



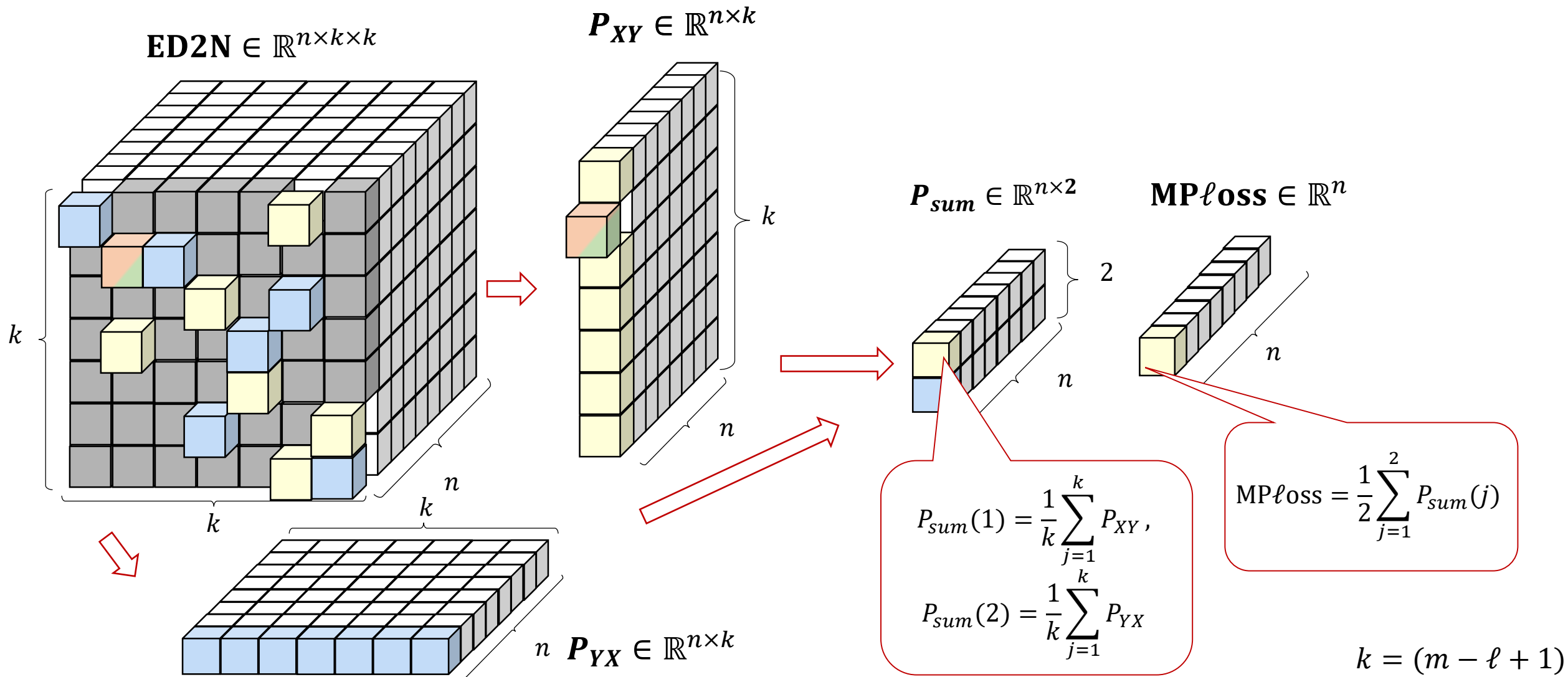
$$k = (m - \ell + 1)$$

MP ℓ oss: вычисление матричных профилей

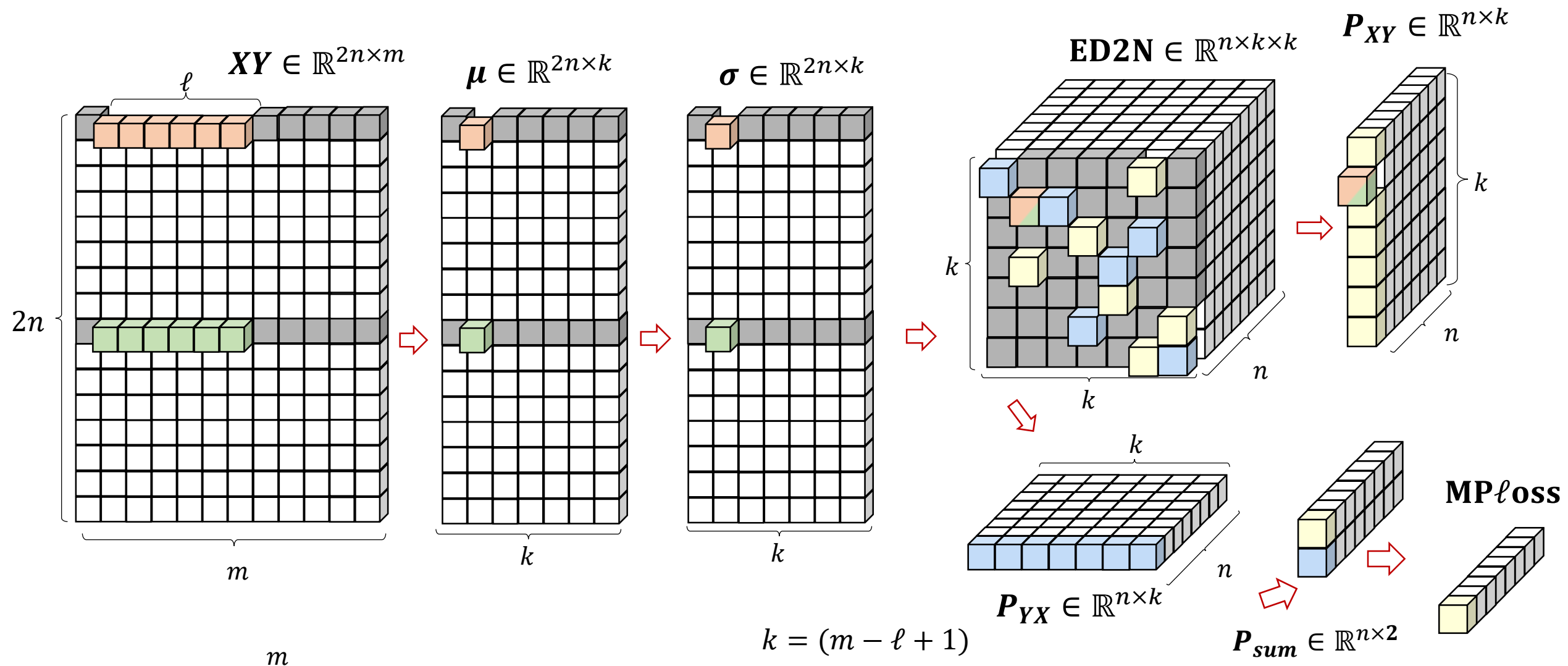


$$k = (m - \ell + 1)$$

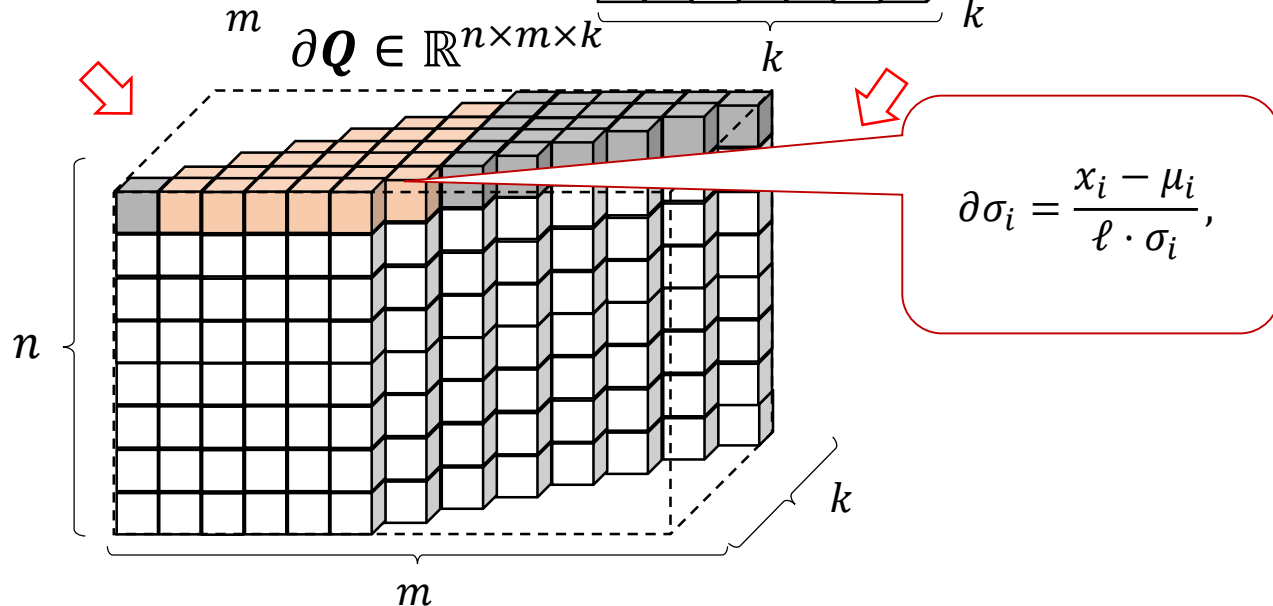
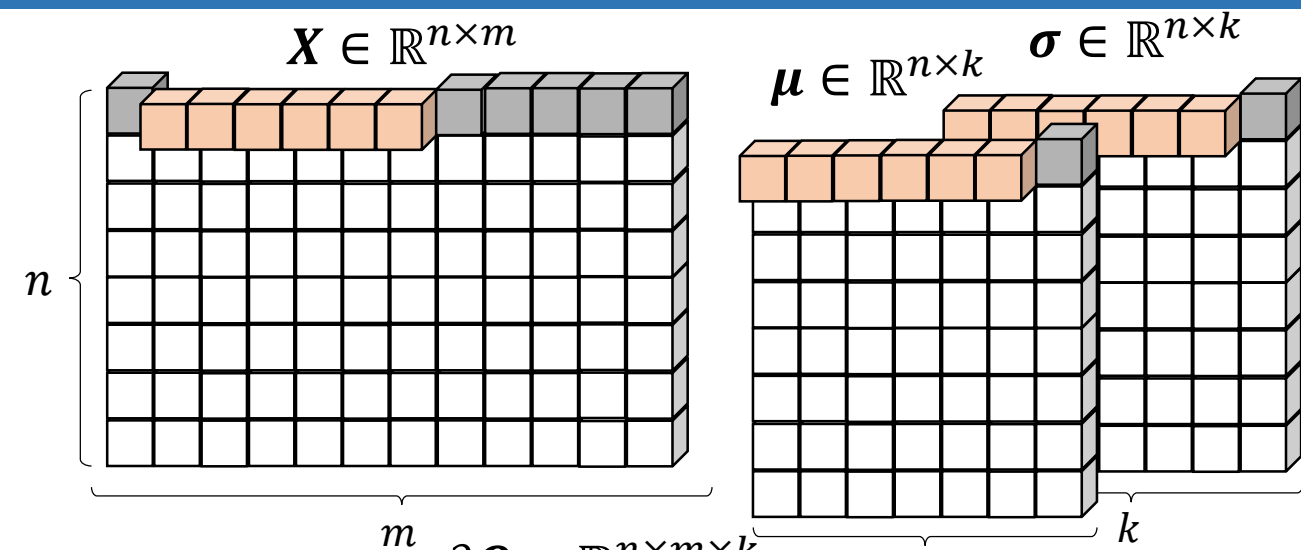
MPloss: вычисление ошибки



MPloss: параллельное вычисление ошибки

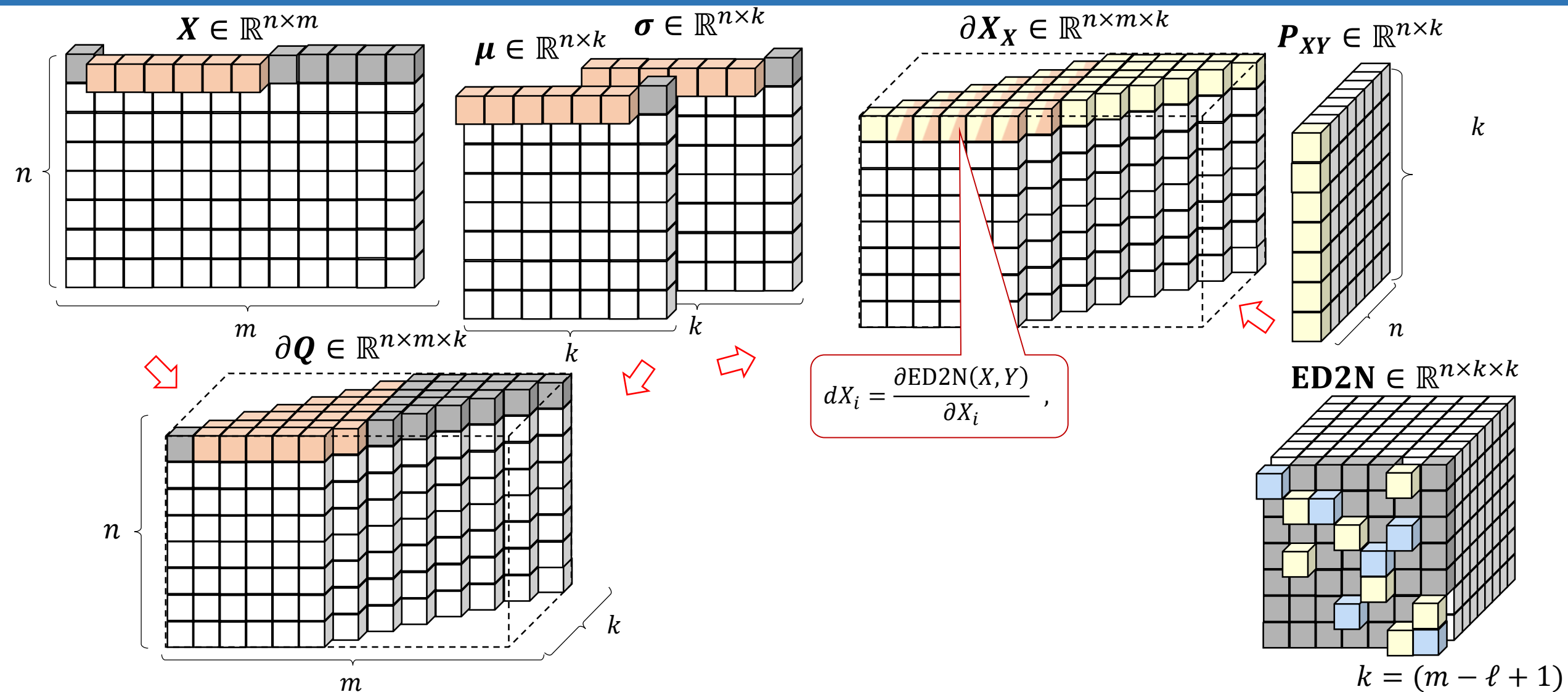


Описание параллельной реализации

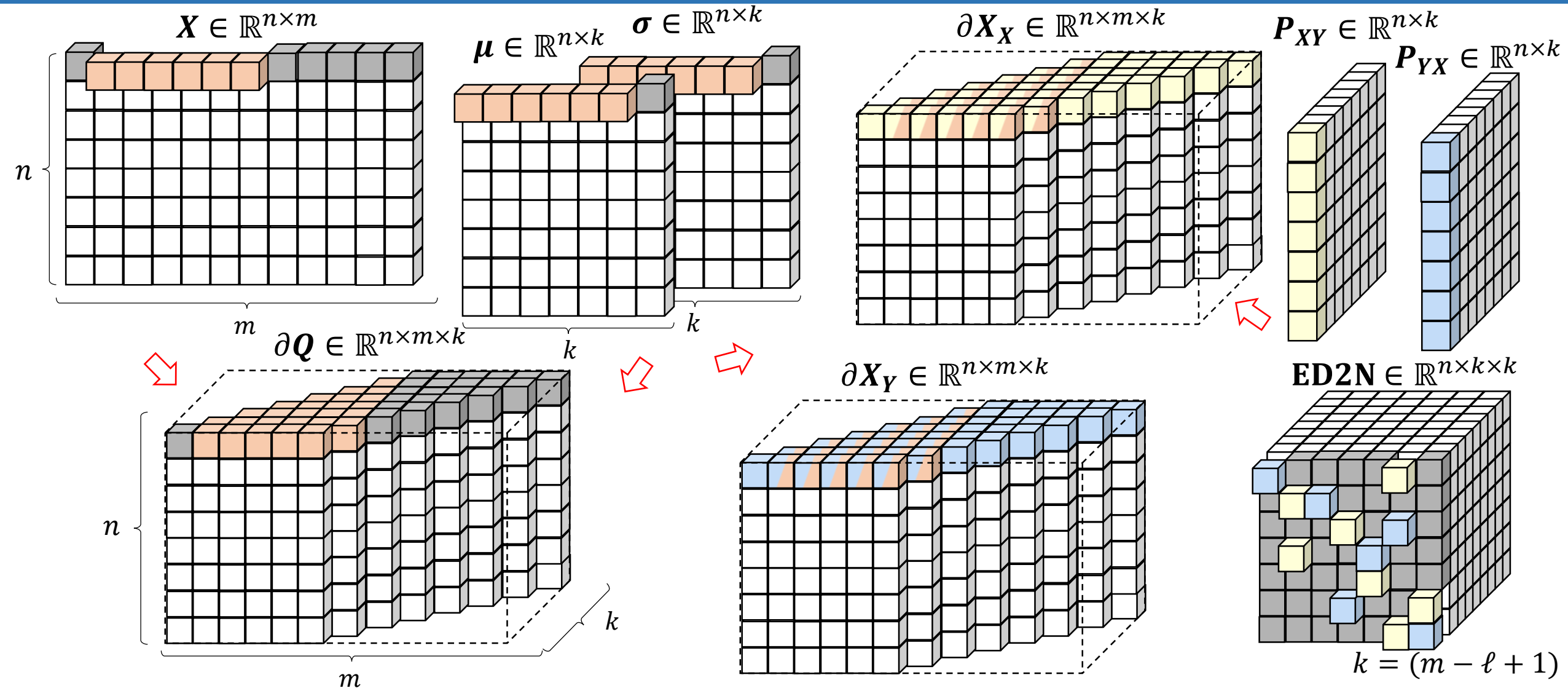


$$k = (m - \ell + 1)$$

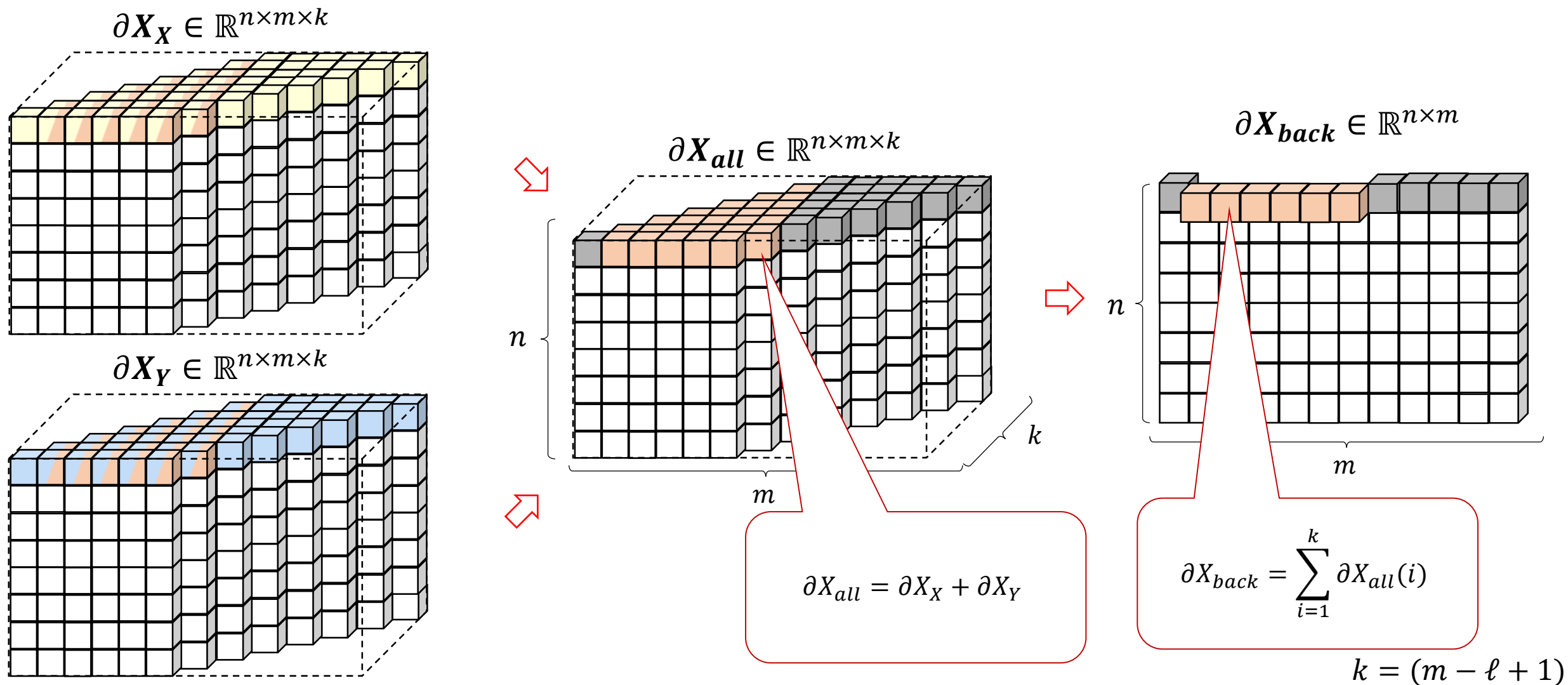
Описание параллельной реализации



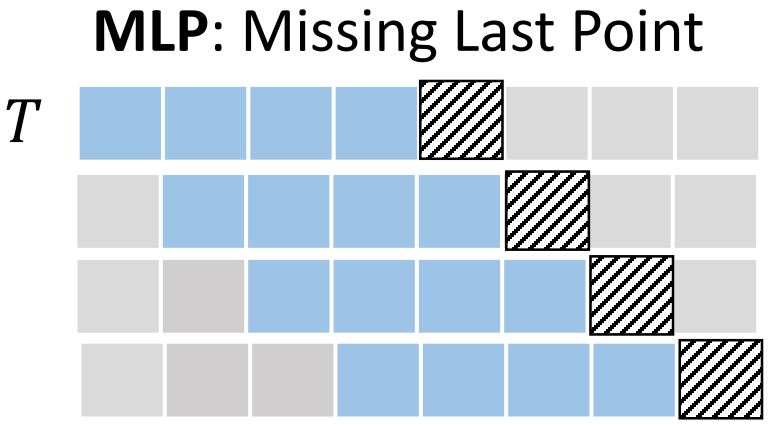
Описание параллельной реализации



MP ℓ oss: параллельное вычисление градиента ошибки



Вычислительные эксперименты: Наборы, модели и сценарий



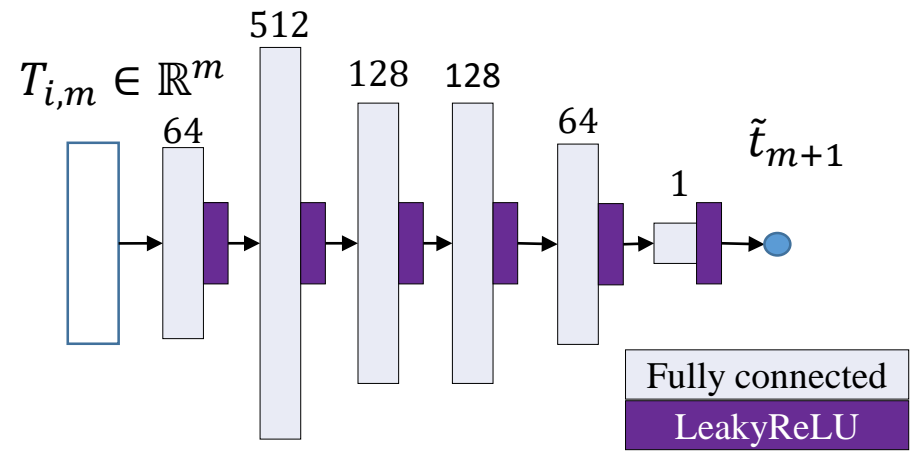
Наборы

Набор	Длина	Семантика
PAMAP	10 000	Показания носимого акселерометра во время физической активности человека
MADRID	25 000	Дорожное движение в Мадриде
BAFU	50 000	Сброс воды в реках Швейцарии
ElectricCity	5 000	Потребление электроэнергии домохозяйством

Параметры обучения

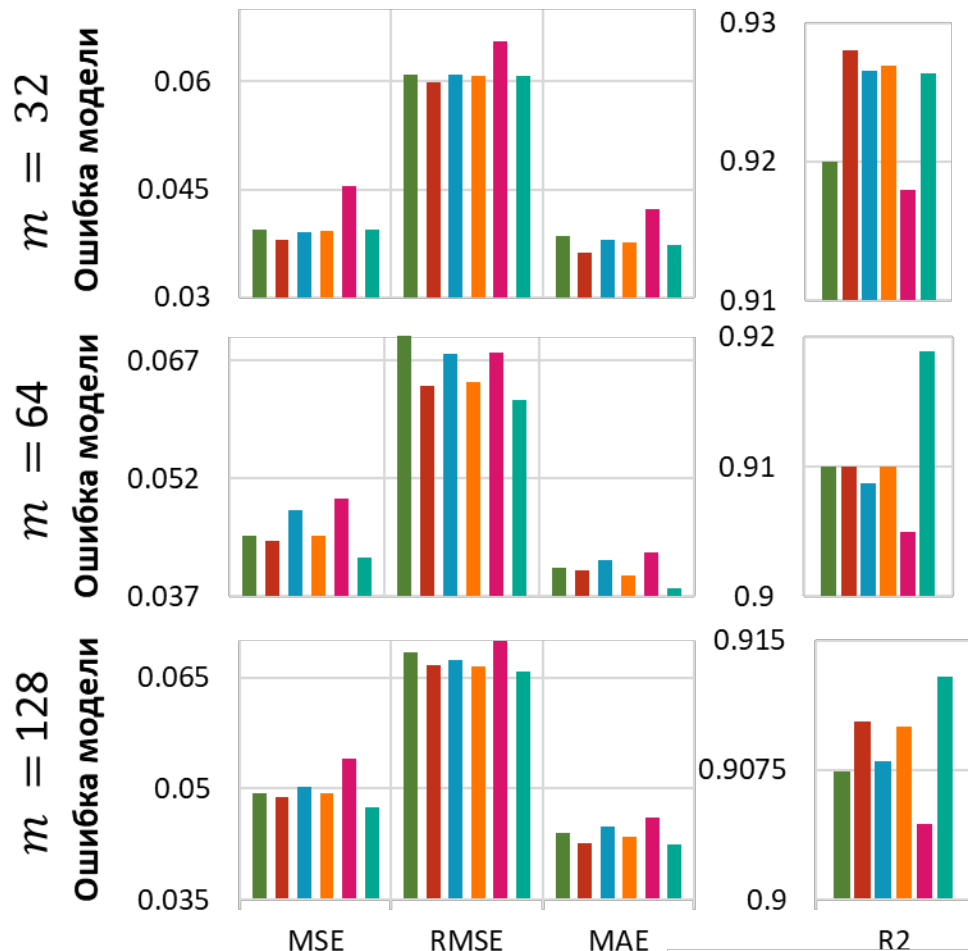
Параметр	Значения
Длина подпоследовательности, m	32, 64, 128
Скорость обучения	0.0001, 0.0005, 0.001, 0.0015, 0.002
Оптимизатор	Adam
Максимальное количество эпох	1000
Размер партии (batch size)	64

SFFN: Simple Feed-forward Network

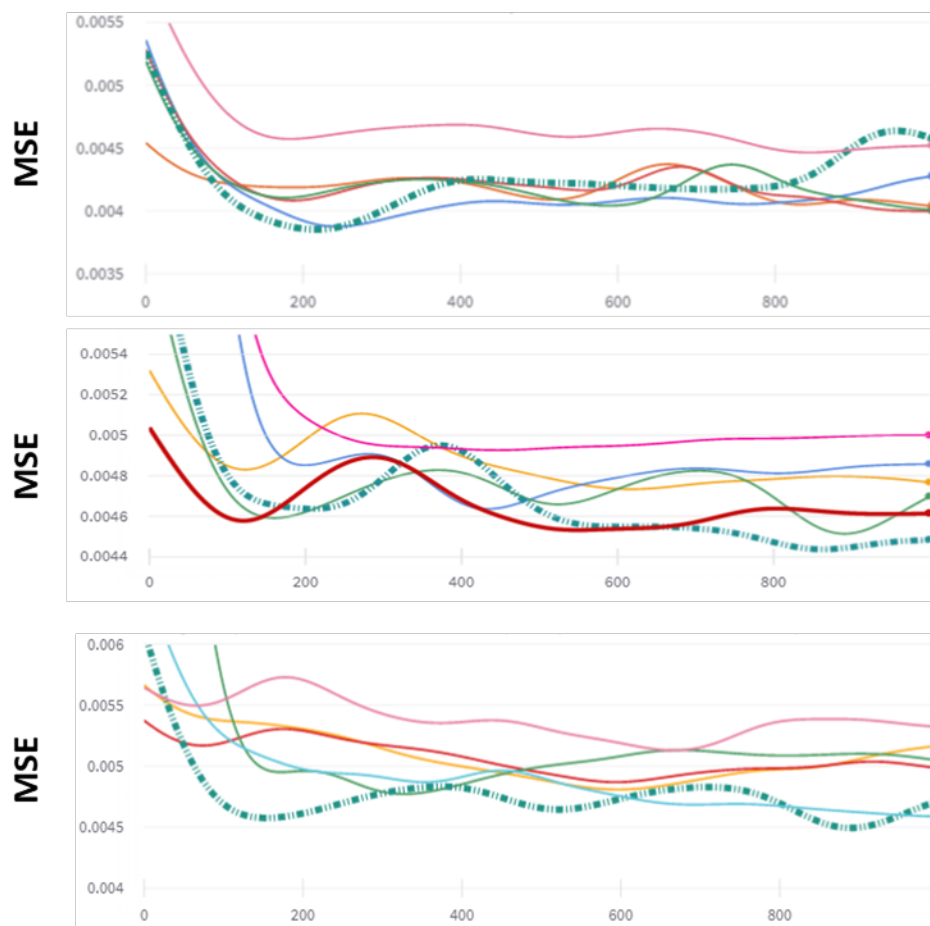


Вычислительные эксперименты: RAMAP

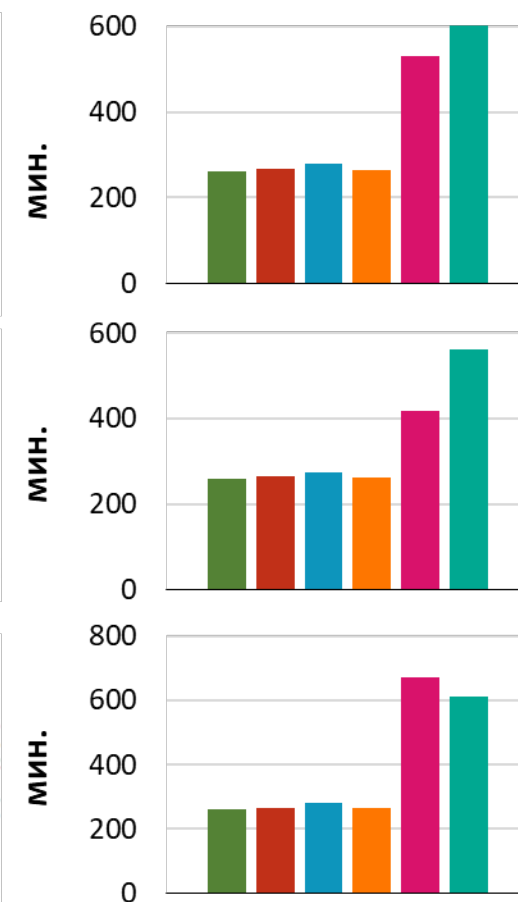
Ошибка на тестовой выборке после обучения



Ошибка на валидационной выборке во время обучения

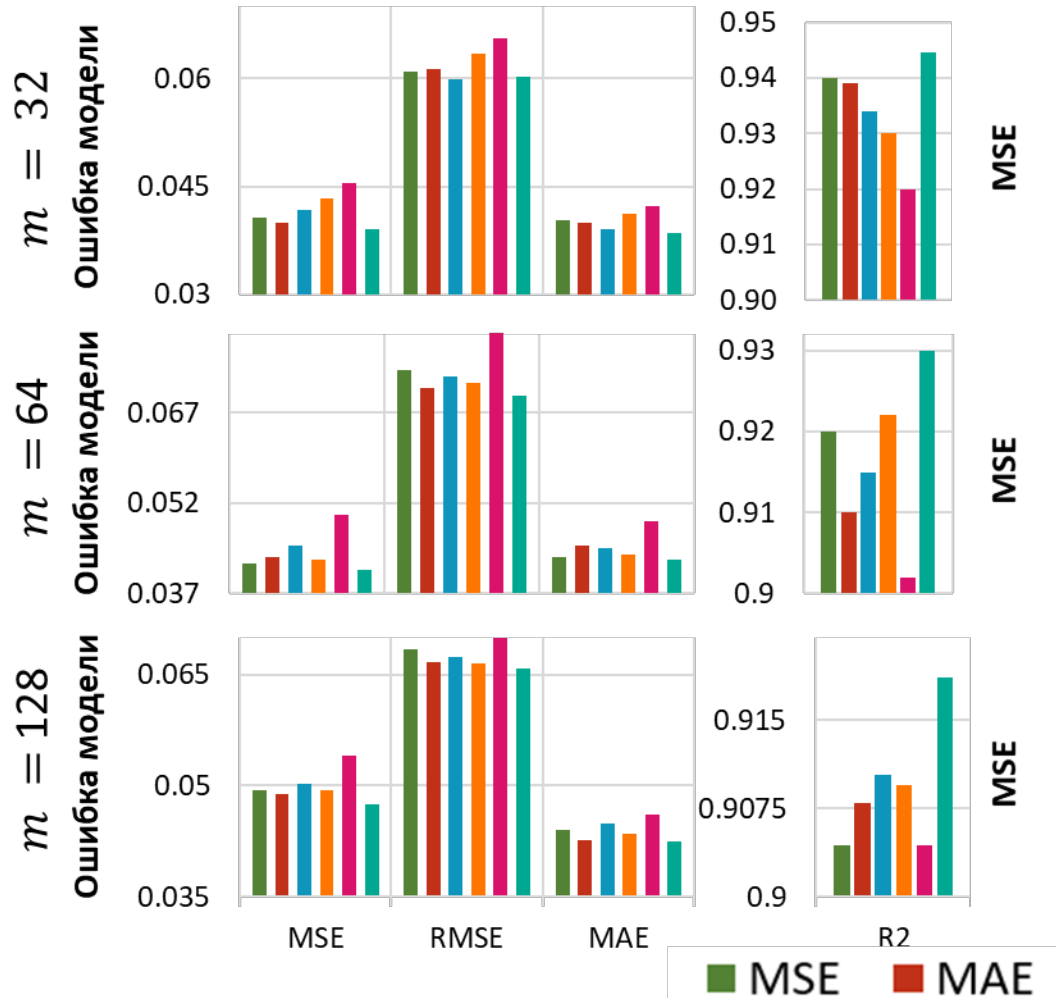


Время обучения

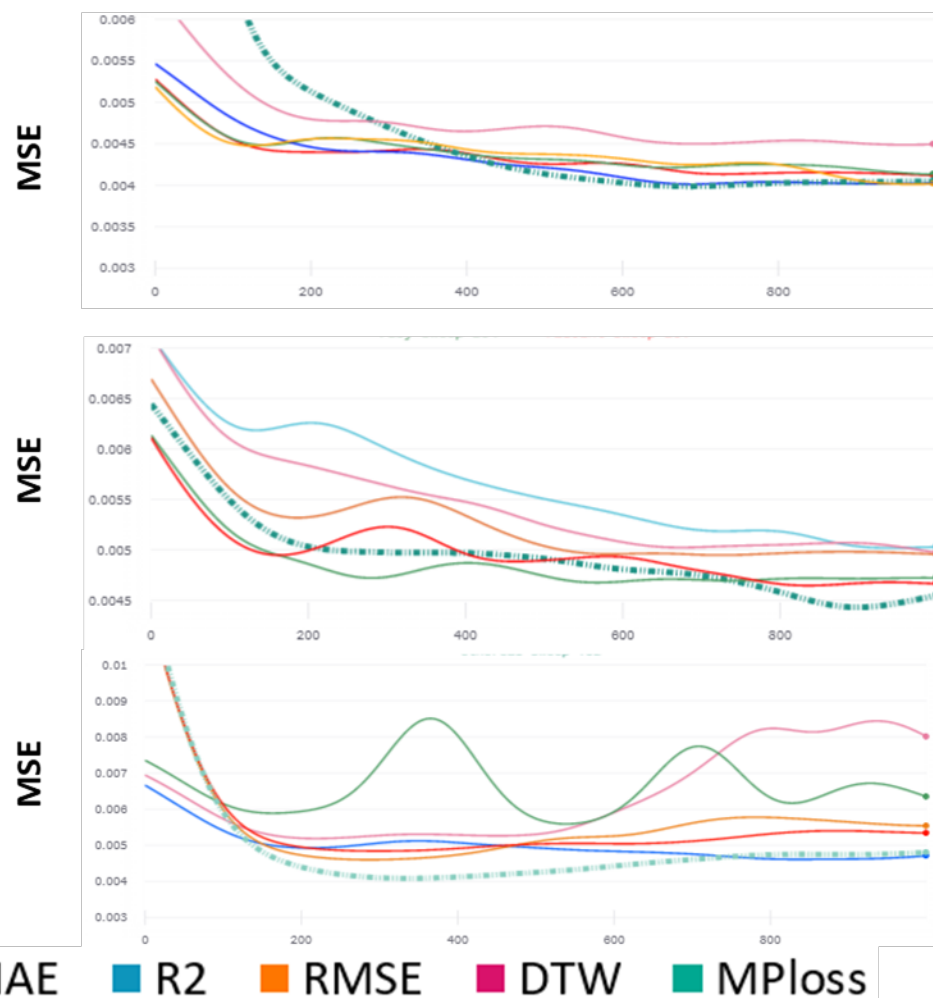


Вычислительные эксперименты: MADRID

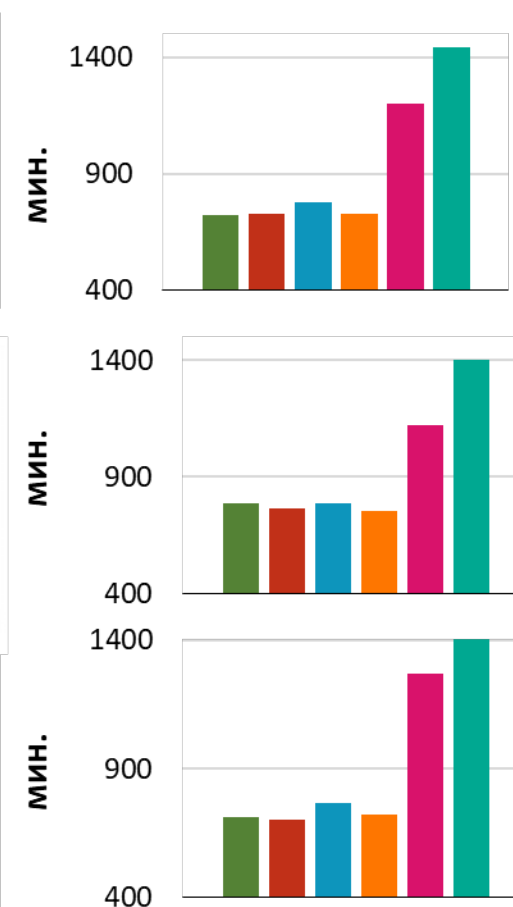
Ошибка на тестовой выборке после обучения



Ошибка на валидационной выборке во время обучения



Время обучения

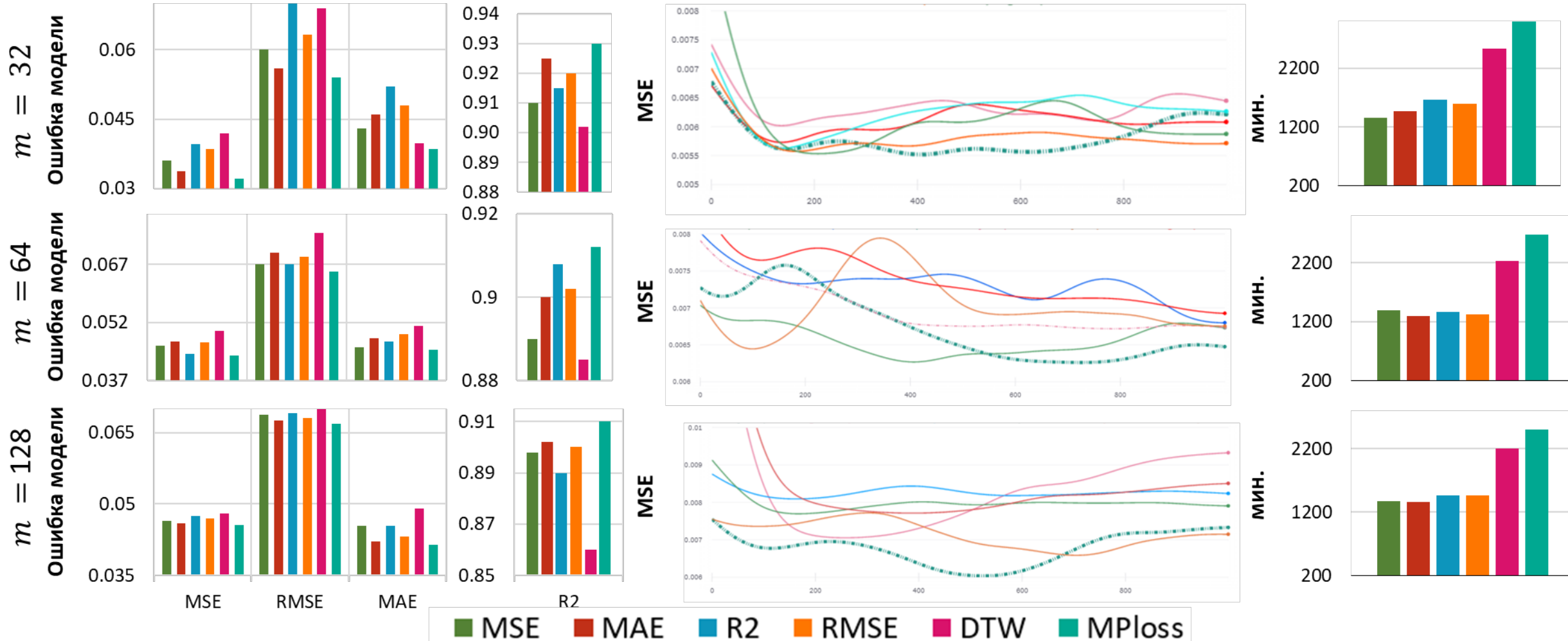


Вычислительные эксперименты: BAFU

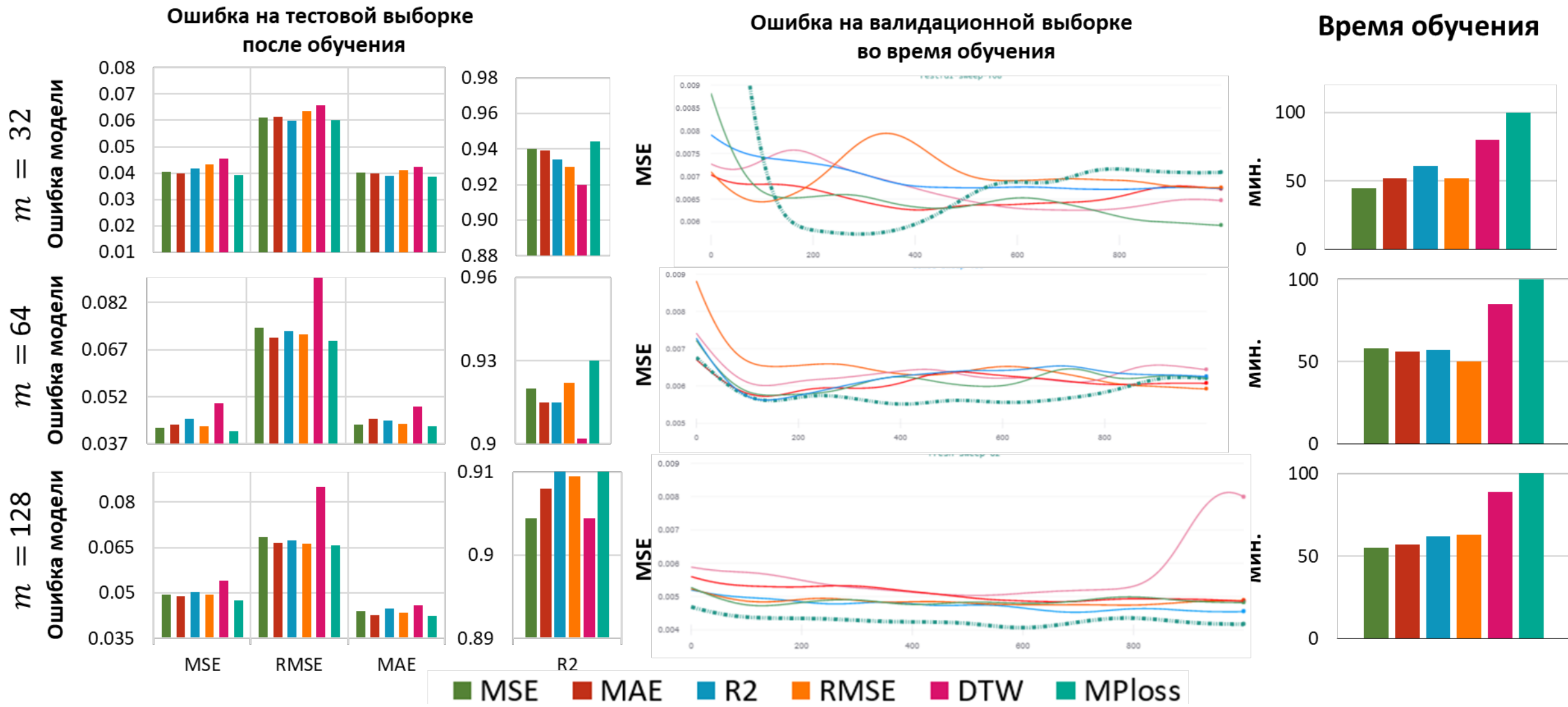
Ошибка на тестовой выборке после обучения

Ошибка на валидационной выборке во время обучения

Время обучения



Вычислительные эксперименты: ElectricCity



Заключение

- Предложены новая функция потерь $MP\ell_{oss}$ и метод ее параллельного вычисления, позволяющие увеличить точность нейросетевых моделей восстановления данных временных рядов при адекватном уменьшении производительности обучения моделей
- Будущие исследования: применение $MP\ell_{oss}$ в других архитектурах нейронных сетей
- Спасибо за внимание! Вопросы?
Алексей Юртин, yurtinalexei@yandex.ru