



ПОЛИТЕХ

Санкт-Петербургский
политехнический университет
Петра Великого

COMPUTO ERGO SUM

XXXV МЕЖДУНАРОДНАЯ НАУЧНАЯ КОНФЕРЕНЦИЯ ММТТ-35»

**УПРАВЛЕНИЕ ПРОЦЕССАМИ ПЛАНИРОВАНИЯ ЗАДАНИЙ
В ГЕТЕРОГЕННОМ СУПЕРКОМПЬЮТЕРНОМ ЦЕНТРЕ
С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ**

Работа выполнена в рамках
госзадания ФГАОУ ВО
СПБПУ FSEG-2022-0001

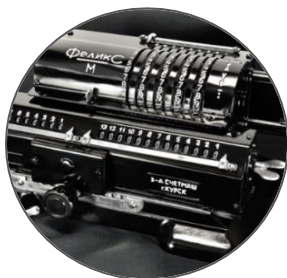
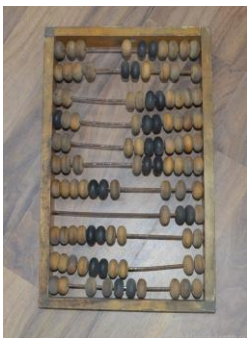
Минск,
25 октября 2022 г.

Заборовский В.С., Уткин Л. В.
Политехнический университет Петра Великого,
Санкт-Петербурга

«МИРОВАЯ ЛИНИЯ» ЭВОЛЮЦИИ: ОТ ПРОГРАММНЫХ АВТОМАТОВ К «УМНЫМ» ВЫЧИСЛИТЕЛЬНЫМ ПЛАТФОРМАМ

Физика достигла таких высот, что мы можем вычислить даже то, что невозможно себе представить
Л. Ландау

Эра
механических автоматов, исполняющих алгоритм, представленный в структуре самого «вычислителя»



эра
конечных автоматов, которые реализуют заранее составленные программы вычислений

Программа – мыслимая последовательность команд вычислений

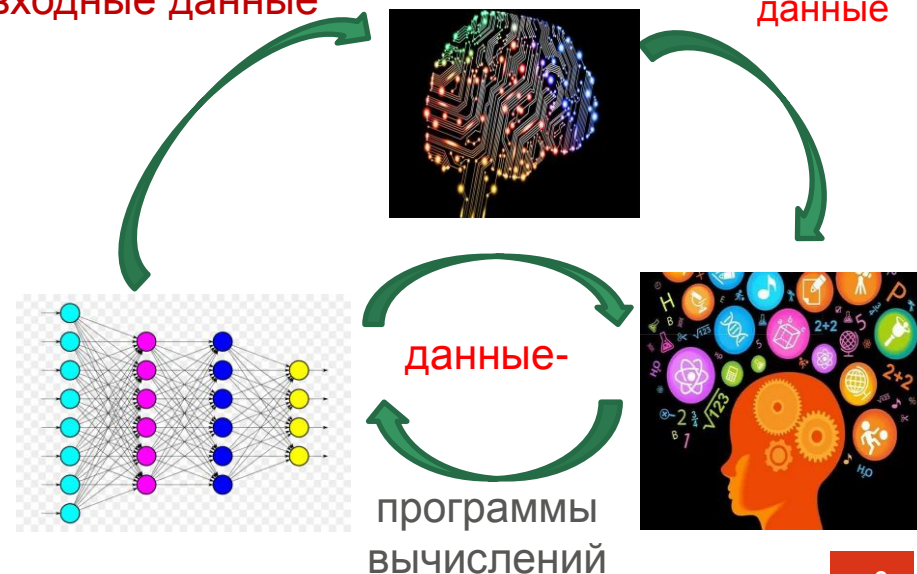


Суть «умного» подхода: «бесконечную ленту» из счетного множество команд «машины Тьюринга» заменить на «функцию распределения» множества входных X и целевых переменных уактуальных в момент времени t - оператор $P(X, y, t)$

эра
«умных» компьютерных платформ, которые реализуют алгоритмы вычислений, формируемые «в процессе» обработки входных и целевых данных

X-входные данные

y –целевые данные



АКТУАЛЬНЫЕ АСПЕКТЫ РАЗВИТИЯ СУПЕРКОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ: TOP 500

год	число ядер	$R_{\text{реак}}$, ПФлопс	R_{max} , ПФлопс	эл. мощность, МВт
2022 Frontier	8,700,XXX	1680.XX (1.7 ЭФлопс)	1100.XX (1.1 ЭФлопс)	21
2020 Fugaku	7,300,XXX	513.XX	415.XX	28
2010 Tianhe-1	186,XXX	4.7X	2.6X	4
2000 ASCI Intel	9,6XX	0.03	0.02	-

- 1 кг угля -> 3 кВтч = 0.003 МВтч
- 1 тонна угля -> 3 МВтч

21 МВт -> 21/3 =

- 7 тонн в час
- 168 тонн в день
- 60480 тонн в год

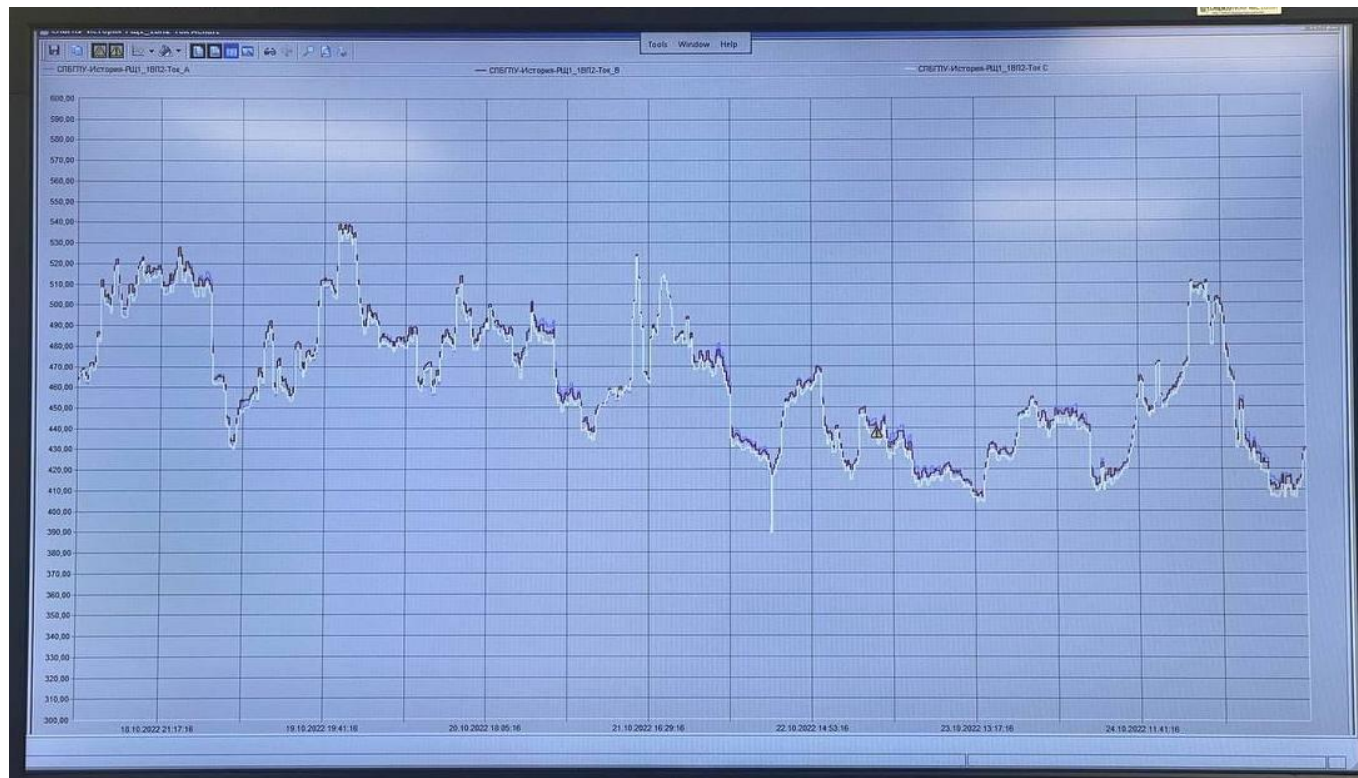
Вся Европа **потребляет**
400 000 МВт.

По линии электропередачи (500 кВ) можно передать
500 МВт

Проблемы:

- **УПРАВЛЕНИЕ ПРОЦЕССАМИ ПЛАНИРОВАНИЯ ЗАДАНИЙ** так, чтобы **снизить ПОТРЕБЛЯЕМУЮ МОЩНОСТЬ** (уменьшить «углеродный след цифровизации»)

ГРАФИК ПОТРЕБЛЕНИЯ ЭЛ. МОЩНОСТИ ПРИ ВЫПОЛНЕНИИ ЗАДАНИЙ



Ошибки ПЛАНИРОВАНИЯ ЗАДАНИЙ

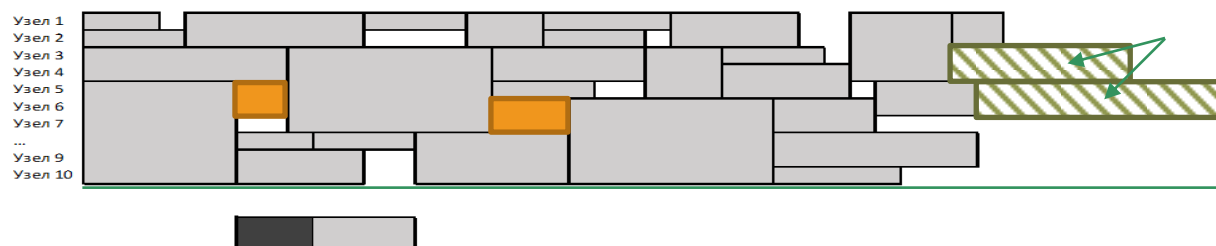
ПРИВОДЯТ К НЕРАВНОМЕРНОСТИ
ЗАГРУЗКИ ВЫЧИСЛИТЕЛЬНЫХ
РЕСУРСОВ

Новая стратегии развития КТ:

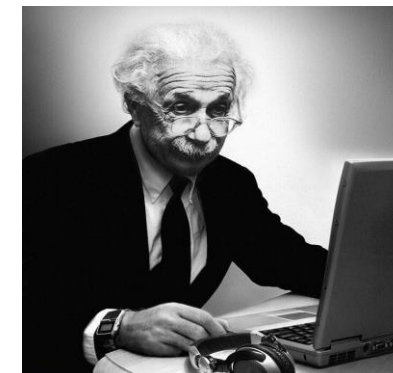
Less Moore, More Brain
Меньше Мура, Больше Мозга

«More Brain» – это
как и ...куда?

«визуализация»
очереди исполняемых
заданий в узлах
кластера :



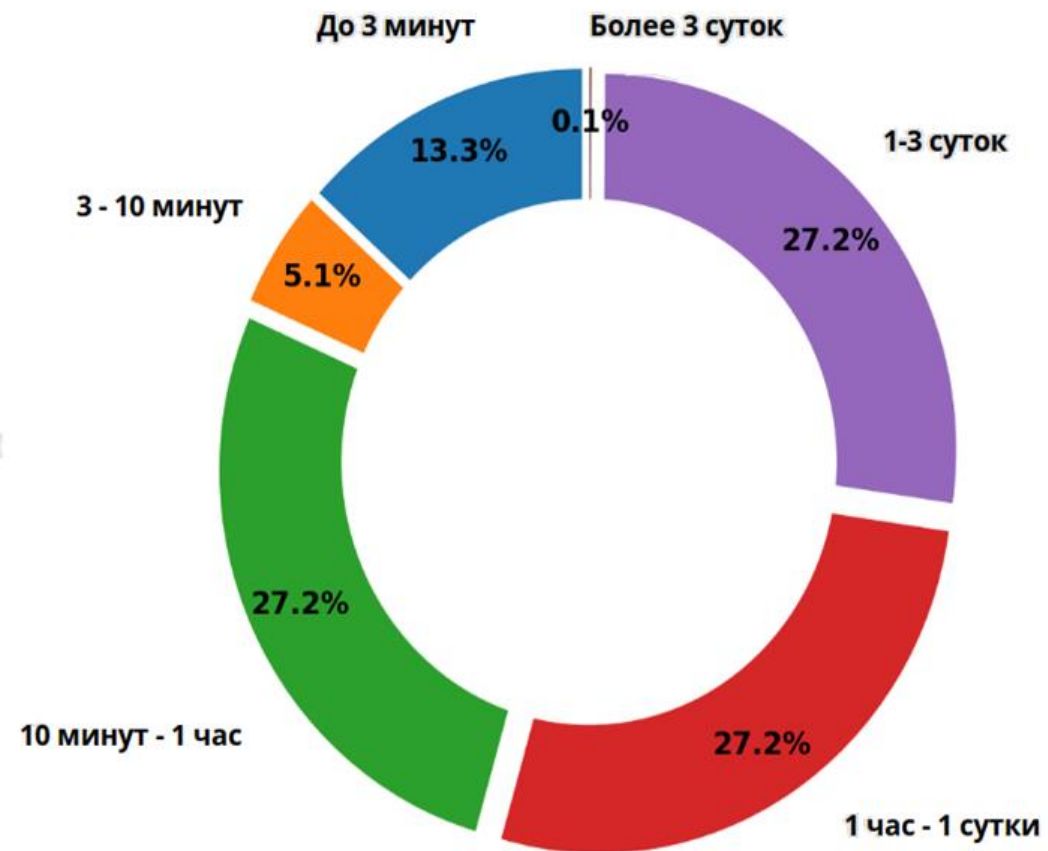
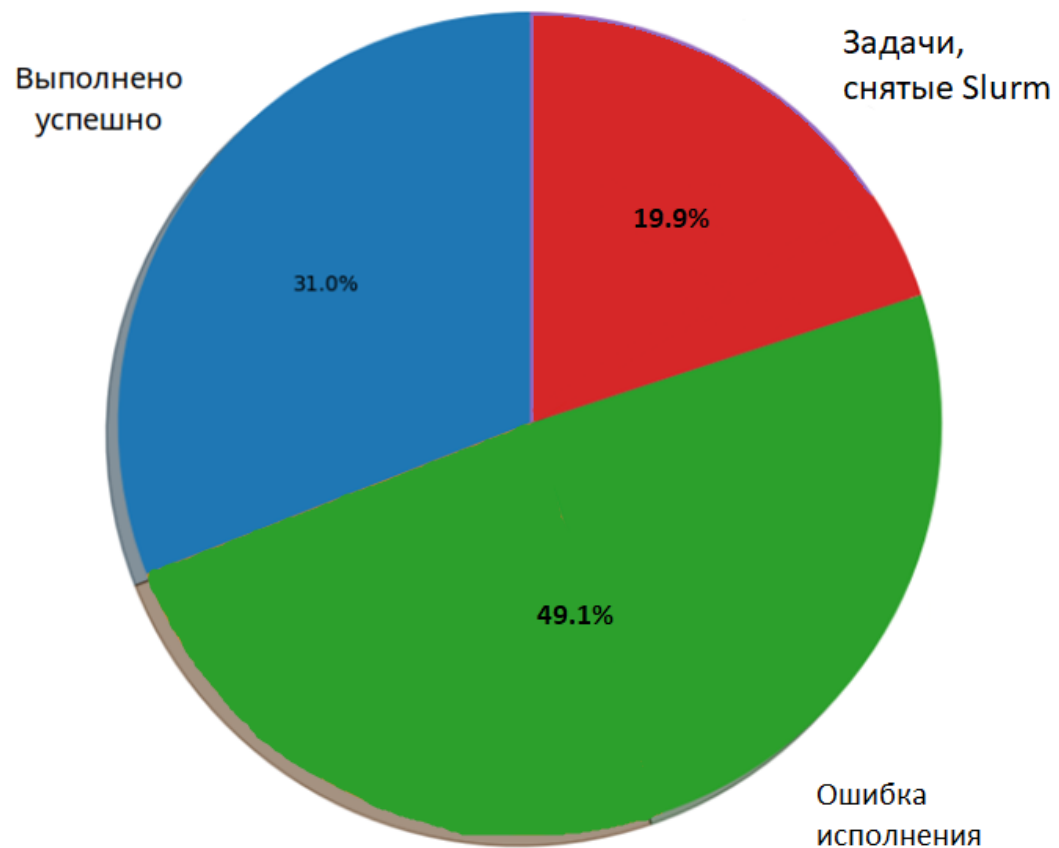
- Масштаб и сложность вычислений продолжают увеличиваться, а планирование процессов выполнения заданий **становится все более сложным процессом**, поэтому :
 - особую актуальность приобретают подходы к созданию систем управления, обеспечивающих **точное прогнозирование времени выполнения заданий** с целью формирования равномерной загрузки имеющихся вычислительных ресурсов: &...снижения потребляемой электрической мощности.
- Исследование «лучших практик» показали, пользователь — неотъемлемая часть СК системы в целом, а в **функции планировщика заданий** должны входить не только анализ данных
 - о состоянии ресурсов **непосредственно самого СК** (загрузка узлов, наличие свободных ресурсов памяти и др.), **но и**
 - предоставляемых **пользователем** при формировании заданий



«умный»
планировщик
заданий



МОТИВАЦИЯ ПЕРЕХОДА К «УМНЫМ» РЕШЕНИЯМ. СТАТИСТИКА «ВЫЖИВАНИЯ» И ПРОДОЛЖИТЕЛЬНОСТИ

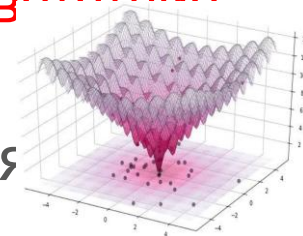


«Успешно выполнено» - составляет **меньше 1/3** от общего числа обработанных заявок пользователей

«MORE BRAIN» - ЭТО КАК — ИСПОЛЬЗОВАНИЕ НОВЫХ «МЕТРИК»

Так как запуск задачи пользователя на исполнение не гарантирует ее успешное завершение, нужны метрики «умных» решений:

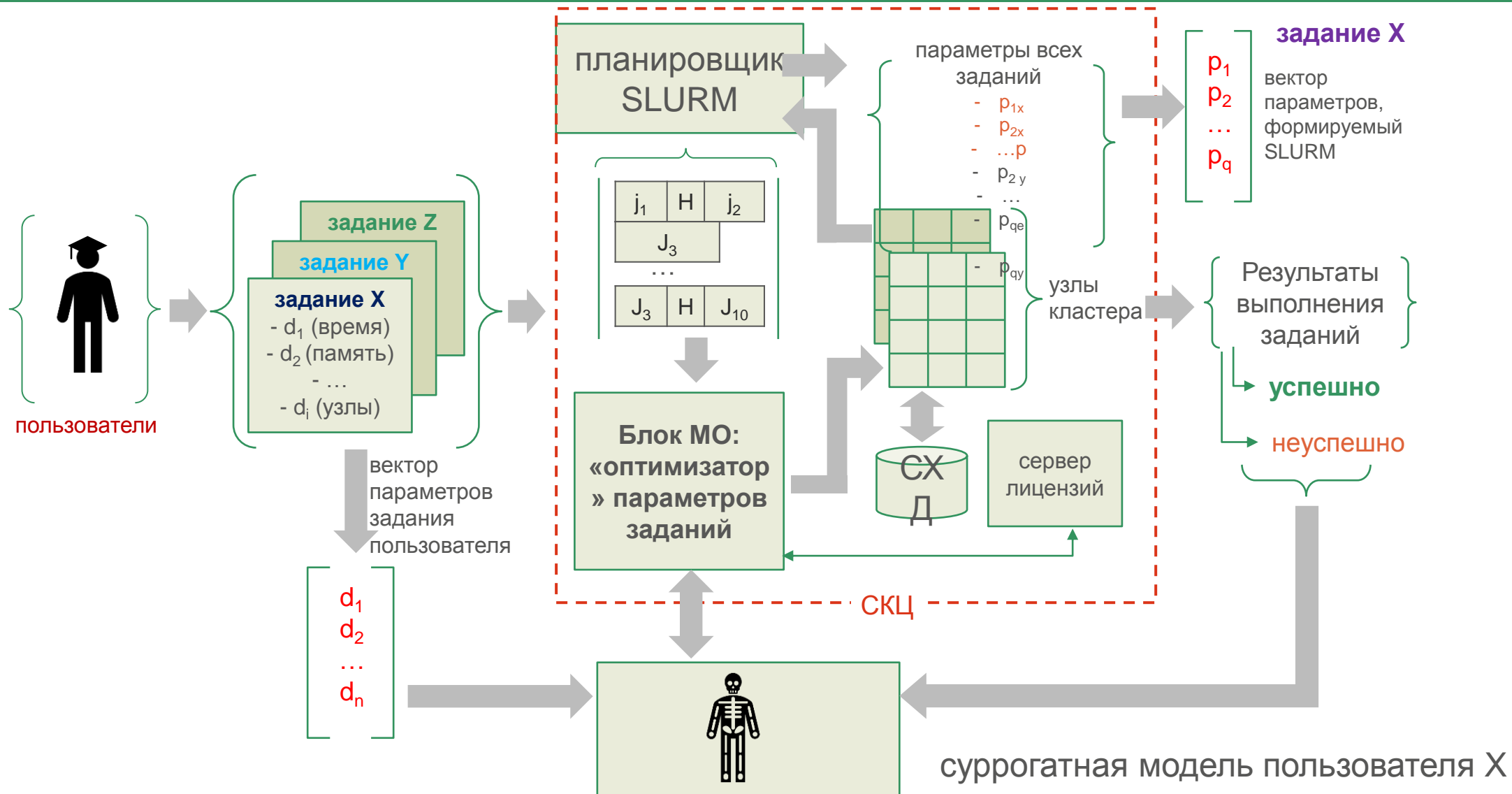
- **«тонкая» настройка** заданий пользователей с использованием механизма внимания, когда каждому параметру задания ставится в соответствие «эмбединг» вектор внимания
- **«объяснимость полученных результатов»** (успешное завершение, снятие диспетчером, ошибка в процессе выполнения)

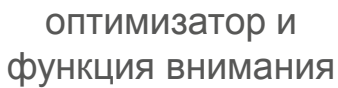


Необходимо использовать также **метрики производительности** СК, которые позволяют определять какие:

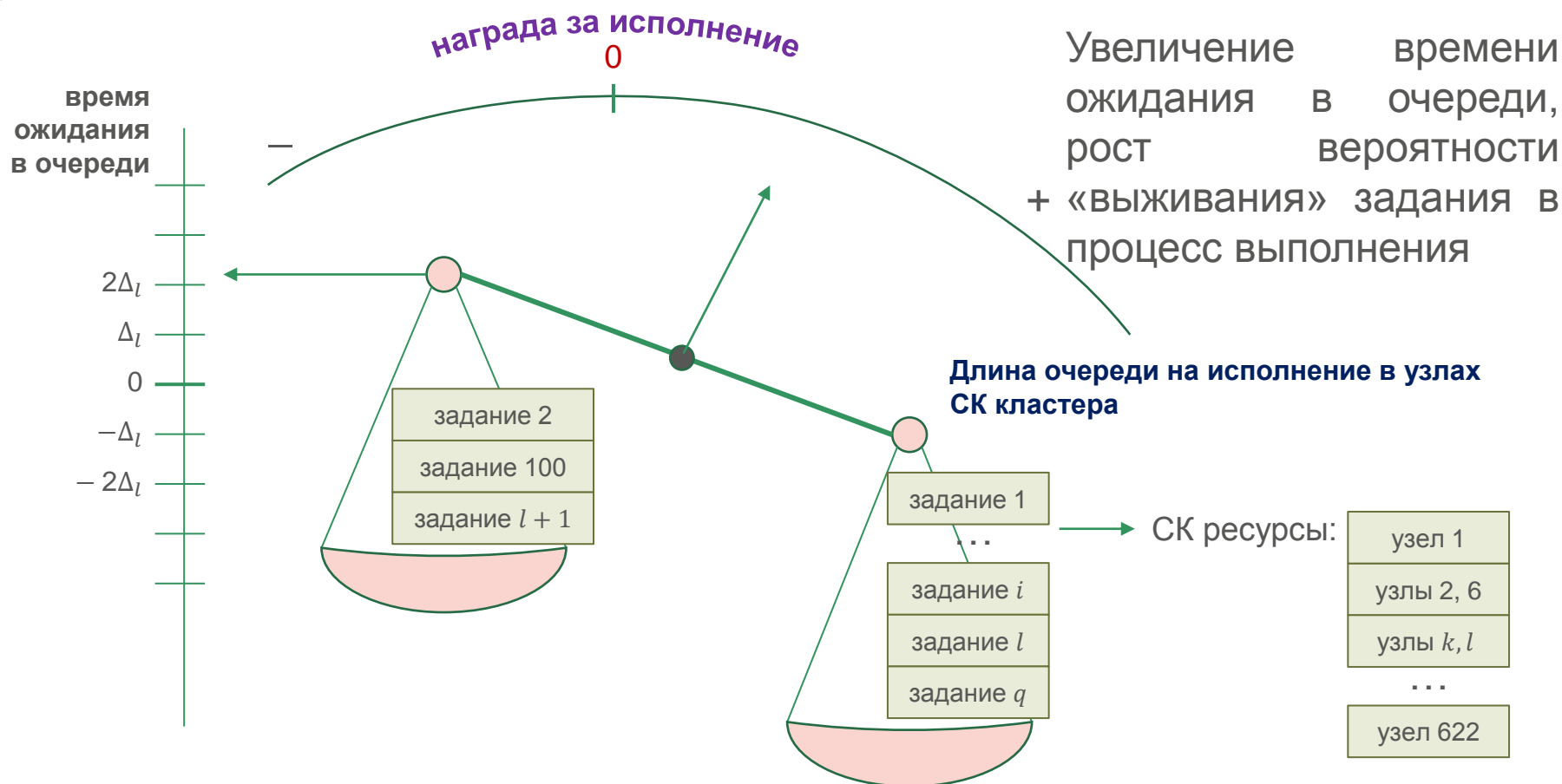
- параметры заданий **следует принять**, чтобы **повысить вероятность «выживания»** задания в процессе его исполнения
- стратегии исполнения заданий **могут обеспечить равномерную загрузку** имеющихся аппаратных ресурсов СК

ВАРИАНТ РЕШЕНИЯ “MORE BRAIN” С ИСПОЛЬЗОВАНИЕМ БЛОКА «ОПТИМАЛЬНОГО» ПРЕДСКАЗАНИЯ ПАРАМЕТРОВ ЗАДАНИЙ ПОЛЬЗОВАТЕЛЕЙ





«БАЛАНС» АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ : КРИТЕРИИ НА ОСНОВЕ ФУНКЦИИ ВНИМАНИЯ



Алгоритмы балансирования очереди могут меняться в зависимости критериев

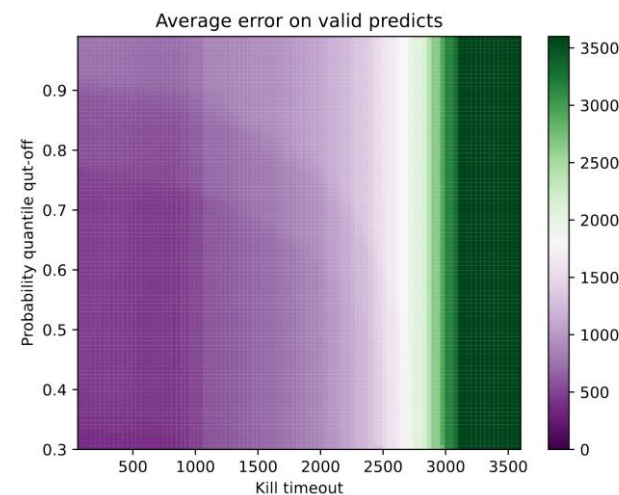
- 1) min среднего времени «простоя» (ожидания)
- 2) min «время выполнения» задания
- 3) min время ожидания + время выполнения

И

эффективность работы СК платформы



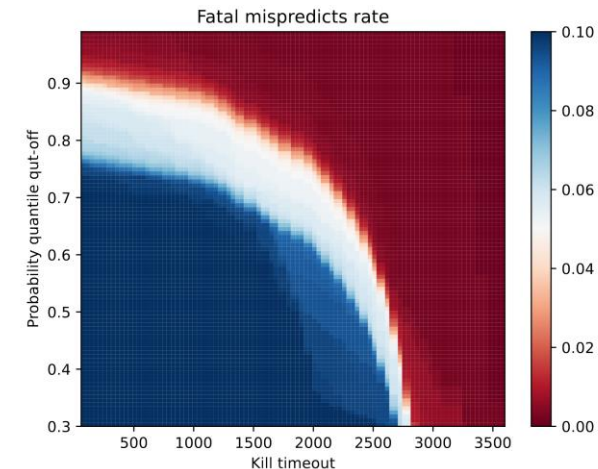
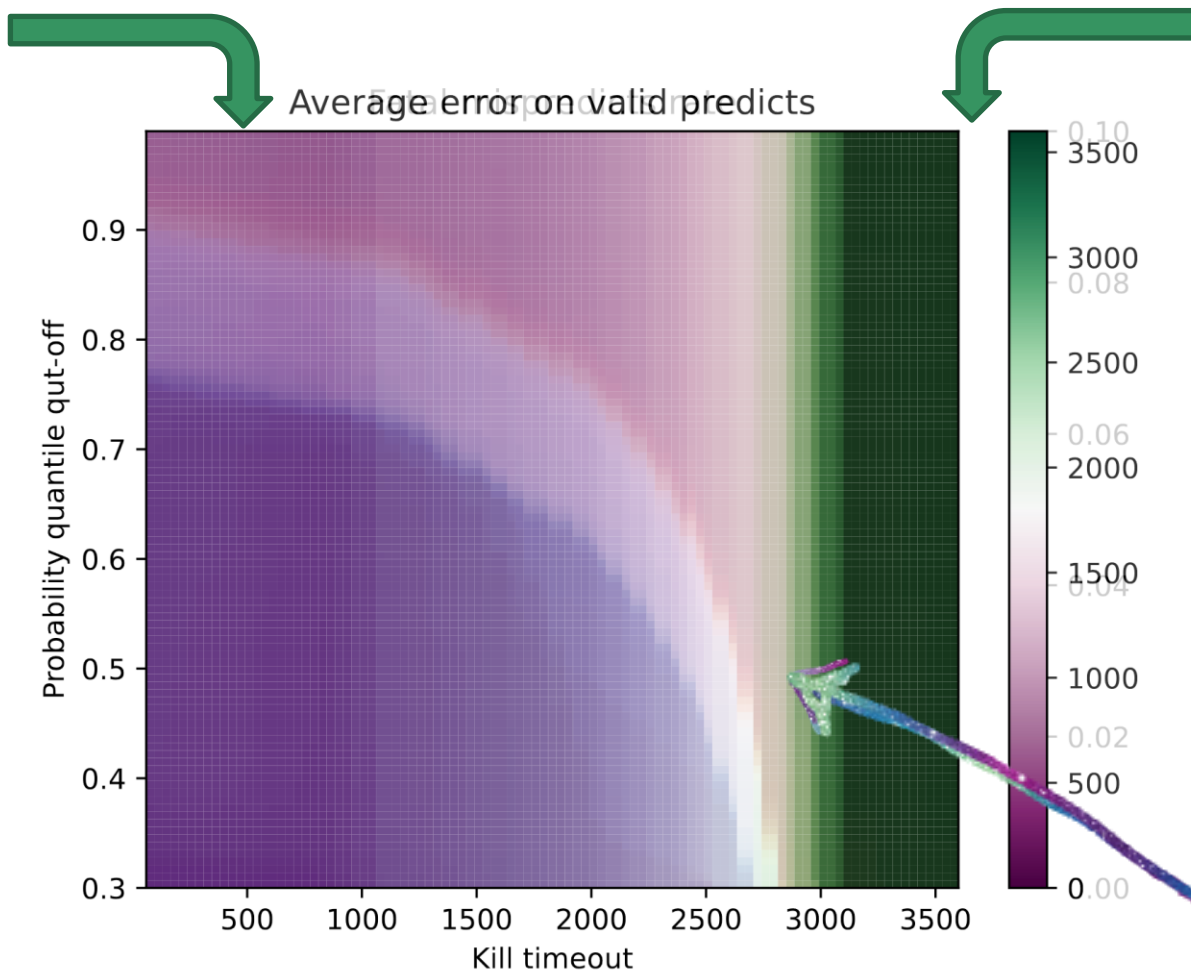
«УСТОЙЧИВОЕ НЕРАВНОВЕСИЕ» ОШИБОК ПРЕДСКАЗАНИЙ



Комментарий:

Слева – средняя ошибка
«**правильных**»
предсказаний (время
равно или больше
фактического)

Справа – доля
«**неправильных**»
предсказаний (с
занижением времени
выполнения)



Комментарий:

В качестве критерия
«выживаемости» заданий
нужна функция «засорения»
множества «правильных»
множеством «неправильных»
ошибок

Такая область на среднем
рисунке имеет «розовый
цвет» (выделил красным
контуром)

Требуется динамическая корректировка параметров приоритета заданий

1. Существующие методы машинного обучения СК **не стоит переоценивать**, однако их «корректное» использование **позволяет реально повысить эффективность** работы СК платформы, если в систему управления исполнения заданий включить «обучаемую» суррогатную модель пользователя.
2. Задача **«обучения» СК точному планированию времени** поддается эффективному решению, но требует встраивания в контур исполнения заданий «умного» блока (с точки зрения выбранной политики выполнения заданий) **оценки корректности данных**, представленных пользователем.
3. Так как пользователь непрерывно взаимодействует СК платформой и постоянно совершенствует свой собственный опыт, основываясь на результатах вычислений с учетом значений выбранных параметров заданий, то суррогатная модель должна носить объяснительный характер и учитывать **фактор «внимания» (мульти-внимания)** по отношению к вектору параметров исполняемого задания
4. Эффективная работы «умного» блока **требует постоянного уточнения «суррогатной модели пользователей»** СК платформы, поэтому для использования «умного» блока оценки времени исполнения для других СК платформ необходимо проводить процесс «дообучения» моделей с учетом конкретного разнообразия классов решаемых прикладных задач