

ПРИМЕНЕНИЕ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ ДЛЯ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ РАСПРОСТРАНЕНИЯ СОЦИАЛЬНО ЗНАЧИМЫХ ЗАБОЛЕВАНИЙ

Леоненко В. Н., Университет ИТМО, Санкт-Петербург
**Перцев Н. В., Омский филиал Института математики
им. С. Л. Соболева СО РАН, Омск**

ТЕМЫ ИССЛЕДОВАНИЙ

- Социально значимые заболевания – заболевания, отличающиеся существенным отрицательным влиянием на организм человека, на человеческий капитал в целом, на социально-экономическое развитие.
- Цель исследований: изучение динамики социально значимых заболеваний и оценка эффективности программ обследования индивидов методами имитационного моделирования.



МОДЕЛИРОВАНИЕ ДИНАМИКИ ТУБЕРКУЛЁЗА ОРГАНОВ ДЫХАНИЯ В РЕГИОНАХ РФ

- Удельный вес ТОД в структуре смертности от инфекционных заболеваний в РФ составляет 70-85%.
- Уровень заболеваемости ТОД существенно отличается в различных регионах РФ.
- Задача: построить имитационную модель распространения ТОД для поиска причин различий в уровнях заболеваемости ТОД в регионах РФ

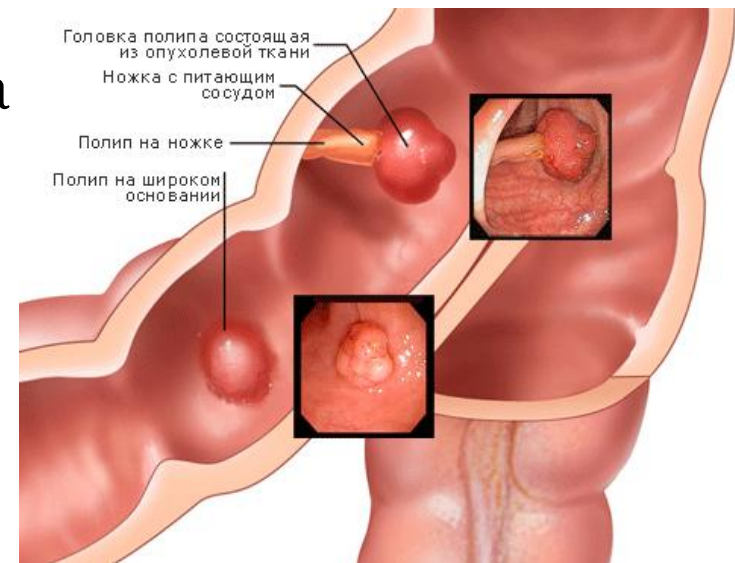


СРАВНЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДОВ ВЫЯВЛЕНИЯ ИНДИВИДОВ, ПРЕДРАСПОЛОЖЕННЫХ К КОЛОРЕКТАЛЬНОМУ РАКУ

- Колоректальный рак (КРР) вызывается перерождением полипов на внутренних стенках толстой кишки.

- Опасность КРР:

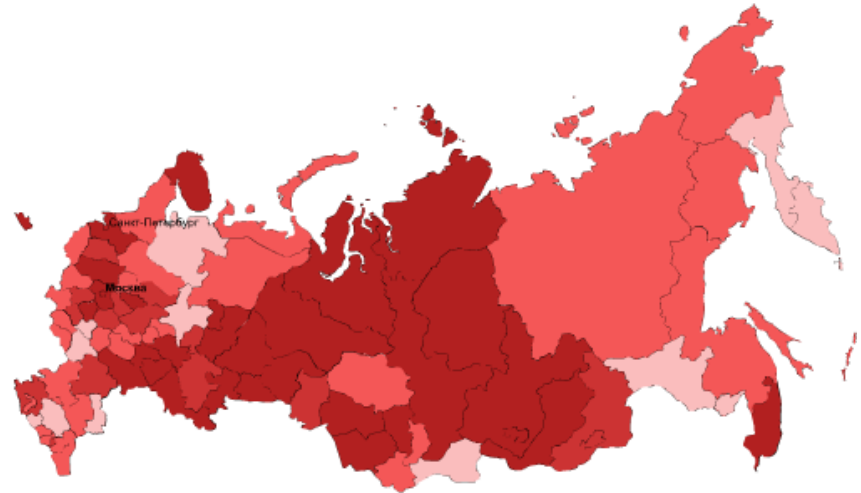
- На ранних стадиях протекает бессимптомно;
- Около 80% больных не имели случаев заболевания в семье.



- Задача: оценить количество потенциальных лет потерянной жизни в популяции при использовании различных методов выявления полипов

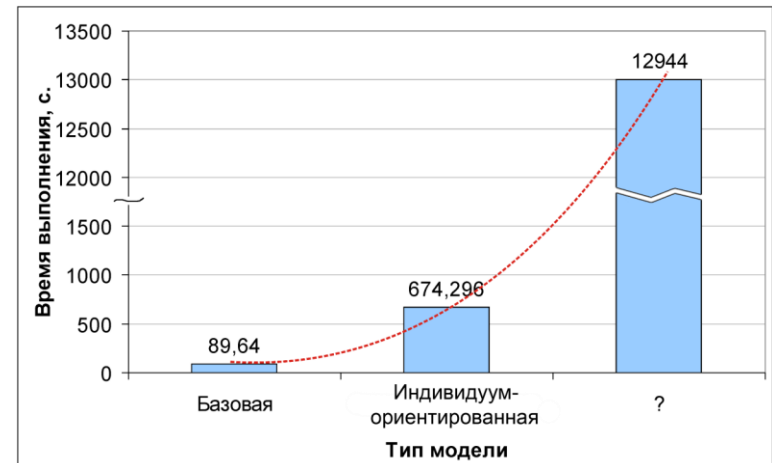
ИЗУЧЕНИЕ РАСПРОСТРАНЕНИЯ ВИЧ СРЕДИ ГРУПП ПОТРЕБИТЕЛЕЙ ИНЪЕКЦИОННЫХ НАРКОТИКОВ

- За 2013 год в РФ зарегистрировано 77 896 новых случаев ВИЧ, 57% случаев — из-за употребления наркотиков нестерильным инструментарием
- Практически нет моделей распространения ВИЧ среди ПИН, учитывающих российскую специфику
- Задача: создать модель распространения ВИЧ в группе ПИН и проанализировать влияние различных исходных параметров на динамику ВИЧ среди ПИН



ПОСТАНОВКА ПРОБЛЕМЫ

- В задачах имитационного моделирования возникает проблема ускорения вычислительных экспериментов



- Возможные способы ускорения вычислений:
 - Оптимизация исходного алгоритма
 - Переход к параллельным и распределённым вычислениям
 - Использование более мощных вычислительных систем

ИСПОЛЬЗУЕМЫЕ ТИПЫ МОДЕЛЕЙ

- Способ задания модели: цепи Маркова, системы стохастических разностных уравнений, дискретно-событийное описание
- Время: дискретное, непрерывное
- Учёт неоднородности популяции: популяционные модели, индивидуум-ориентированные модели, гибридные двухкомпонентные модели (совмещение обоих подходов)
- Порядок численностей популяций: от 10^2 до 10^6
- Язык моделирующих программ: C++, MATLAB



ОБЩИЕ СВОЙСТВА МОДЕЛЕЙ

- Все модели — имитационные и стохастические
- Существенная часть вычислительного времени тратится на генерацию псевдослучайных чисел
- Для получения оценок на выходные величины требуется многократное повторение вычислений с разными начальными значениями датчика псевдослучайных чисел
- Потребление памяти и количество операций записи-считывания данных невелико



ТРЕБОВАНИЯ К СПОСОБУ УСКОРЕНИЯ

- Желательные характеристики способа ускорения исходного алгоритма:
 - применимость к широкому классу задач (универсальность)
 - высокий прирост быстродействия моделирующей программы (эффективность)
 - низкие затраты на оптимизацию алгоритма (простота)

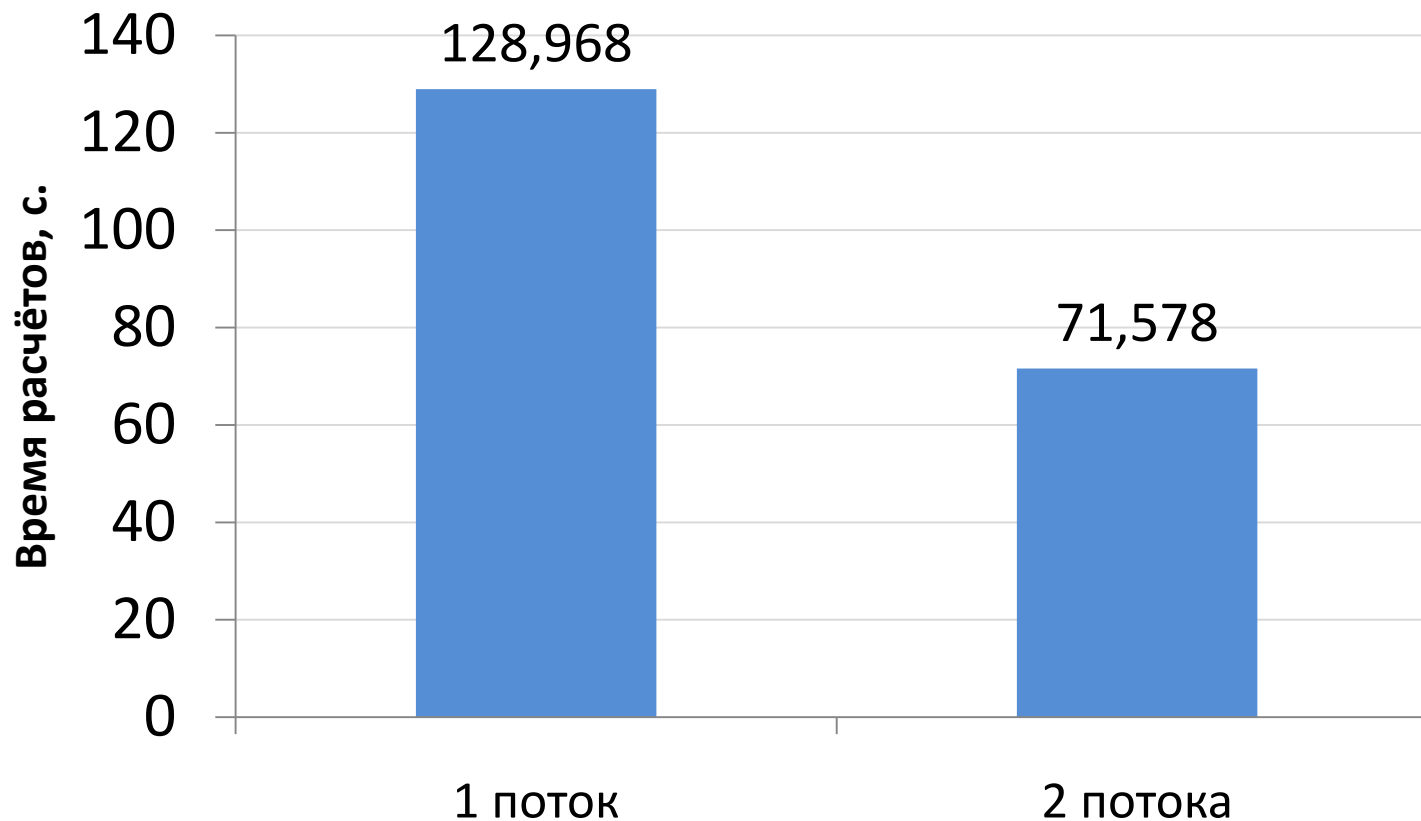


«УНИВЕРСАЛЬНЫЙ ПОДХОД»: ПАРАЛЛЕЛИЗМ ПО РЕАЛИЗАЦИЯМ

- Реализации распределяются между ядрами (вычислительными узлами) с помощью технологий OpenMP или MPI
- Возможно использование отдельных программ для распараллеливания (MONC, PARMONC)
- Подходит для большинства задач, связанных с имитационным моделированием методом Монте-Карло
- Вычисления проводятся на многоядерных ПК или многопроцессорных суперкомпьютерах



МАСШТАБИРУЕМОСТЬ АЛГОРИТМА НА ОСНОВЕ ТЕХНОЛОГИИ OPENMP

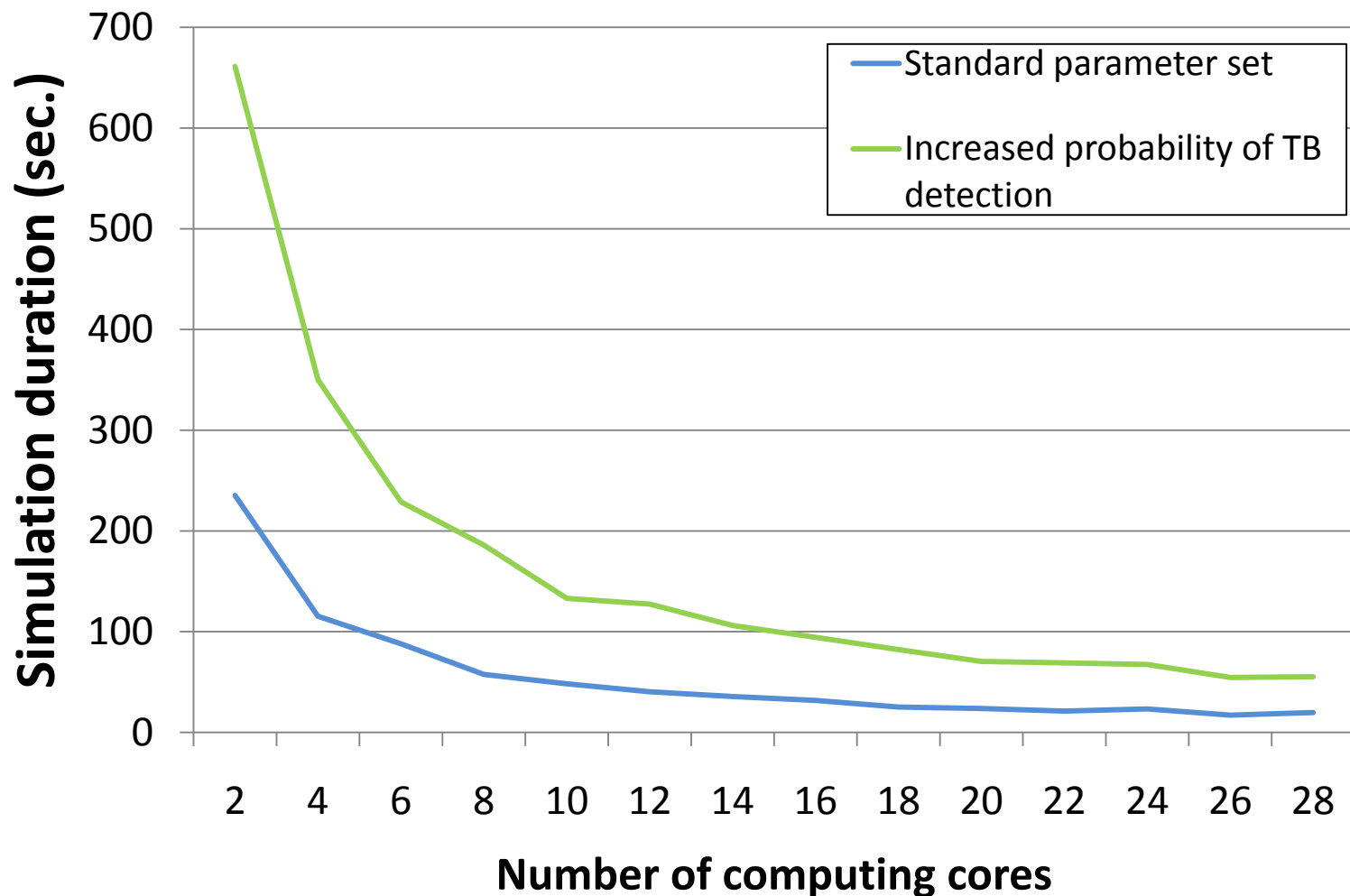


Количество потоков

Время вычислений для популяционной модели
распространения туберкулёза, ПК Intel Core2 Duo 2,8 ГГц



МАСШТАБИРУЕМОСТЬ АЛГОРИТМА НА ОСНОВЕ ТЕХНОЛОГИИ MPI



Время вычислений для гибридной модели распространения туберкулёза, суперкомпьютер НКС-30Т Сибирского суперкомпьютерного центра (ядра Intel Xeon E5450 3 GHz)



«ЭФФЕКТИВНЫЙ ПОДХОД»: ПАРАЛЛЕЛЬНЫЕ ВЫЧИСЛЕНИЯ НА GPU

- Алгоритм моделирования разбивается на большое количество «легковесных» потоков, выполняемых на GPU
- Рост производительности на среднем ПК – на порядок и выше
- Алгоритм с требуемыми свойствами построен для класса моделей, в которых:
 - большие индивиды не влияют друг на друга и на восприимчивых индивидов (неинфекционные заболевания);
 - демографические процессы (рождение, миграция) отсутствуют (чисто дискретно-событийные модели без популяционной составляющей).

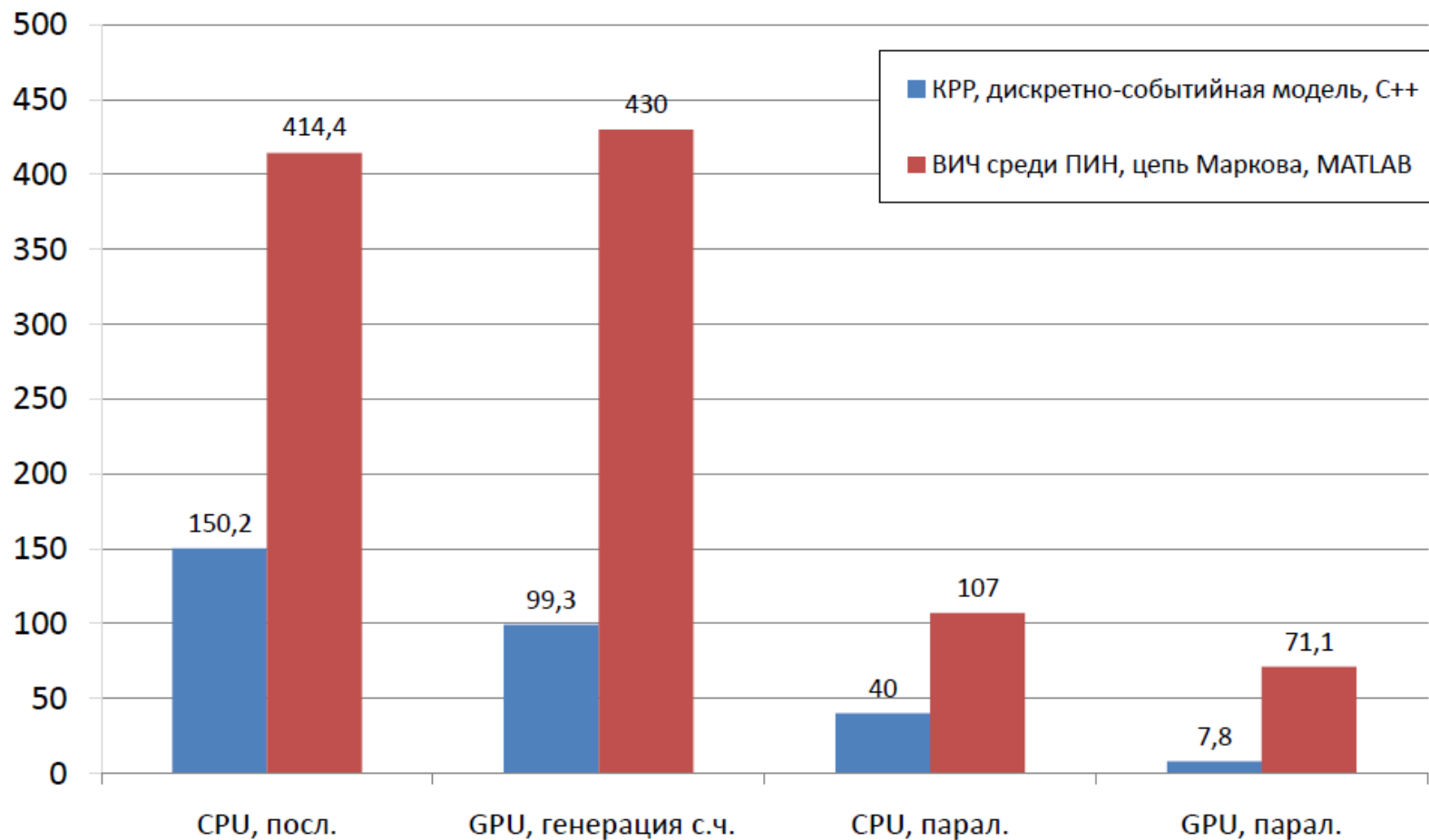


«ПРОСТОЙ ПОДХОД»: ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ЧИСЕЛ НА GPU

- GPU используется для генерации массивов псевдослучайных чисел (стандартные команды библиотеки CURAND для CUDA C++, команды для gpuArray в MATLAB)
- Сам вычислительный алгоритм выполняется на CPU
- В идейном плане подход очень прост, но не всегда даёт выигрыш в производительности



ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ



- Время вычислений на ПК с CPU Intel Core i5-3450, GPU NVIDIA GeForce GTX 550 Ti

ВОЗМОЖНЫЕ ПЕРСПЕКТИВЫ

- Новые варианты параллельных алгоритмов на GPU для конкретных классов моделей (рост универсальности «эффективного» алгоритма)
- Совместное применение OpenMP, MPI и CUDA для высокодетализированных моделей и вычисления на суперкомпьютерных кластерах с гибридными узлами CPU+GPU



СПАСИБО ЗА ВНИМАНИЕ!

E-mail: VNLeonenko@yandex.ru

