ПОСТРОЕНИЕ ЭФФЕКТИВНЫХ МЕТОДОВ ПАРАЛЛЕЛЬНОГО ВЫПОЛНЕНИЯ ЧИСЛЕННЫХ РАСЧЕТОВ НА ОСНОВЕ ПОЛИНОМИАЛЬНОЙ СИСТЕМЫ КЛАССОВ ВЫЧЕТОВ

К.С. Исупов

В настоящее время в различных областях науки и техники возрастает потребность решения задач большой размерности, которые требуют обработки огромных массивов цифровой информации. К таким задачам относятся, например, в климатологии - проведение численного эксперимента с глобальной моделью атмосферы, в энергетике - расчет режима энергосистемы, в аэродинамике – исследование аэродинамических характеристик летальных аппаратов [1]. Основу таких задач составляют численные расчеты в базисе матричной алгебры.

Одной из существенных проблем решения таких задач на ЭВМ является накопление ошибок округления в ходе вычислительного процесса, поэтому для получения корректных результатов при решении вышеприведенных задач широко используются вычисления с многократной точностью над массивами чисел, выходящих за пределы разрядной сетки ЭВМ, которые так же называются *большими числами*.

Выполнение даже элементарных операций с большими числами становится затруднительным и часто требует специального подхода. Как правило, современные процессоры оптимизированы для операций с 32- или 64-битными операндами, все подсистемы ввода/вывода также имеют такую разрядность, поэтому производительность при работе с типичными N-разрядными числами, где N больше 64, катастрофически падает. Ввиду этого ставится задача решения так называемой проблемы больших чисел.

Имеющиеся на сегодняшний день готовые решения для выполнения численных расчетов в базисе матричной алгебры, такие как SMath Studio, FreeMat, MathAdd, Maxima, MatLab, MAPLE, Mathematica и др. опираются на парадигмы программирования 80х-90х годов и не предназначены для выполнения на мультипроцессорных ЭВМ, процессоры которых имеют многоядерную архитектуру. Кроме этого их базовые структуры данных не ориентированы на решение проблемы больших чисел.

Поэтому необходимы принципиально новые программные продукты, использующие специальные подходы для эффективного использования ресурсов современных мультипроцессорных ЭВМ с кластерной архитектурой и максимально быстрого выполнения операций над большими числами.

Одним из перспективных и многообещающих направлений в исследованиях по данной проблеме является внедрение методов вычислений на основе нетрадиционных способов кодирования и соответствующих им вариантов машинной арифметики. Особую роль в развитии указанного направления играют числовые системы с параллельной структурой и, в первую очередь, модулярные системы счисления, в которых целые числа представляются наборами остатков от деления на выбранные модули (основания).

Одной из самых известных таких систем является полиномиальная система классов вычетов, так же известная как система остаточных классов (СОК). Здесь в отличие от обобщенной позиционной системы счисления (ПСС) образование цифры каждого разряда проводится независимо друг от друга [2].

Каждый разряд в системе остаточных классов представляет собой наименьший положительный остаток от деления на соответствующее основание самого числа, а не предыдущего частного, как это имеет место в позиционной системе.

Если же касаться задач матричной алгебры, то здесь каждый массив, представленный в позиционной системе, в СОК будет представляться множеством независимых друг от друга массивов, элементы которых представляют собой наименьший положительный остаток от деления на выбранные основания соответствующих позиционных элементов. Очевидно, что мощность данного множества совпадает с числом оснований (модульностью) системы остаточных классов.

Остаток от деления меньше делителя, поэтому, если в качестве оснований полиномиальной системы выбрать числа, не выходящие за пределы разрядной сетки ЭВМ, то и вычеты, взятые по этим основаниям, не будут выходить за пределы разрядной сетки. К тому же, выполнение операций в полиномиальной системе практически ничем не отличается от выполнения операций в традиционных позиционных системах счисления.

В отличие от позиционных систем, где заложена фиксированная градация типов числовых данных (Byte, Int32, Int64 и т.д.), система остаточных классов может быть настроена на конкретную разрядность числовых данных. Данная настройка осуществляется путем выбора оптимальных для работы с конкретными числовыми данными оснований СОК.

Оптимальными в данном случае можно считать взаимно-простые основания, произведение которых (а, следовательно, и верхний предел диапазона представления числовых данных) максимально приближено к конкретной разрядности чисел, работа с которыми осуществляется в определенный момент времени. С другой стороны, оптимальными для использования полиномиальной системы на вычислительных машинах, основанных на двоичной арифметике, являются числа вида "2^N-1", где число N не выходит за пределы разрядной сетки ЭВМ [3].

При этом рекомендуется, чтобы основания имели одинаковую разрядность, так как только в этом случае, при параллельной обработке массивов данных на многоядерных процессорах мультипроцессорной системы удастся достигнуть наиболее сбалансированной полезной загрузки всех ядер. При этом каждое ядро будет производить работу с фрагментами массивов данных только по соответствующему ему основанию системы остаточных классов. Т.е. предполагается, что ядра всех процессоров отдельно взятого вычислительного узла в каждый момент времени будут выполнять одни и те же операции над различными данными одинаковой разрядности.

Из этого следует, что полиномиальная система классов вычетов по своему существу предполагает SIMD (Simple Instruction Multiple Data) модель вычислений.

Из всего сказанного можно судить о ценности исследований, посвященных оптимальному использованию системы остаточных классов и определению возможности ее внедрения в качестве базиса для выполнения высокоточных параллельных численных расчетов на современных высокопроизводительных вычислительных системах с кластерной архитектурой.

На основании предложенного подхода был разработан инструментальный комплекс для проектирования параллельных масштабируемых программ численных расчетов в базисах матричной алгебры и системы остаточных классов.

Ядром комплекса является вычислительный модуль, непосредственно выполняющий численные расчеты в базисе матричной алгебры над данными, представленными в полиномиальной системе. Вычислительный модуль производит масштабирование вычислений и выполнение одного из численных расчетов, согласно заданному алгоритму и входным параметрам.

Особенностью комплекса является обособленность исполнительных модулей от пользовательского интерфейса, что позволяет выполнять их в виде динамических библиотек и встраивать в современные промышленные математические системы и специализированные программные комплексы.

Для оптимизации использования системы остаточных классов и определения возможности ее внедрения в качестве базиса для выполнения высокоточных параллельных численных расчетов была проведена серия экспериментов, включающая в себя следующие тесты: - исследование влияния числа элементов массивов данных на время решения задачи; - исследование влияния разрядности числовых данных на время решения задачи; - исследование эффективности применения грубой аппроксимации оснований СОК вида "2^N-1" вокруг диапазона представления числовых данных; - исследование эффективности применения сглаженной аппроксимации оснований СОК вокруг диапазона представления числовых данных.

Все эксперименты производились на кластере Вятского Государственного Университета ENIGMA (HP Hewlett-Packard Cluster Platform 3000 BL460c, Xeon 5345 2.33GHz, Infiniband). В распоряжении находился один вычислительный узел, содержащий два процессора, по четыре ядра каждый.

В качестве выполняемой операции была выбрана операция сложения двумерных массивов.

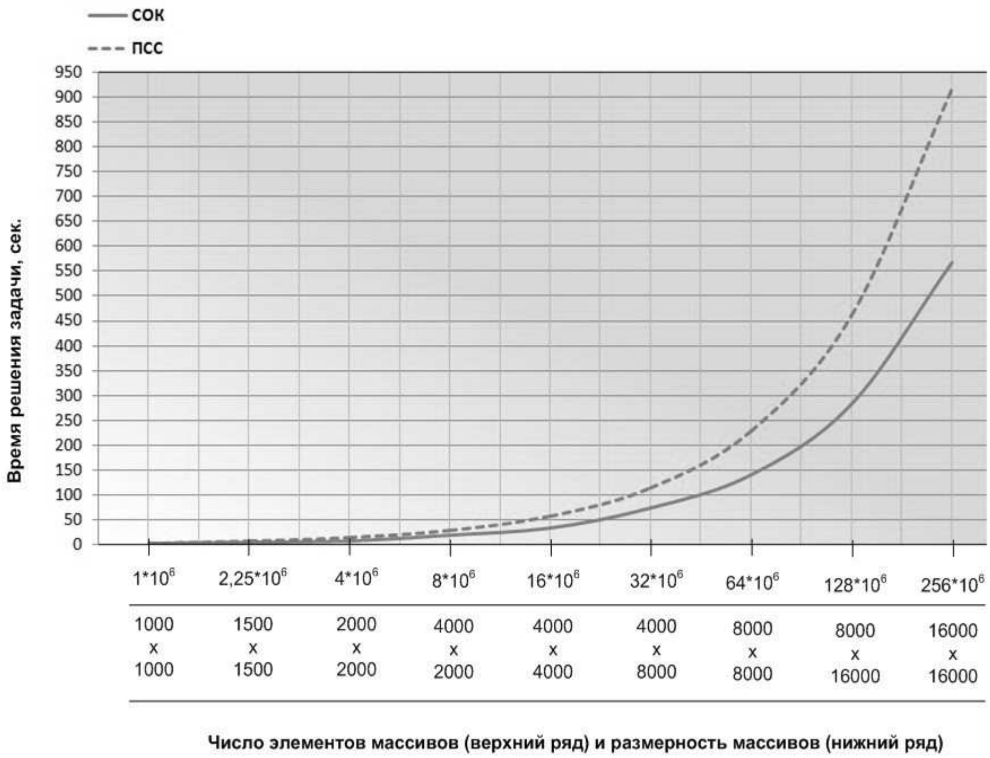

Рис. 1 График зависимости времени решения задачи сложения массивов в полиномиальной и позиционной системах от числа элементов массивов данных

Как при использовании СОК, так и при использовании десятичной системы исходные массивы разбивались на горизонтальные полосы, обрабатываемые параллельно, т.е. алгоритм выполнения расчетов подразумевает параллелизм на уровне элементов данных.

Перед вычислениями производилась подготовка массивов данных (разбиение на горизонтальные полосы и преобразование в полиномиальную систему, в случае ее использования в качестве базиса). Для каждой конфигурации производилось три тестовых запуска. В качестве результата принималось усредненное значение исследуемого параметра.

График зависимости времени решения задачи сложения массивов в полиномиальной и позиционной системах от числа элементов массивов данных представлен на рисунке 1.

В качестве элементов массивов были использованы псевдослучайные числа из неизменного диапазона (от 1x10^9 до 2,147x10^9).

Из графика можно заметить, что время решения задачи связано с размерностью массивов входных данных квадратичной зависимостью, т.е. при неизменной аппаратно-программной платформе, на которой производятся параллельные вычисления, время решения задачи линейно возрастает с увеличением числа обрабатываемых элементов (число обрабатываемых элементов квадратично возрастает с увеличением размерности массивов данных).

Можно предположить, что время решения задачи будет линейно уменьшаться с увеличением числа вычислительных узлов.

На всем диапазоне данных, при неизменной разрядности чисел использование полиномиальной системы сокращает время решения задачи по отношению к позиционной (десятичной) системе счисления в среднем на 40%.

Средняя полезная загрузка процессорных элементов при использовании системы остаточных классов составляет 100% (используется все 8 ядер, миграции процессов между ядрами не происходит), а при использовании позиционной системы – приближенно 30% (используется 3 ядра, наблюдается миграция процессов между ядрами).

График зависимости времени решения задачи сложения массивов 4000x8000 в полиномиальной и позиционной системах от разрядности числовых данных представлен на рисунке 2.

Основания СОК выбирались из поля чисел "2^N-1". В качестве чисел N были выбраны значения: 11, 9, 8 и 7. Верхний предел диапазона представления чисел (максимальная разрядность) при выбранных основаниях составляет 3,387x10^10, что в восемь раз превосходит предел типа unsigned long.

В общем случае, в каждом новом эксперименте разрядность псевдослучайных чисел увеличивается на порядок (в 10 раз).

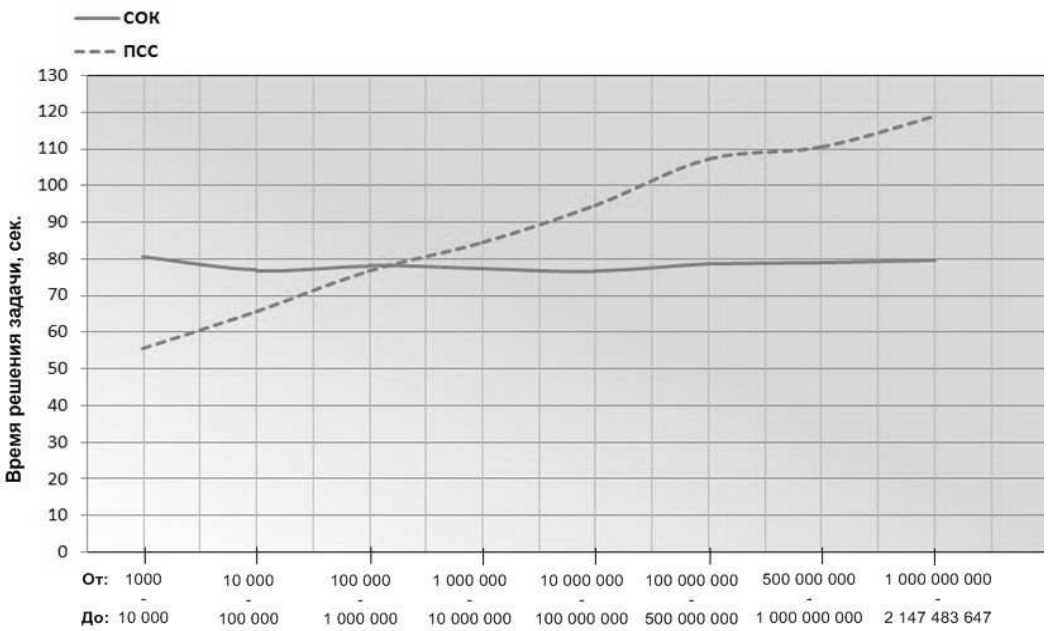

Рис. 2 График зависимости времени решения задачи сложения массивов в полиномиальной и позиционной системах от разрядности числовых данных

При выполнении вычислений в СОК время решения задачи практически не зависит от разрядности обрабатываемых чисел, тогда как при использовании в качестве базиса для вычислений позиционной системы, с увеличением нижнего и верхнего пределов диапазона представления на порядок время решения задачи существенно увеличивается. Можно сделать вывод, что в данном случае позиционную систему выгоднее использовать для работы с семи- и менее разрядными числами (от 0 до 3x10^6). Данная зависимость вызвана различным числом обращений к файловой системе при работе с различными числовыми данными.

В данном случае диапазон представления чисел, обеспечиваемый выбранными основаниями полиномиальной системы, постоянен и верхний его предел равен 3,387x10^10. Очевидно, что при выборе других оснований СОК данная зависимость и диапазон представления будут выглядеть иначе.

Стоит заметить, что комплекс рассчитан на выполнение операций над массивами данных порядка 10^6 и более, поэтому работа с файловой системой в данном случае неизбежна.

Расстояние степеней - максимальная разность между степенями N двух соседних оснований СОК вида "2^N-1".

Грубая аппроксимация подразумевает приближение оснований к требуемому диапазону с учетом постоянного максимального расстояния степеней, равного двум, т.е. разность степеней двух соседних оснований полиномиальной системы может равняться "1", либо "2".

Основания СОК выбирались из поля чисел "2^N-1". В качестве степеней N выбирались взаимно-простые числа, с таким условием, чтобы произведение всех оснований максимально близко приближалось к верхнему пределу требуемого диапазона представления числовых данных в конкретном эксперименте, т.е. аппроксимировало данный диапазон сверху.

График приближения оснований системы остаточных классов к требуемому диапазону представления числовых данных для постоянного максимального расстояния степеней N, равного двум, представлен на рисунке 3. Используется логарифмическая шкала.

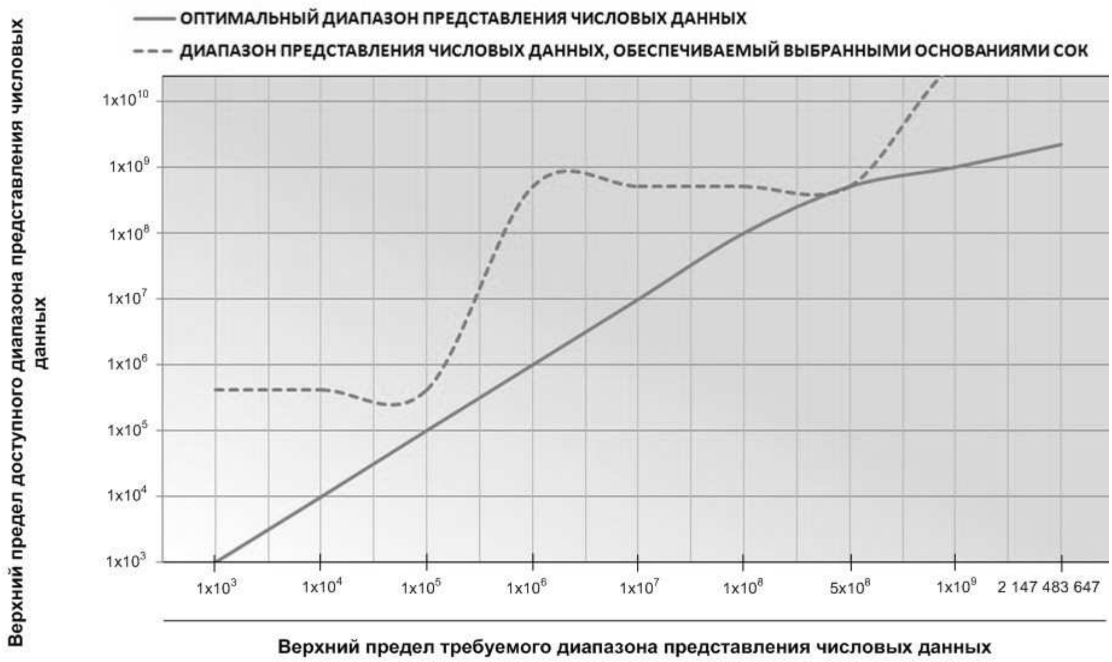

Рисунок 3 - График приближения оснований системы остаточных классов к требуемому диапазону представления числовых данных для постоянного максимального расстояния степеней N, равного двум

Из графика видно, что приближение (пунктирная линия) к верхнему пределу требуемого диапазона представления числовых данных (сплошная линия) является достаточно грубым. Так, при работе с трех- и менее разрядными числами (от 0 до 1x10^3) согласно алгоритму грубой аппроксимации были сгенерированы основания полиномиальной системы, обеспечивающие корректную работу с шести-разрядными числовыми данными (с числами от 0 до 4,134x10^5). При работе с числами из диапазона от 1x10^6 до 1x10^7 были сгенерированы основания, позволяющие работать с девяти-разрядными числовыми данными (с числами, не превышающими 5,13x10^8). Наиболее близкое приближение наблюдается при работе с числовыми данными, верхний предел диапазона представления которых равняется 5x10^8. При этом были сгенерированы основания, обеспечивающие работу с числами, не превышающими 5,13x10^8.

График зависимости времени решения задачи сложения двумерных массивов размерности 4000x8000 от разрядности числовых данных после применения грубой аппроксимации оснований вокруг диапазона представления числовых данных представлен на рисунке 4.

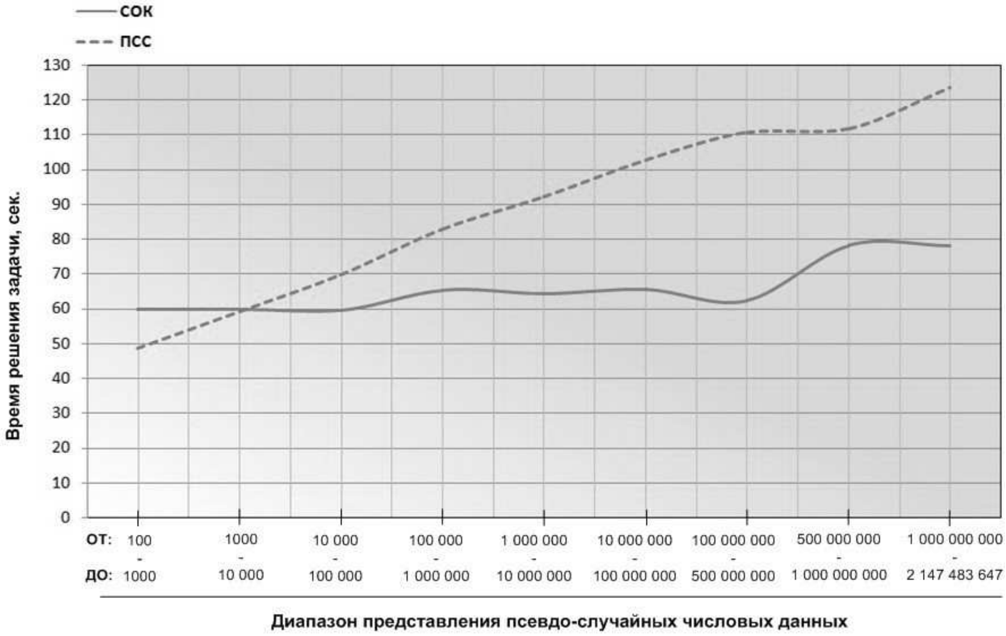

Рис. 4 График зависимости времени решения задачи сложения двумерных массивов от разрядности числовых данных после применения грубой аппроксимации оснований

Путем применения грубой аппроксимации оснований полиномиальной системы удалось повысить эффективность ее использования относительно десятичной системы счисления. Если без аппроксимации оснований, позиционную систему было выгодно использовать для работы с семи- и менее разрядными числовыми данными (с числами, не превышающими 3x10^6), то теперь она выгоднее лишь при работе с трех- и менее разрядными числовыми данными (с числами, не превышающими 1x10^3). Из этого следует, что даже грубое приближение верхнего предела диапазона представления числовых данных к оптимальному диапазону обеспечивает существенное сокращение времени решения задачи, а выбранный базис "подстраивается" под конкретные числовые данные.

Сглаженная аппроксимация подразумевает приближение оснований к требуемому диапазону с учетом изменяющегося максимального расстояния степеней N, при этом в качестве оптимального выбирается расстояние, при котором произведение оснований СОК наиболее приближенно к верхнему пределу диапазона представления числовых данных.

График приближения оснований полиномиальной системы к верхнему пределу диапазона представления числовых данных при использовании сглаженной аппроксимации (динамически изменяющегося расстояния степеней) представлен на рисунке 5. Используется логарифмическая шкала.

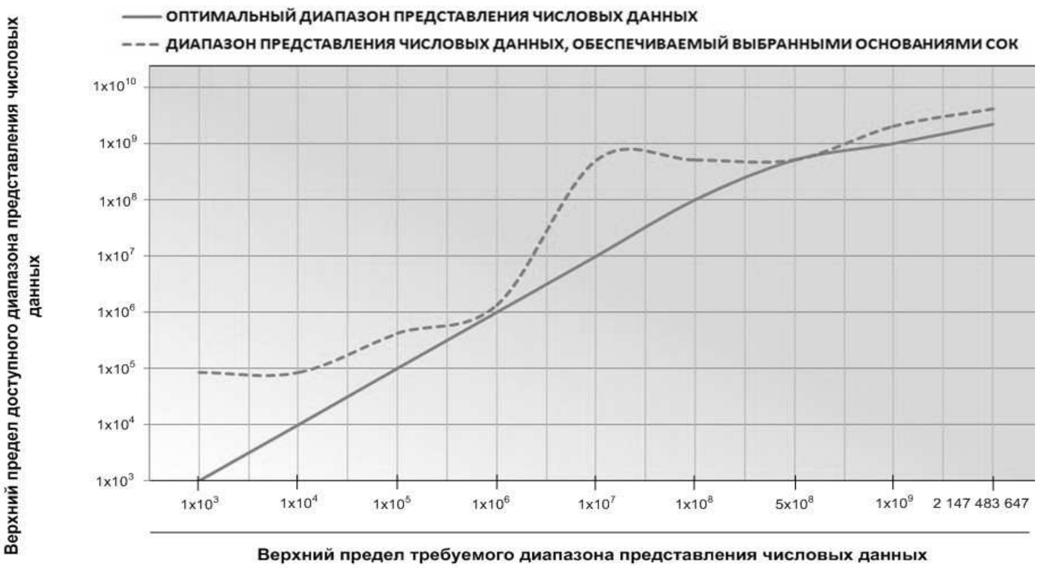

Рисунок 5 - График приближения оснований полиномиальной системы к верхнему пределу диапазона представления числовых данных при использовании сглаженной аппроксимации

Можно заметить, что путем применения сглаженной аппроксимации удалось существенно сократить разность между верхним пределом требуемого диапазона представления числовых данных (сплошная линия) и приближением к требуемому верхнему пределу диапазона, обеспечиваемым выбираемыми основаниями полиномиальной системы (пунктирная линия).

В данном случае, наиболее эффективное использование полиномиальной системы наблюдается при работе с пяти и восьми-разрядными числовыми данными (с числами из диапазона от 1x10^5 до 1x10^6, и от 1x10^8 до 1x10^9 соответственно), а наиболее неэффективное – при работе с семи-разрядными данными (с числами из диапазона от 1x10^7 до 1x10^8), т.к. здесь разность между приближением и оптимальным верхним пределом диапазона наибольшая.

Уменьшить эту разность для четырехмодульной полиномиальной системы, в которой основания выбираются из поля чисел "2^N-1" невозможно, т.к. не существует набора из четырех взаимно-простых чисел вида "2^N-1", произведение которых составляло бы менее 5,13x10^8, но более 1x10^6.

Поэтому для дальнейшего повышения эффективности использования СОК необходимо выбирать ее взаимно-простые основания не ограничиваясь при этом полем чисел "2^N-1", например, из последовательности Фибоначчи, либо из других "примечательных" последовательностей.

График зависимости времени решения задачи сложения двумерных массивов 4000x8000 от разрядности числовых данных при использовании сглаженной аппроксимации оснований представлен на рисунке 6.

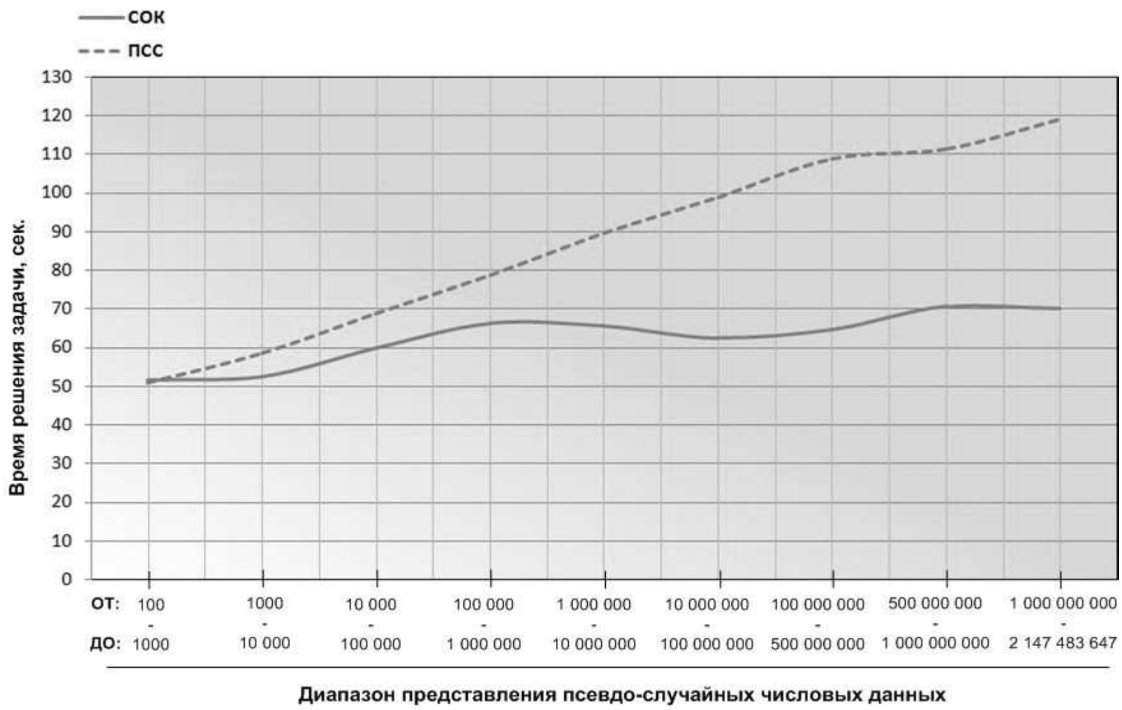

Рисунок 6 - График зависимости времени решения задачи сложения двумерных массивов от разрядности числовых данных при использовании сглаженной аппроксимации оснований

После применения сглаженной аппроксимации, работа с позиционной системой продолжает быть эффективной лишь для двух- и менее разрядных числовых данных (для чисел, не превышающих 100). В остальных случаях эффективность выполнения расчетов в полиномиальной системе возрастает с увеличением разрядности числовых данных и при работе с числами из диапазона от 10^9 до 2,147x10^9 время решения задачи сокращается на 42%.

Сводный график времени решения задачи сложения двумерных массивов размерности 4000x8000 при использовании системы остаточных классов: без аппроксимации оснований вокруг диапазона представления числовых данных, с использованием грубой аппроксимации и с использованием сглаженной аппроксимации представлен на рисунке 7.

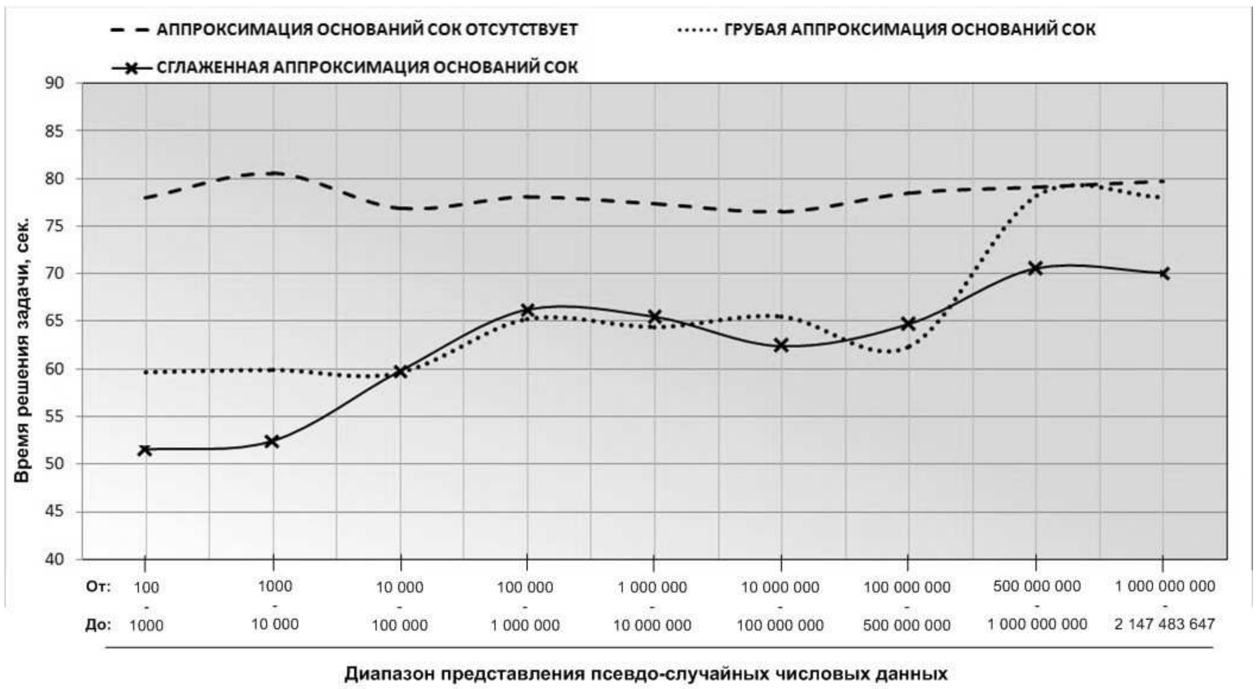

Рисунок 7 - Сводный график времени решения задачи сложения двумерных массивов в полиномиальной системе классов вычетов

Сводный график наглядно показывает уменьшение времени решения задачи с введением того или иного вида аппроксимации оснований СОК вокруг диапазона представления числовых данных. Принимая тот факт, что в данном случае эффективность обратно пропорциональна времени решения задачи, самым эффективным является использование полиномиальной системы с применением к ее основаниям сглаженной аппроксимации. При этом система наиболее точно "подстраивается" под конкретную разрядность числовых данных, что позволяет избежать накладных расходов по обработке оснований, разрядность которых превышает требования конкретного диапазона представления.

С помощью проведенных экспериментов удалось оптимизировать и показать целесообразность использования полиномиальной системы с взаимно-простыми основаниями вида "2^N-1" в качестве базиса для выполнения численных расчетов.

ЛИТЕРАТУРА:

1. Ю.Д. Валеев Система распараллеливания алгоритмов компьютерной алгебры на основе арифметики полиномов : автореф. дис. …канд. физ-мат. наук: 05.13.11 / Тамбовский. гос. ун-т. им. Г.Р. Державина. – Тамбов, 2008. – 24 с.
2. И.Я. Акушский Машинная арифметика в остаточных классах / И.Я. Акушский, Д.И. Юдицкий. – М.: Сов. Радио, 1968. – 440 с.
3. Д. Кнут Сортировка и поиск / Дональд Кнут. – М.: "Вильямс", 2007. – 824с. – (Искусство программирования /Дональд Кнут; том 3).