

## 各種ベンチマーク

### Parsec

The PARSEC benchmark suite: characterization and architectural implications

Christian Bienia, et al, 2008

Unlike Woo et al. who measured actual concurrency on an abstract machine, we therefore decided to analyze inherent concurrency and its limitations. Our approach is based on the number of executed instructions in parallel and serial regions of the code. We neglect any delays due to blocking on contended locks and load imbalance. This methodology is feasible because we do not study performance, our interest is in fundamental program characteristics.

- blackscholes
  - calculates the prices for a portfolio of European options analytically with the Black-Scholes partial differential equation(PDE)
- bodytrack
  - tracks a human body with multiple cameras through an image sequence
- canneal
  - uses cache-aware simulated annealing(SA) to minimize the routing cost of a chip design
  - uses fine-grained parallelism with a lock-free algorithm and a very aggressive synchronization strategy that is based on data race recovery instead of avoidance
- dedup
  - compresses a data stream with a combination of global and local compression that is called 'deduplication'
  - the kernel uses a pipelined programming model to mimic real-world implementations
- facesim
  - computes a visually realistic animation of the modeled face by simulating the underlying physics
- ferret
  - is based on the Ferret toolkit which is used for content-based similarity search.
  - represents emerging next-generation search engines for non-text document data types
  - is parallelized using the pipeline model
- fluidanimate
  - uses an extension of the Smoothed Particle Hydrodynamics(SPH) method to simulate an incompressible fluid for interactive animation purposes
- freqmine
  - employs an array-based version of the FP-growth(Frequent Pattern-growth) method for Frequent Itemset Mining(FIMI)
- streamcluster
  - solves the online clustering problem
  - the prevalence of problems with streaming characteristics
- swaptions
  - uses the Health-Jarrow-Morton(HJM) framework to price a portfolio of swaptions
  - employs Monte Carlo(MC) simulation to compute the prices
- vips
  - is based on the VASARI Image Processing System(VIPS)
  - includes fundamental image operations such as an affine transformation and a convolution
- x264
  - is an H.264/AVC(Advanced Video Coding) video encoder

### ALPBench

## The ALPBench Benchmark Suite for Complex Multimedia Applications

Man-lap Li, et al, 2005

- MPEG-2 Encoder(MPGenc)
- MPEG-2 Decoder(MPGdec)
- RayTracing(RayTrace)
- Speech Recognition(SpeechRec)
- Face Recognition (FaceRec)

## MediaBench II Video

### MediaBench II video: Expediting the next generation of video systems research

Jason E. Fritts, et al, 2009

- H.263
  - A video coder(h263enc) and decoder(h263dec) based on the ITU H.263 standard targeting video compression for transmission over ISDN networks.
  - Source code produced by Telenor R&D.
- H.264
  - A video coder(h264enc) and decoder(h264dec) based on the forthcoming joint ISO/ITU H.264 standard(also known as MPEG-4 part 10) for very low bit-rate video coding.
  - Source code is the test model produced by the H.264 working group.
- Motion-JPEG
  - A video coder(jpegenc) and decoder(jpegdec) based on the ISO JPEG standard for image compression.\*\* Source code produced by the Independent JPEG Group.
- Motion-JPEG2000
  - A video coder(jp2Kenc) and decoder(jp2Kdec) based on the recent ISO JPEG standard for wavelet-based image compression.
  - Source code is the JasPer library for JPEG-2000.
- MPEG-2
  - A video coder(mpeg2enc) and decoder(mpeg2dec) based on the ISO MPEG-2 standard for high-quality video coding.
  - Source code produced by the MPEG Software Simulations Groups(MSSG)
- MPEG-4
  - A video coder(mpeg4enc) and decoder(mpeg4dec) based on the recent ISO MPEG-4 standard for object-based and very-low bit-rate video coding.
  - Source code is the ffmpeg library for audio/video coding.

## NU-MineBench

### NU-MineBench 2.0

J. Pisharath, et al, 2005.

- k-Means(clustering)
  - is a partition-based method and is arguably the most commonly used clustering technique.
  - "hard" clusters, since any data object either is or is not a member of a particular cluster
- Fuzzy k-Means(clustering)
  - K-means algorithm relaxes the condition.
- BIRCH(clustering)
  - is one of the hierarchical clustering methods that employ a hierarchical tree to represent the closeness of data objects.
- HOP(clustering)

- Density-based methods grow clusters according to the density of neighboring objects or according to some other density function.
- highly scalable when applied to large databases
- Naive Bayesian(classification)
  - a simple statistical classifier uses an input training data-set to build a predictive model.
  - based on Bayes' Theorem
- ScalParC(classification)
  - a scalable decision tree based classifier
  - builds the decision tree by recursively splitting the training data-set
- Apriori(Association Rule Mining)
  - explores the level-wise mining using the Apriori property
- Eclat(Association Rule Mining)
  - uses a vertical database format instead of the hash trees(horizontal format) as in apriori
  - enables breaking the search space into small, independent, and manageable chunks
- Utility(Association Rule Mining)
  - based data mining technique where higher "utility" item-sets are identified from a database by considering different values of individual items as utilities.
- SNP(Bayesian network)
  - Single nucleotide polymorphisms(SNP) are DNA sequence variations that occur when a single nucleotide is altered in the genome sequence.
- GeneNet(Bayesian network)
  - uses a similar hill climbing algorithm as in SNP
  - the input data is the micro-array data
- SEMPHY(Expectation Maximization)
  - is a structure learning algorithm that is based on phylogenetic trees.
  - uses Structural Expectation Maximization(probability estimation) algorithm to address this complication
- Rsearch(Pattern Recognition)
  - uses a grammar based approach to achieve this goal.
  - uses SCFG(Stochastic Context-Free Grammar)
- SVM-RFE(Support Vector Machine)
  - is a feature selection method that uses SVM techniques to refine and identify the optimum feature set in the feature data.
- PLSA(Dynamic Programming)
  - uses a dynamic programming approach to identify the similar and diverged regions between two sequences(eg. biological DNA/protein sequence, or text strings)
  - since this method is dependent on the sequence length, it is computationally very intense.

## SPLASH-2

The SPLASH-2 programs: characterization and methodological considerations

Steven Cameron Woo, et al, 1995

- Barnes
  - the interaction of a system bodies(galaxies or particles, etc) in 3D over a number of time-steps, using the Barnes-Hut hierarchical N-body method
- Cholesky
  - The blocked sparse Cholesky factorization kernel factors a sparse matrix into the product of a lower triangular matrix and its transpose
  - LU との違い :sparse matrices, not globally synchronized between steps
- FFT
  - a complex 1-D version of the radix- $n$  six-step FFT algorithms
- FMM
  - simulates a system of bodies over a number of time-steps
  - in 2-D using a different hierarchical N-body method(the adaptive Fast Multipole Method)

- LU
  - factors a dense matrix into the product of a lower triangular and an upper triangular matrix
  - $n \times n$  の行列  $A$  は  $B \times B$  ブロックの  $N \times N$  の array に分割される . ( $n = N \times B$ )
- Ocean
  - studies large-scale ocean movements based on eddy and boundary currents
- Radiosity
  - computes the equilibrium distribution of light in a scene using the iterative hierarchical diffuse radiosity method
- Radix
  - The integer radix sort kernel
  - the permutation step in the final step requires all-to-all communication.
- Raytrace
  - renders a 3-D scene using ray tracing.
- Volrend
  - renders a 3-D volume using a ray casting technique.
  - data access are input-dependent and irregular, and no attempt is made at intelligent data distribution.

## BioBench

BioBench: A Benchmark Suite of Bioinformatics Applications

K. Albayraktaroglu, et al, 2005

- blastn
  - DNA sequence searching
- blastp
  - Protein sequence searching
- clustalw
  - Multiple sequence alignment
- fasta\_dna
  - DNA sequence searching
- fasta\_prot
  - Protein sequence searching
- hmmer
  - Sequence profile searching
- mummer
  - Genome-level alignment
- protpars
  - Phylogenetic analysis
- tigr
  - Sequence assembly

## BioParallel(in C++)

Last level cache (LLC) performance of data mining workloads on a CMP - a case study of parallel bioinformatics workloads

A. Jaleel, et al, 2006

- GeneNet
  - is used to measure the regulatory relationship between genes.
  - hill-climbing as its main search algorithm
- SNP

- is used to measure and understand the patterns of Single Nucleotide Polymorphisms.
- SEMPHY
  - is a tool for constructing phylogenetic trees.
- Support Vector Machines Recursive Feature Elimination
  - is used to eliminate gene redundancy from a given input data set in order to provide compact gene subset.
- PLSA
  - is used to identify the similarities or differences between two genetic sequences.

## 一言メモ

- たんたんめん食べた (Wed Jan 13 13:37:25 2010 +0900)