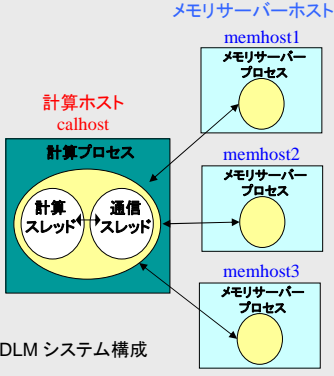


遠隔メモリを利用する大容量メモリシステム DLM(Distributed Large Memory)とDLMコンパイラ

緑川 博子(成蹊大), 黒川 原佳, 姫野 龍太郎(理研)

現状: 64BitOSの普及 広大なメモリ空間 X86_64(AMD64, Intel64) 現実装で 256tebibytes (2⁴⁸)
大容量データプログラムの実行: 物理メモリ、swapファイルサイズで制限を受ける

DLM システム ネットワークで結ばれたコンピュータの遠隔物理メモリを集め、**仮想的な大規模メモリを構築** するユーザレベルソフトウェア → 大規模データを扱うプログラムの実行が可能



DLM設定ファイル

使用ホストと使用メモリ容量を指定する

```
hostfile
calhost 32768 //32GB 計算ローカルホスト
memhost1 32768 //32GB メモリサーバホスト1
memhost2 65536 //64GB メモリサーバホスト2
memhost3 65536 //64GB メモリサーバホスト3
:
```

先頭行はプログラム実行ホストDLMに使用するメモリサイズを指定
2行目以降はメモリサーバに使用する遠隔ホストとDLMに提供できるメモリサイズ

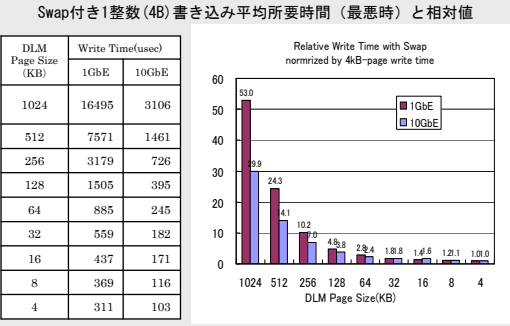
プログラム実行コマンド例

実行例 prog -- -n 4 -f hostfile

hostfileの先頭4行を用い、計算ホストとメモリサーバ3台を使用する指定例

DLM コンパイラ, DLMライブラリ dlmc test0.c -ldlm
ユーザ透過性: ユーザに遠隔メモリ利用を意識させない
通常プログラムに dlm を加えるだけで利用可能

性能評価1 DLMページサイズとswapオーバーヘッド



例文1 静的データ宣言 dlm double data[Size1][Size2][Size3];
例文2 動的データ割付 int *p; p = (int *) dlm_malloc (sizeof(int) * SIZE);

DLMプログラム例 (行列ベクトル積 DLM静的割当て例)

```
#include <stdio.h>
#define N 16384 // 2^14 * 2^15 B, 約2GiB
dlm double a[N][N], x[N], y[N]; // DLM使用

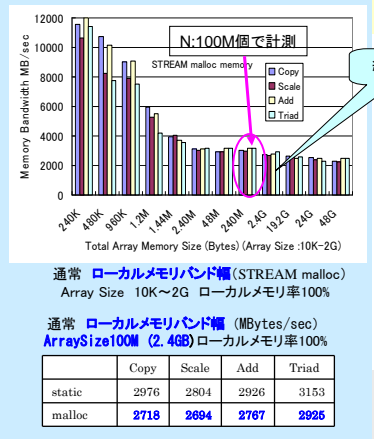
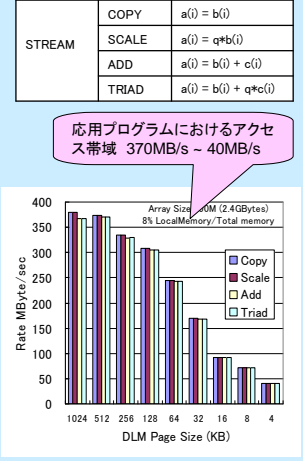
int main(int argc, char *argv[])
{
    int i, j;
    double temp;
    // 行列aを初期化
    for(i = 0; i < N; i++)
        for(j = 0; j < N; j++) a[i][j] = i;
    // ベクトルxを初期化
    for(i = 0; i < N; i++) x[i] = i;
    // a[N][N]*x[N]=y[N] 計算
    for(i = 0; i < N; i++){
        temp = 0;
        for(j = 0; j < N; j++) temp += a[i][j]*x[j];
        y[i] = temp;
    }
    return 0;
}
```

DLM Cコンパイラの構成

コンパイル例 dlmc test0.c -ldlm

コンパイラ構成: プログラム + dlm記述 → プリプロセッサ → トランスレータ → Cコンパイラ → 実行ファイル

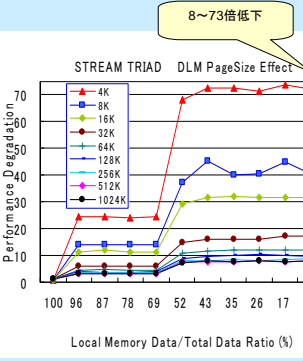
性能評価2 遠隔メモリバンド幅 (STREAMベンチマーク)



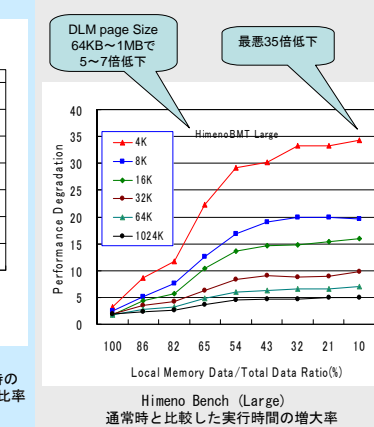
性能評価3 応用プログラム性能 (NPB, Himenoベンチマーク)

1.6~7.8倍低下

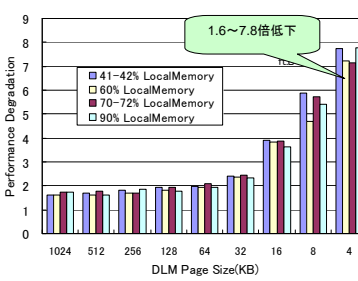
DLM使用時 遠隔メモリバンド幅 (STREAM) Array Size100M (2.4GB) ローカルメモリ率8%



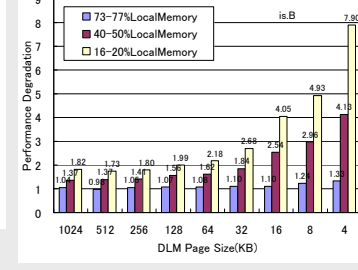
参考バンド幅 (Storage 転送帯域仕様). (実アクセスにはHDのSeek時間などを要す.)
Swapデバイス利用: OSのSwapオーバーヘッド付加
Serial ATA SATA-300 375MB/s
Serial SCSI SAS 375MB/s
Fiber Channel 4GFC (4.25GHz) 425MB/s
(Serial ATA, SATA-600) 600MB/s 予定
(Serial Attached SCSI2) 750MB/s 予定



性能評価3 応用プログラム性能 (NPB, Himenoベンチマーク)



FT.B 通常時と比較した実行時間の増大率



通常時(ローカルメモリ100%)に対するDLM使用時の性能低下度 STREAM (TRIAD) DLMサイズとRM/Total比率データアクセスワーキングセット

IS.B 通常時と比較した実行時間の増大率