

3歩あるいて忘れる鶏の群でなく ～OSとノーマリオフコンピューティング～

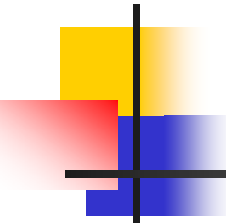
並木美太郎(東京農工大学)



ストアードプログラム型計算機の本質

- 演算系、状態保持系、I/O系
- 演算系は先ほどのとおり
- では？状態保持系は？

(実はこれこそが本質的)



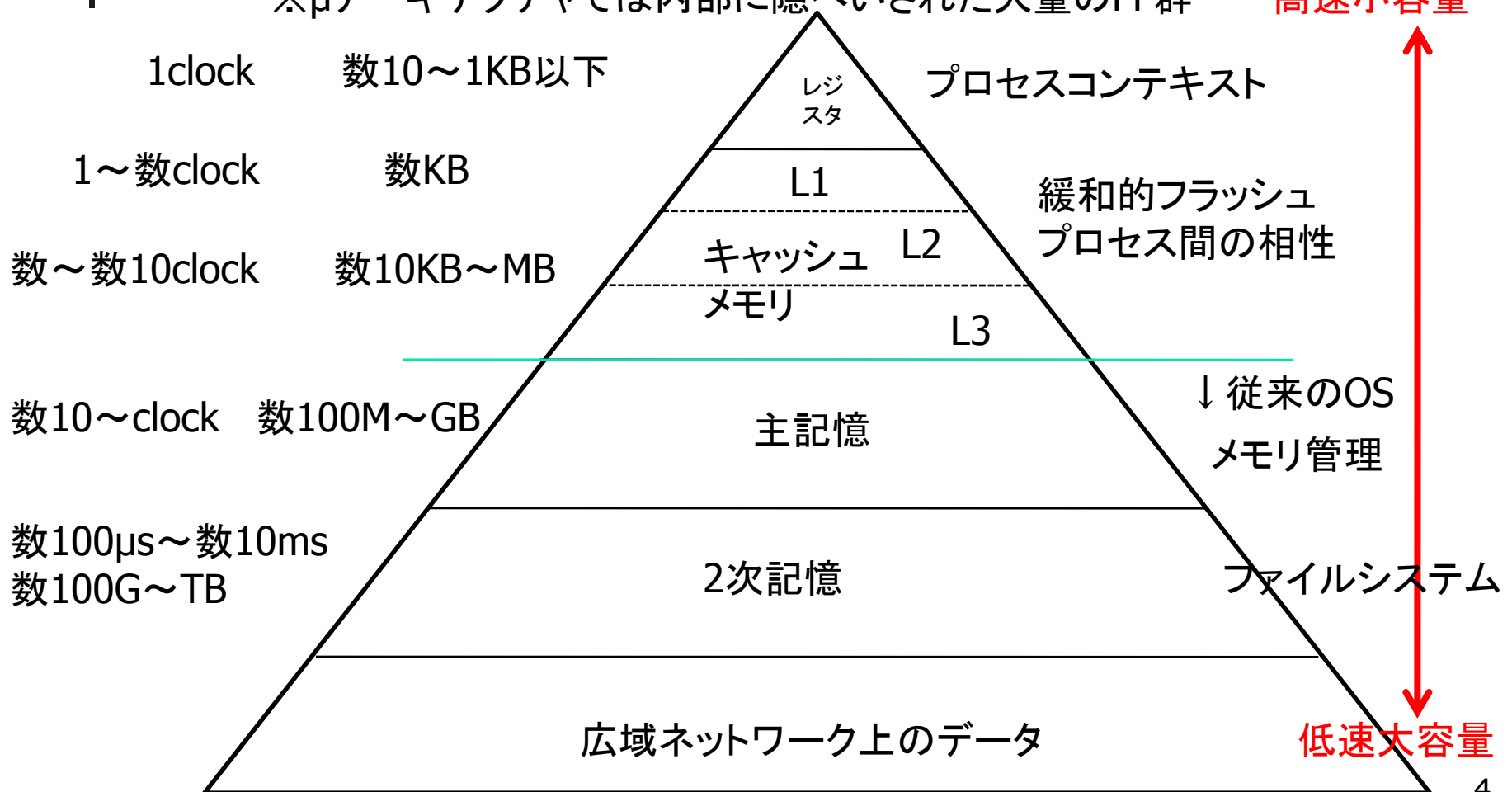
自分のノーマリオフコンピューティングのイメージ

- 普段は通電されてなく、必要になったときに電源が入って利用できる、計算機、ソフトウェア、ネットワークの構成方式
- だから、とっても省電力
- 構成要素、アーキテクチャ、ソフトウェアのいずれにも適用可能な概念
- 対象によって議論が微妙に食い違う部分と、普遍的な本質的なところと両面あると思われる
- 何が犠牲になるかは、まだよくわからない(win-winなのか?)
- ただし、新しいデバイス(不揮発メモリ)を想定すると興味をそそられる
- 多くのソフト屋は、この言葉はまだ知らない

メモリ系：計算機の記憶階層

※ μ アーキテクチャでは内部に隠ぺいされた大量のFF群

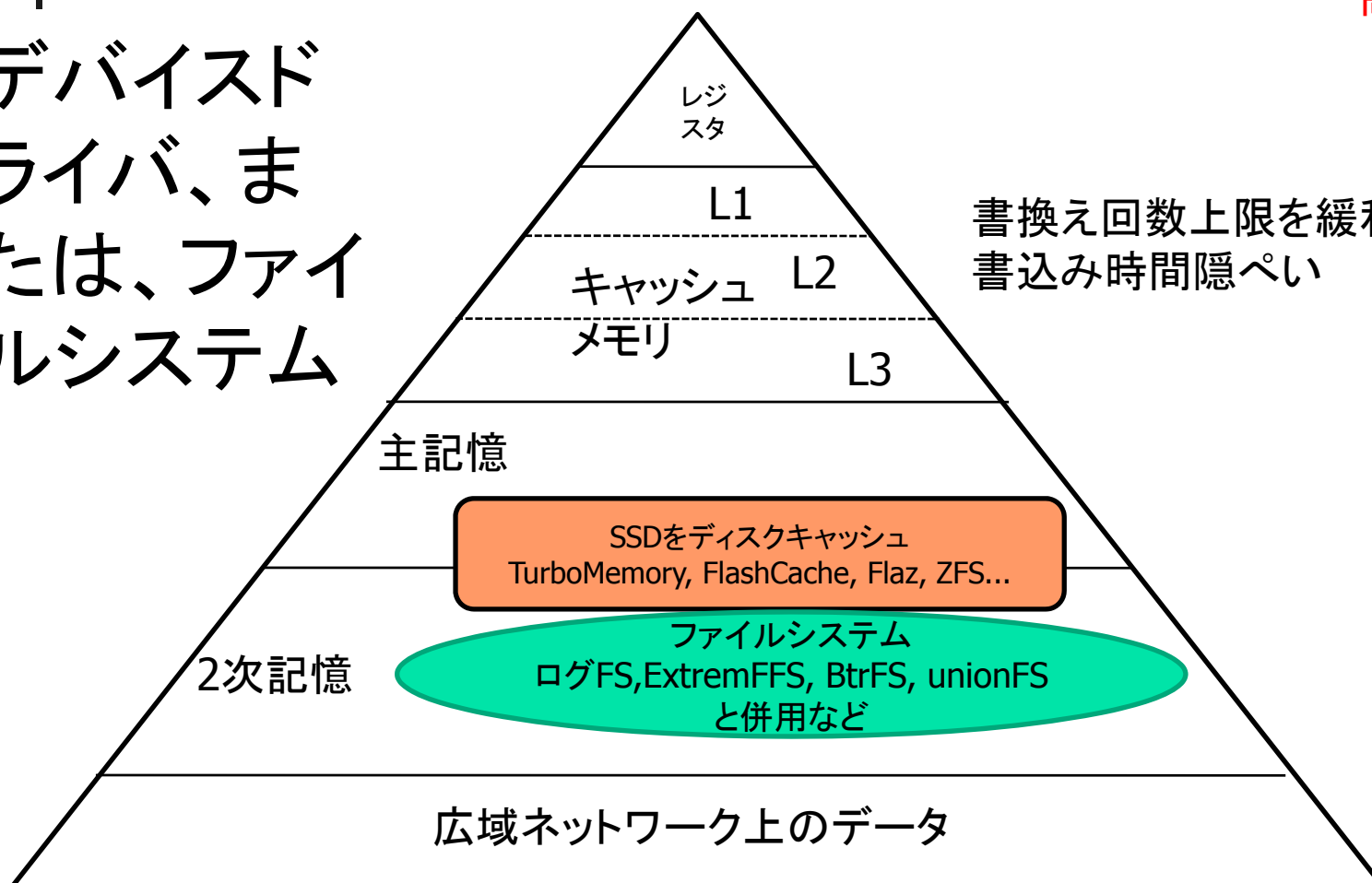
高速小容量



低速大容量

手っとり早く、かつ、実用化されたのはSSD

デバイスドライバ、または、ファイルシステム



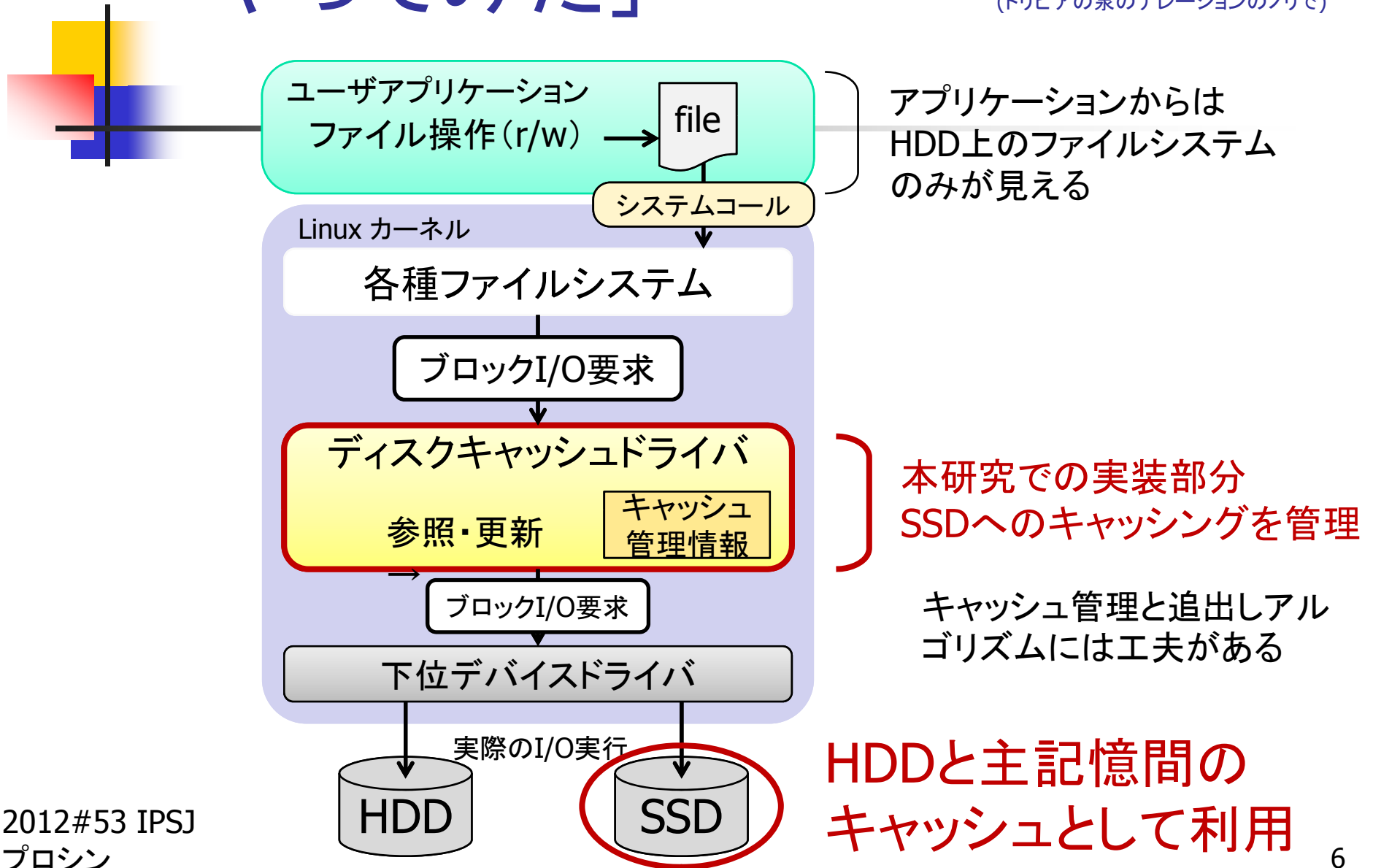
高速小容量



低速大容量

「やってみた」

(トリビアの泉のナレーションのノリで)





ベンチマークによる評価

- 目的
 - サーバーにおけるファイル入出力処理をエミュレーションするベンチマークを実行し、実行性能と消費電力を測定する
- 評価方法
 - ext3ファイルシステムにフォーマットしたストレージに対し、Filebenchを用い、次の4種類のワークロードを実行
 - fileserver - ファイルサーバ(100スレッド, R/W比1:2)
 - varmail - メールサーバ(100スレッド, R/W比1:1)
 - webserver - ウェブサーバ(100スレッド, R/W比10:1)
 - oltp - オンラインランザクシオン処理を行うDBサーバ(R/W比20:1)
 - 各ベンチマークを10分間実行し、オペレーションの処理数とI/Oの時間分布、全体の平均消費電力を測定



性能評価: オペレーション数

1秒当たりの平均オペレーション処理数 (ops/s)

	HDD	SSD	本方式
fileserver	457	9847	8695
webserver	26326	25456	25861
varmail	893	24298	19610
oltp	294	19163	4198

write オペレーションの割合が多いワークロードほど、
高い性能

電力評価(エネルギー効率)

- 平均処理速度 (ops/s) / ストレージ全体の平均消費電力 (J/s)
= 1J当たりのオペレーション処理量 (ops/J)
→ この値が大きいほどエネルギー効率が良い

各ストレージのエネルギー効率 (ops/J)

	HDD	SSD	本方式
fileserver	84	8871	1432
webserver	5161	28284	4474
varmail	171	23821	3127
oltp	56	15088	709

←悪くなってる

SSD とHDDの消費電力を足しても、ほとんどの場合でHDDよりもエネルギー効率が良い。ただし、遊んでいる時間が多いと悪くなる
↑ HDDのspin downはこれから

よみがえる↓オーバヘッド

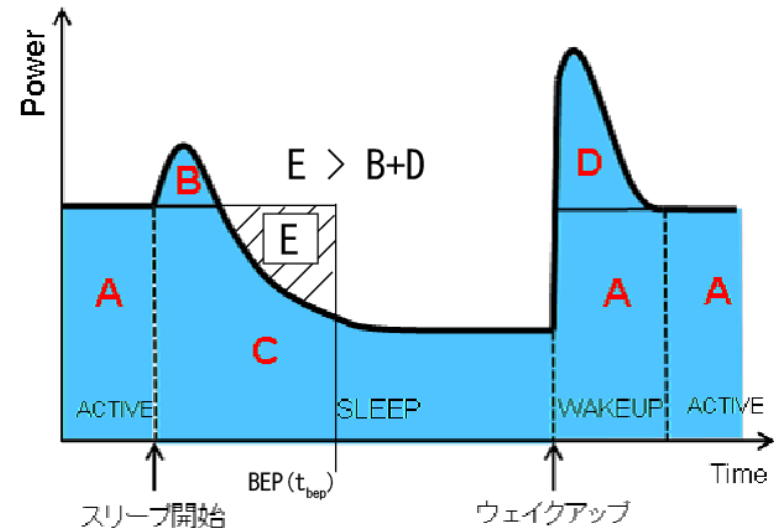
細粒度PGの課題(1)

フリーランチ
はない

(1) 適切なスリープポリシーの設定

- スリープ時の電力モデル
 - オーバヘッド(B,D)よりも削減電力(E)が大きくなる
スリープ期間(損益分岐点)をBEP(Break Even Point)と定義
 - 短い期間でON/OFF繰返すと電力増大→PGせず、
アクティブの方が良い
 - BEPを下回るスリープ(BEPミス)が頻発すると電力ロスにつながる

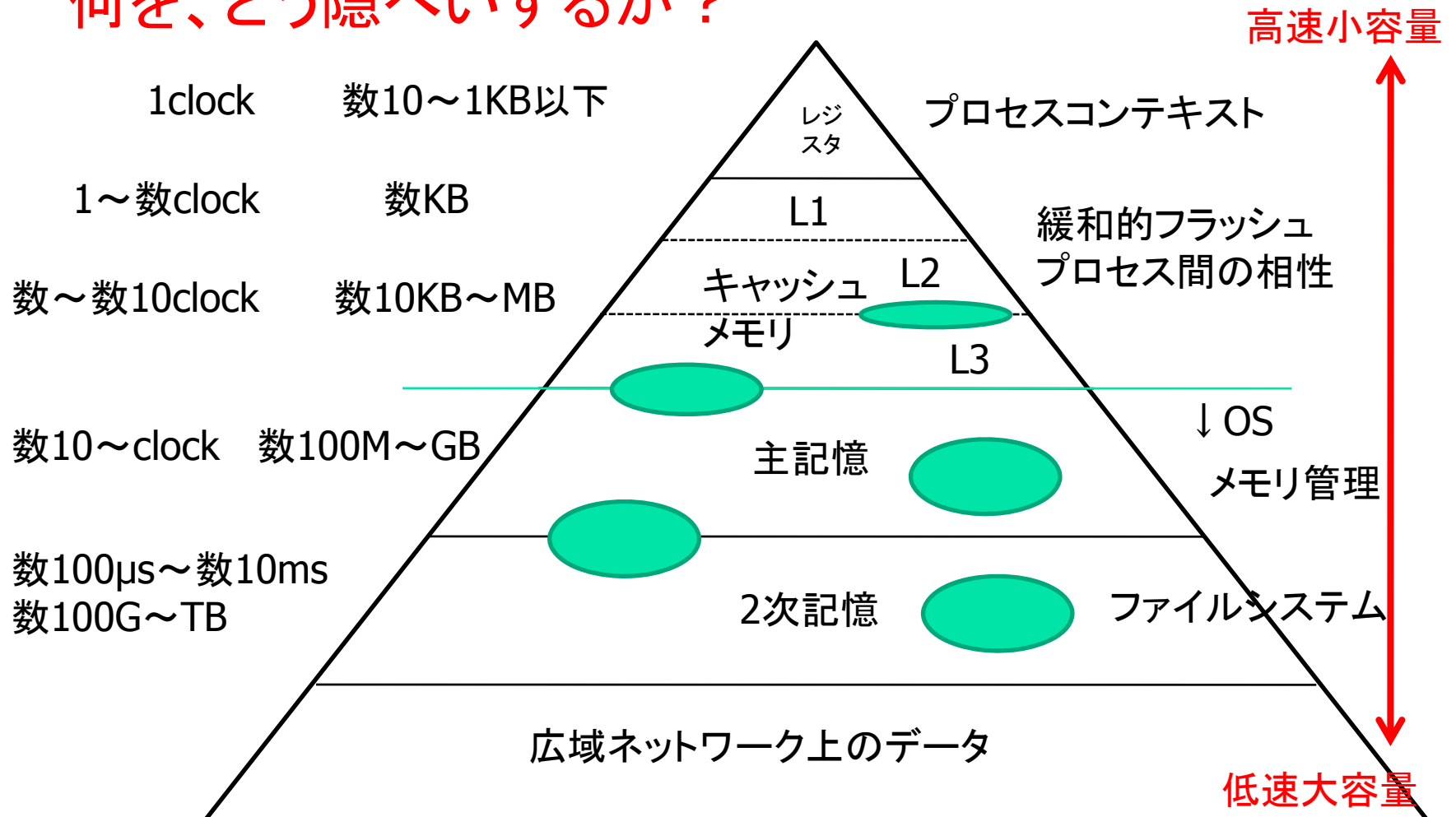
→適切なスリープポリシー、
コードの利用が必要



長いスパンでの観測、アクセス
パターンなどの導入による局所化
が考えられるが...

OSから見ると研究テーマはそれなりにある どこに、何を、どう入れるか？

何を、どう隠ぺいするか？





メモリ管理

- 利用頻度の高いデータと機械語で不揮発記憶素子を利用($R \gg W$)
- 書込み遅延はページングで隠ぺい、とか
(さらに、そこにキャッシュを絡めるとか)
- SPMとの併用
- 省電力キャッシュへの割当て制御
などはすぐに思いつく



ファイルシステム

- SSD、さらにはMRAM/FeRAM/PRAM向けのファイルシステムやデバイスドライバ
- (携帯端末の)組込み機器では必須(データ管理)
- いずれサーバ系もSSD主体になる？
- DB、Webサーバとの連携
- OS研究としては、one level storeの復権も視野に入る(object管理含めて)



アーキテクチャとしては

- SATA I/Fなどではなく、せめて主記憶の一種として使ってみたい
- 内部には、Cache、TLBが考える
隠ぺいはハードが行うが、資源割当てはOS
ネタだろう

←プロセススケジューリング、メモリ割当てで

システム全体で時空間局所化こそが必要不可欠



ただし...

- 新しいデバイスは入手が難しいし
- デバイスの基本データは自分たちでは無理
- オーバヘッド: on/off、書込み、隠ぺい
- デバイス特性
書込み時間、書込み上限、信頼性
- 挙動: ユーザ、プログラムなど
「局所化」のために適切なユースケース、実行
時情報を活用