

---

## GELATO@UNSW VM WORK

Slide 1



Ian Wienand, Paul Davies, Lucy Chubb

This work supported by UNSW and HP through the Gelato Federation

---

---

## ABOUT THE TALK

- Work in progress talk
- Please talk to us!

Slide 2

Topics:

- Superpages
  - Long Format VHPT
  - Page Table Abstraction
- 

---

## SUPERPAGES

What?:

- Use page size > base page size
- HugeTLB
  - ✗ Programmer Visible
  - ✗ Confusing
  - ✓ Suits legacy architectures (x86/64, Power)

Slide 3

Why?:

- More coverage from TLB
  - Transparency is a worthwhile, unsolved problem
  - Itanium provides wide range of interesting MMU options
    - **With LVHPT!**
- 
- 

## OUR APPROACH

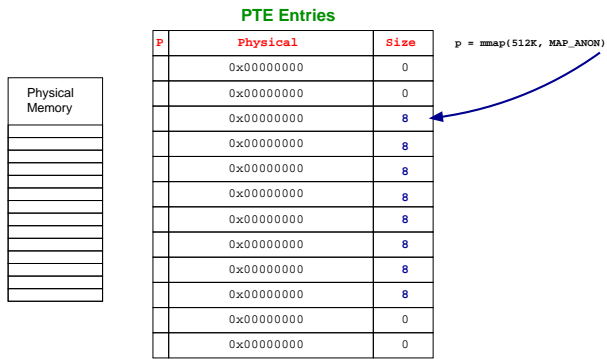
Cluster individual pages:

Slide 4

- ① Tag entries with size when created - `mmap(MAP_ANONYMOUS)`
  - ② On fault, check page size flag
  - ③ Backtrack and **allocate physical superpage**
  - ④ Point each sub-page of superpage to actual physical page
-

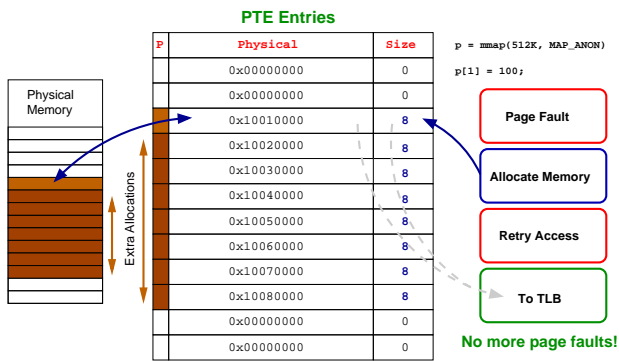
Slide 5

### SUPERPAGE OVERVIEW



Slide 6

### SUPERPAGE OVERVIEW



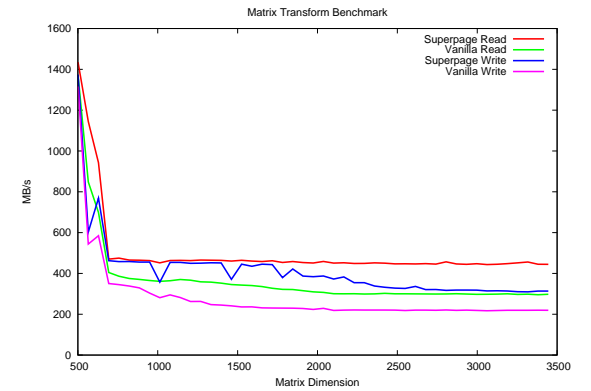
### ISSUES

Slide 7

- Work with Itanium features
- Page Tables
  - Double Size PTE
  - Mapping across PMD's
  - Better approaches
- Fragmentation
  - External Fragmentation (Ozlabs, Rice)
  - Internal Fragmentation (Rice)
- Promotion/Demotion
  - Competitive Algorithms

### MATRIX TRANSFORMATION

Slide 8



### DODGY RESULTS

Slide 9

- lmbench - no massive differences
- SPECint - mcf shows good improvement
- Lee's benchmark?

### VIRTUALLY HASHED PAGE TABLE

Slide 10

Problem:

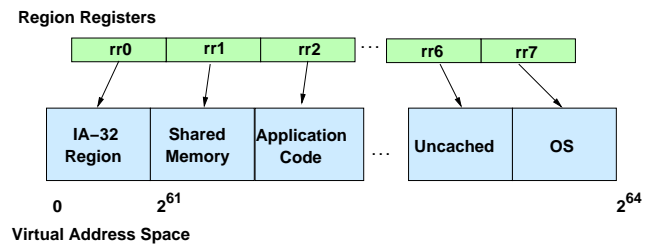
- TLB is **fast** but **small**
- Walking page tables via software is **slow**

Solution:

- Keep a "cache" of translations in RAM
- Let the hardware do the walking!
- The **VHPT**

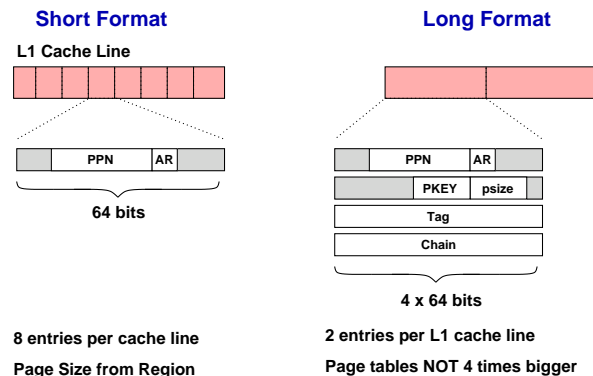
### FAST VM OVERVIEW - REGIONS

Slide 11

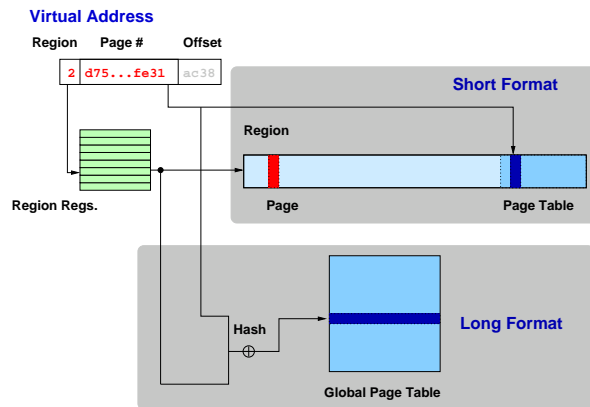


### TWO TYPES OF VHPT

Slide 12



## VHPT DIFFERENCES



Slide 13

## TWO TYPES OF VHPT

- **Short** : Cache Friendly, TLB Unfriendly
  - More entries in a cache line
  - Extra TLB entry per page of PTEs
- **Long** : Cache Unfriendly, TLB Friendly
  - Less entries per cache line
  - Global page table pinned with single entry

Slide 14

Why long?:

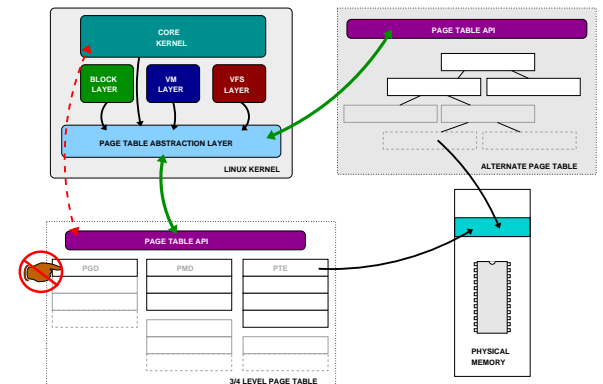
- ✓ Entries can hold page size, protection keys
- ✗ Entries are twice as big.
- Would like to talk to hackers and users.

## PAGE TABLE ABSTRACTION

Slide 15

- Problems with current implementations:
- 3 (4? 5? X?) level page table sub-optimal
  - Main architectures have differing needs

## SOLUTION? PAGE TABLE ABSTRACTION



Slide 16

---

Issues:

Slide 17

- ✓ Performance - *not bad!*
- ✓ Hardware walked page tables - *we have ideas*
- ✓ Alternate Implementations - *GPT*
- ✗ Convincing Linux community to rip apart VM - *hard*

---

---

WHERE DO YOU GET IT?

<http://www.gelato.unsw.edu.au/IA64Wiki>

or

Google for "ia64 wiki"

QUESTIONS?

---

Slide 18

QUESTIONS?