The CP-PACS Machine Plan for 2005-2007

Akira Ukawa
Center for Computational Sciences
University of Tsukuba

- *Introduction*
- *CP-PACS successor*
- *Theoretical Benchmark*
- *Schedule*

Lattice QCD in (Tsukuba) Japan

1980 1985 1990 1995 2000 2005

S810
315Mflops

S820
1Gflops

VPP500/80
128Gflops

SR8000F1/100
1.2Tflops

??
>20Tflops

KEK

JLQCD Collaboration

Supercomputer installation since 1985
Services Japanese LQCD community
Upgrade every 5-6 years

QCDPAX
14Gflops

CP-PACS
614Gflops

post
CP-PACS
10-20Tflops

Univ. of Tsukuba

CP-PACS Collaboration

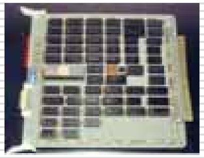
Development of dedicated
computers
for scientific applications

Center for
Computational Physics
1992

Center for
Computational Sciences
2004

25 years of R&D of parallel computers at Tsukuba

PACS-9

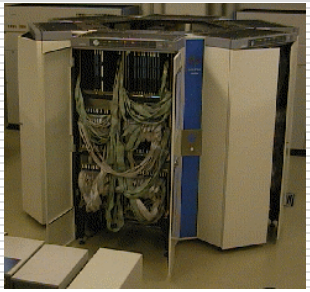


1978

AXS-32

1980

QCDPAX



1989

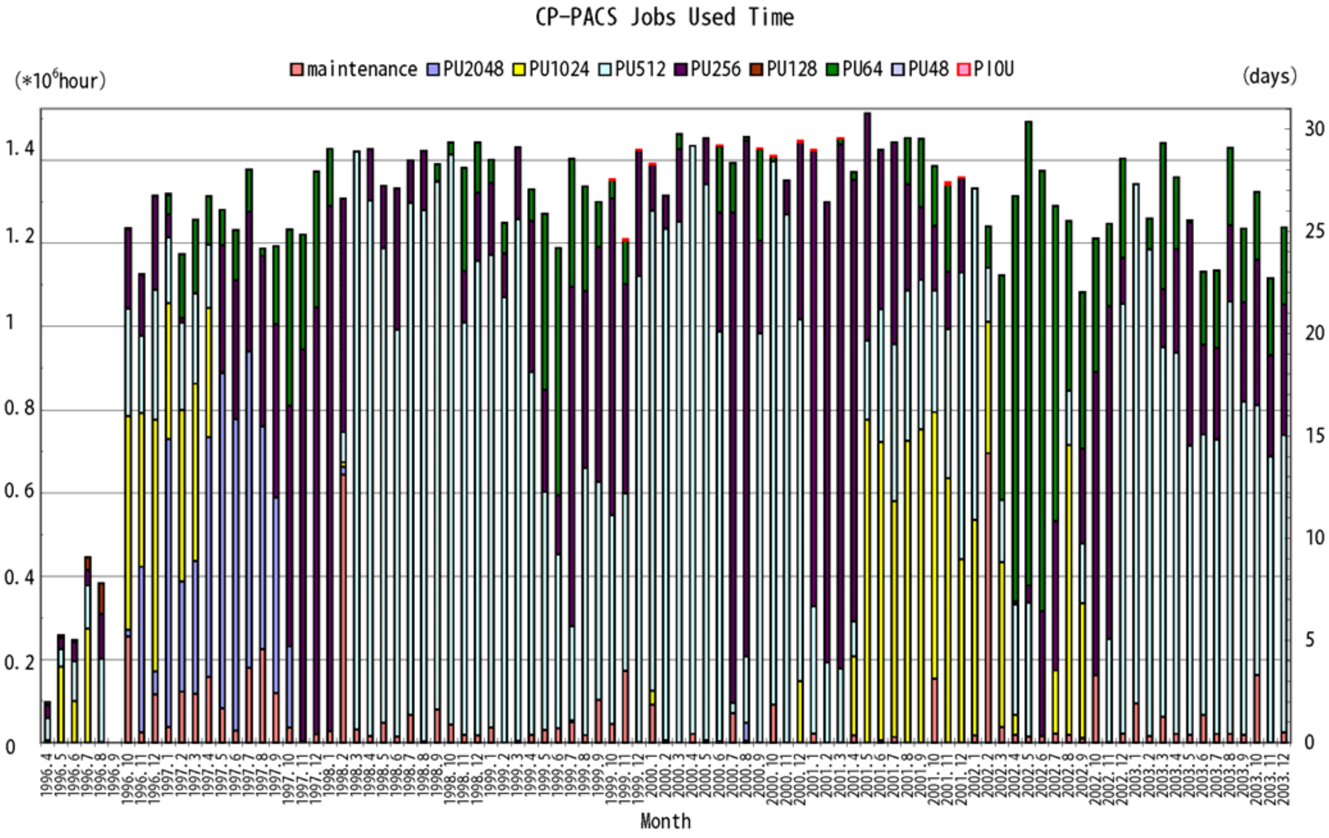
CP-PACS

1996

year	name	speed	funding	金額
1978	PACS-9	7kflops	annual research budget	1.5MYen
1980	PAXS-32	500kflops	annual research budget	1.5MYen
1983	PAX-128	4Mflops	annual research budget	6MYen
1984	PAX-32J	3Mflops	annual research budget + special allocation by the university	30MYen
1989	QCDPAX	14Gflops	Grants-in-aid for scientific research	0.3BYen(3 years)
1996	CP-PACS	614Gflops	Special grants-in-aid	2.2BYen (5 years)

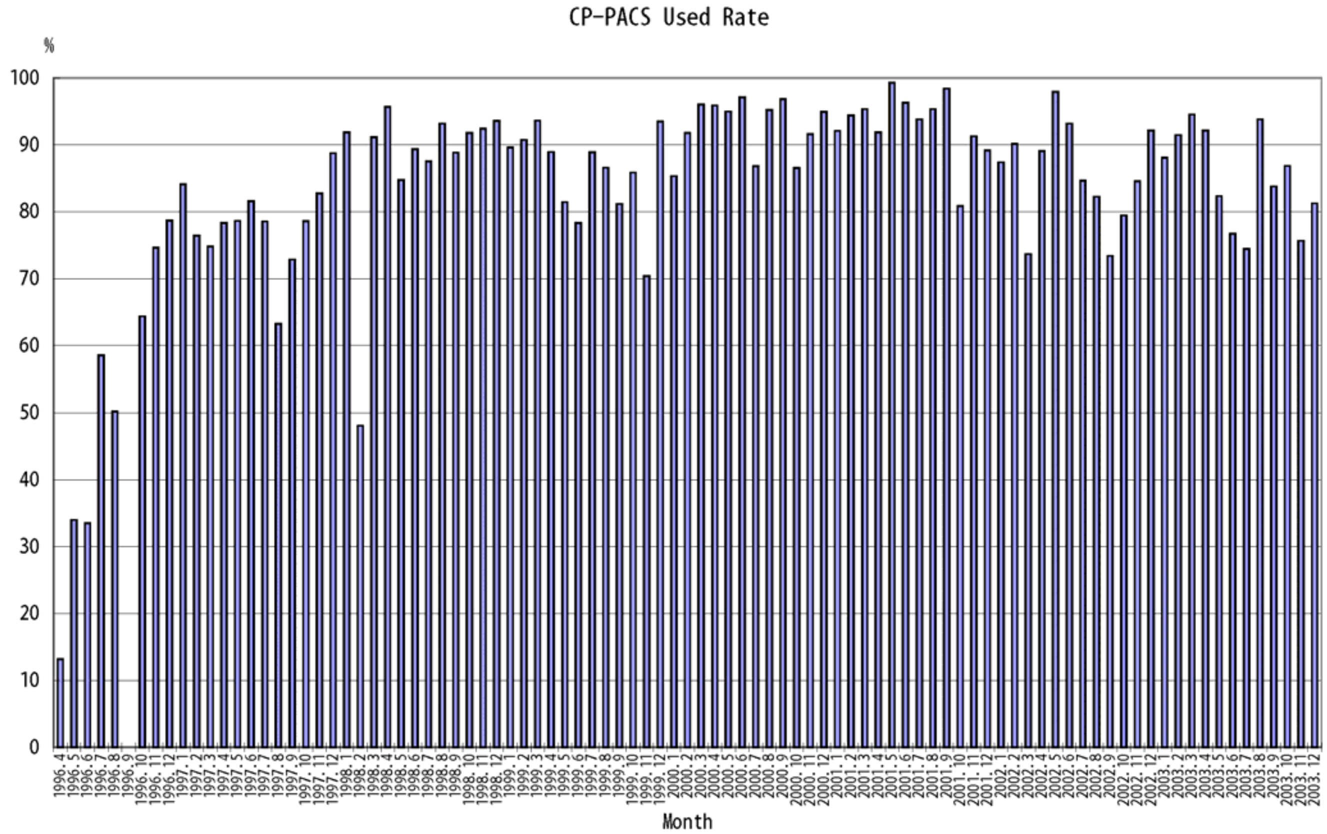
CP-PACS run statistics 1992.4 – 2003.12

Monthly run hours according to partitions

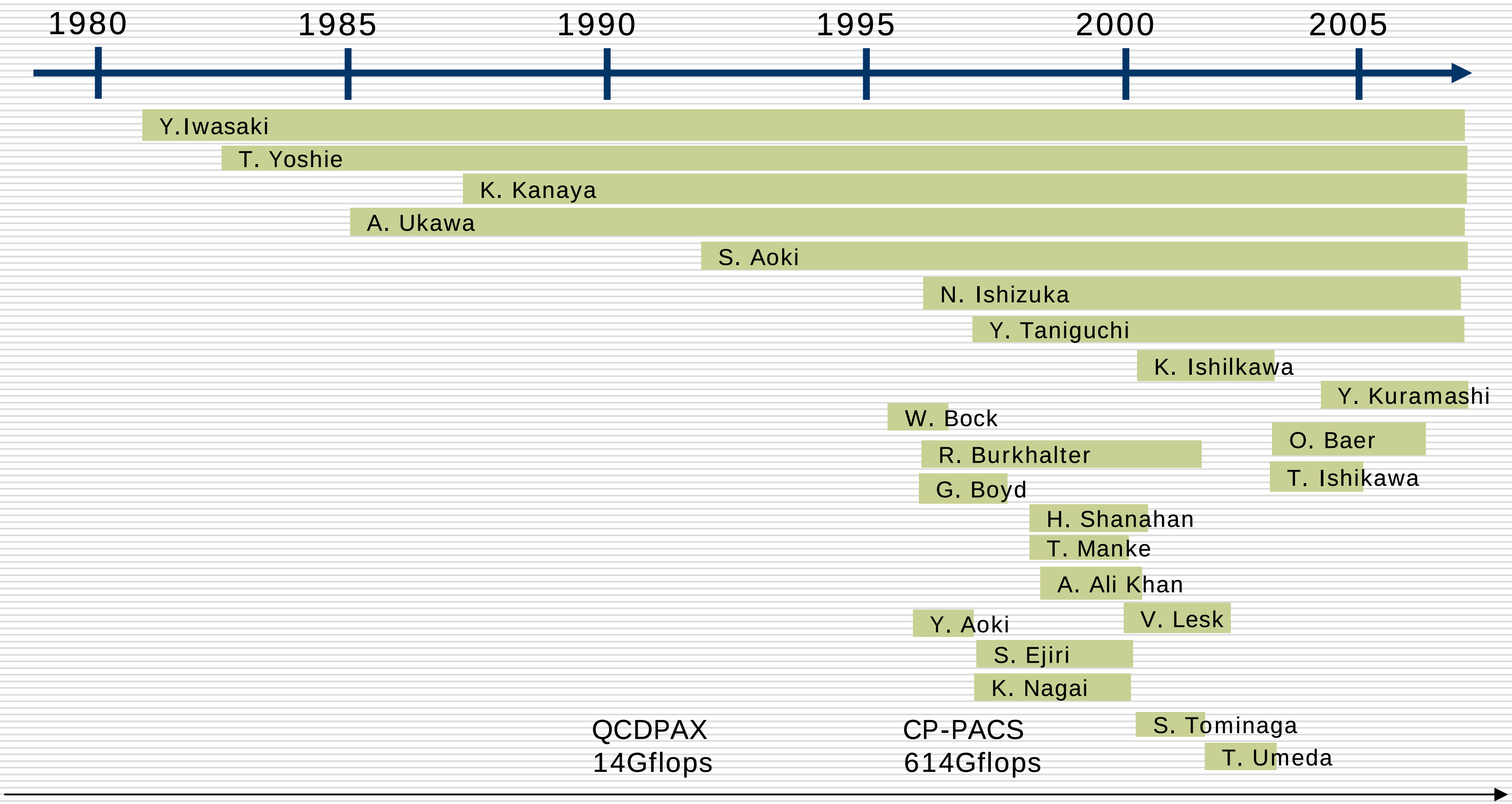


CP-PACS run statistics 1992.4 – 2003.12

Monthly usage rate = (hour used for jobs)/(physical hours of month)



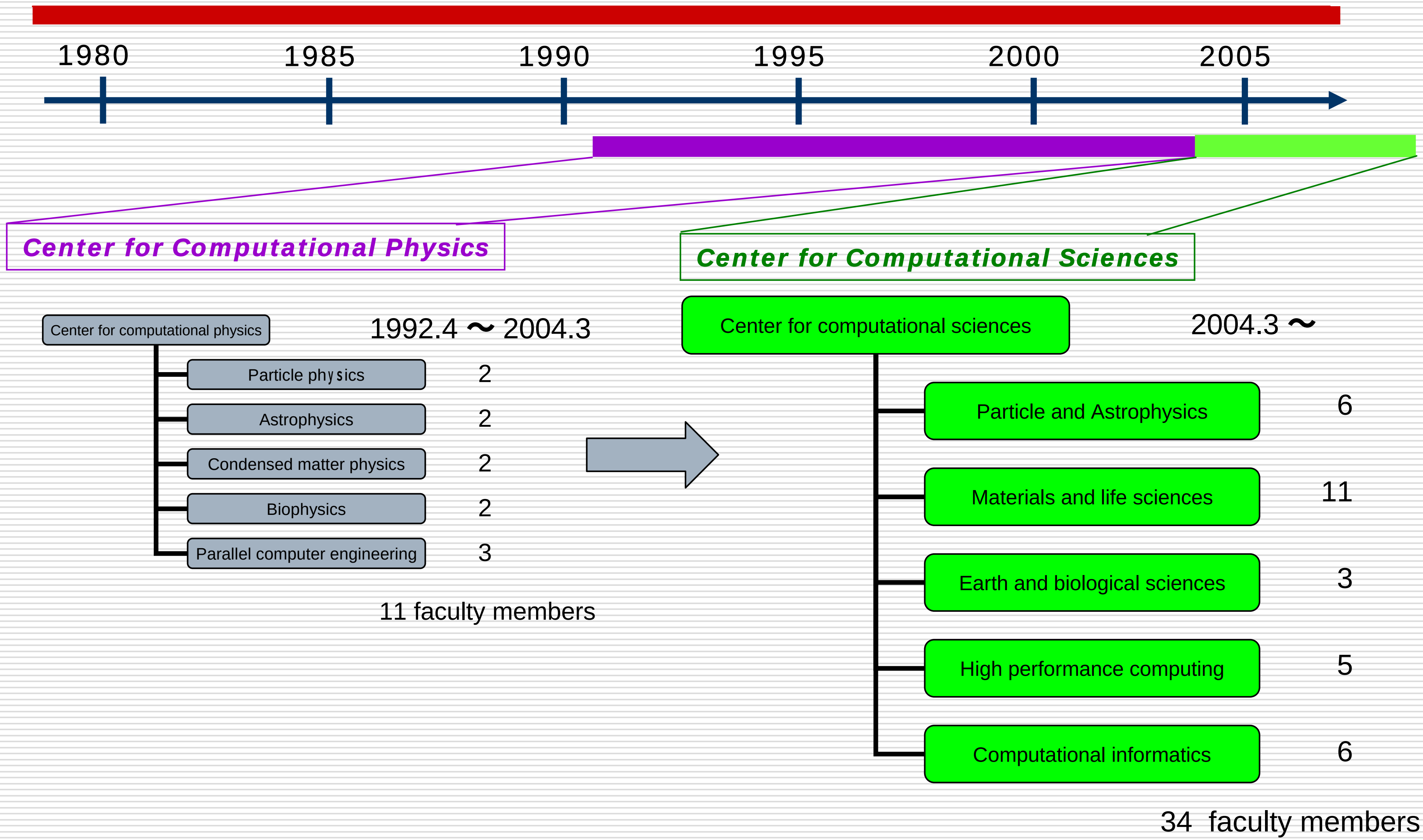
People at University of Tsukuba



Univ. of Tsukuba

CP-PACS Collaboration

Organization



Machine perspective in Japan

- KEK
  - Upgrade of SR8000F1 in March 2006
  - Government supercomputer procurement in progress
  - >20Tflops/10TB peak performance targeted
- Center for Computational Sciences, U. of Tsukuba
  - Successor to CP-PACS planned
  - Funding requested for JFY2005-2007
  - MPP in terms of commodity components
  - At least 12Tflops/4TB system by March 2006; further installation in later years

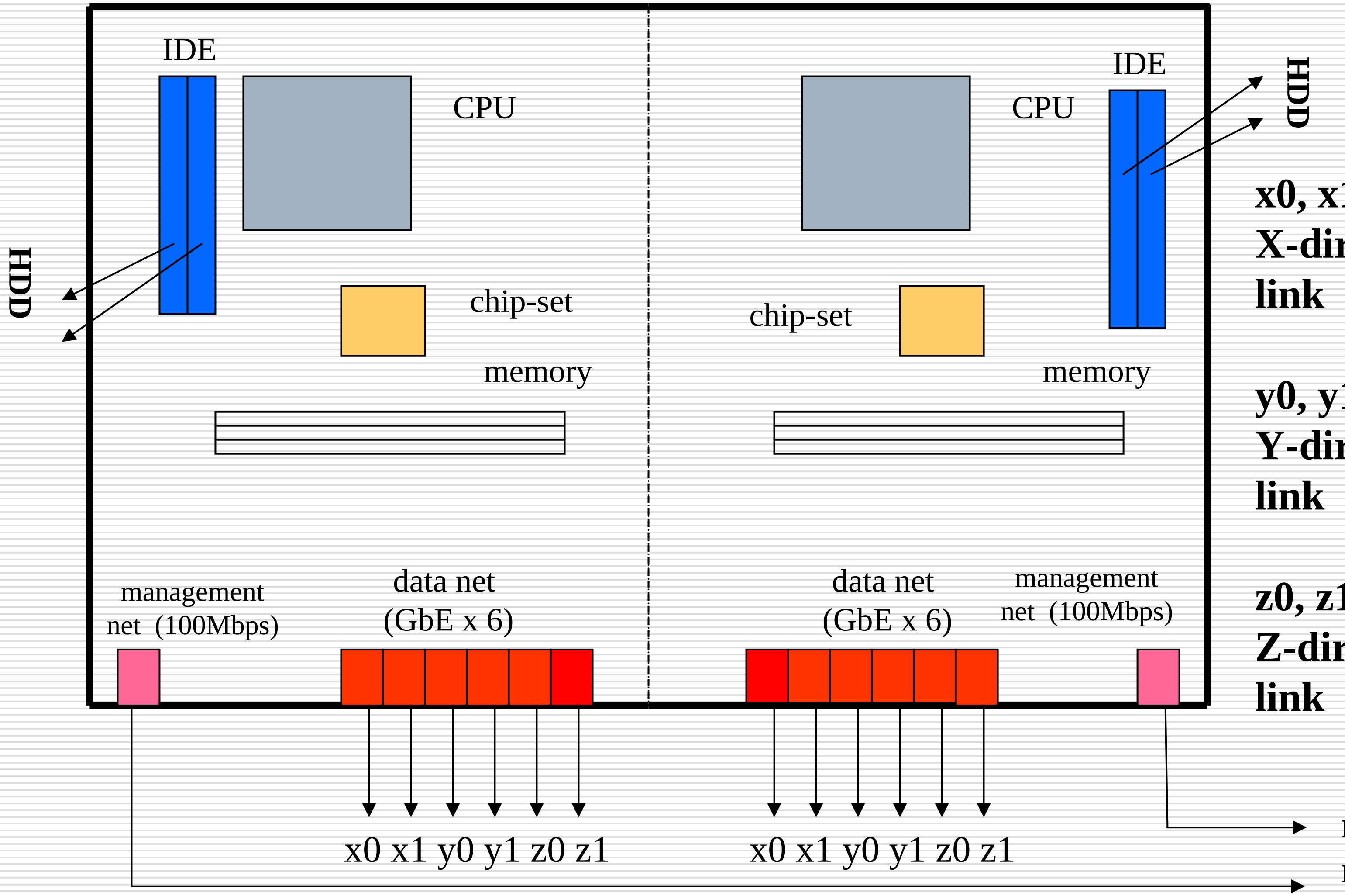
CP-PACS successor outline

- Strategy:MPP in terms of commodity components
  - Node
    - Single CPU (Xeon 3GHz)
    - 2GB PC3200 memory with FSB800

      i.e., 6.4GB/s peak memory bandwidth to each CPU
    - 200GB disk (Raid0 mirror)
  - Network
    - 3-dimensional hyper-crossbar, i.e., crossbar switch in each direction
    - Dual GbEthernet for each direction,

      i.e., 025GB/s/link and an agregate 0.75GB/s/node
  - System size
    - At least 2048 CPU (16x16x8, 12Tflops/4TB), and hopefuly up to 3072 CPU (16x16x12, 18Tflops/6TB)

Board layout

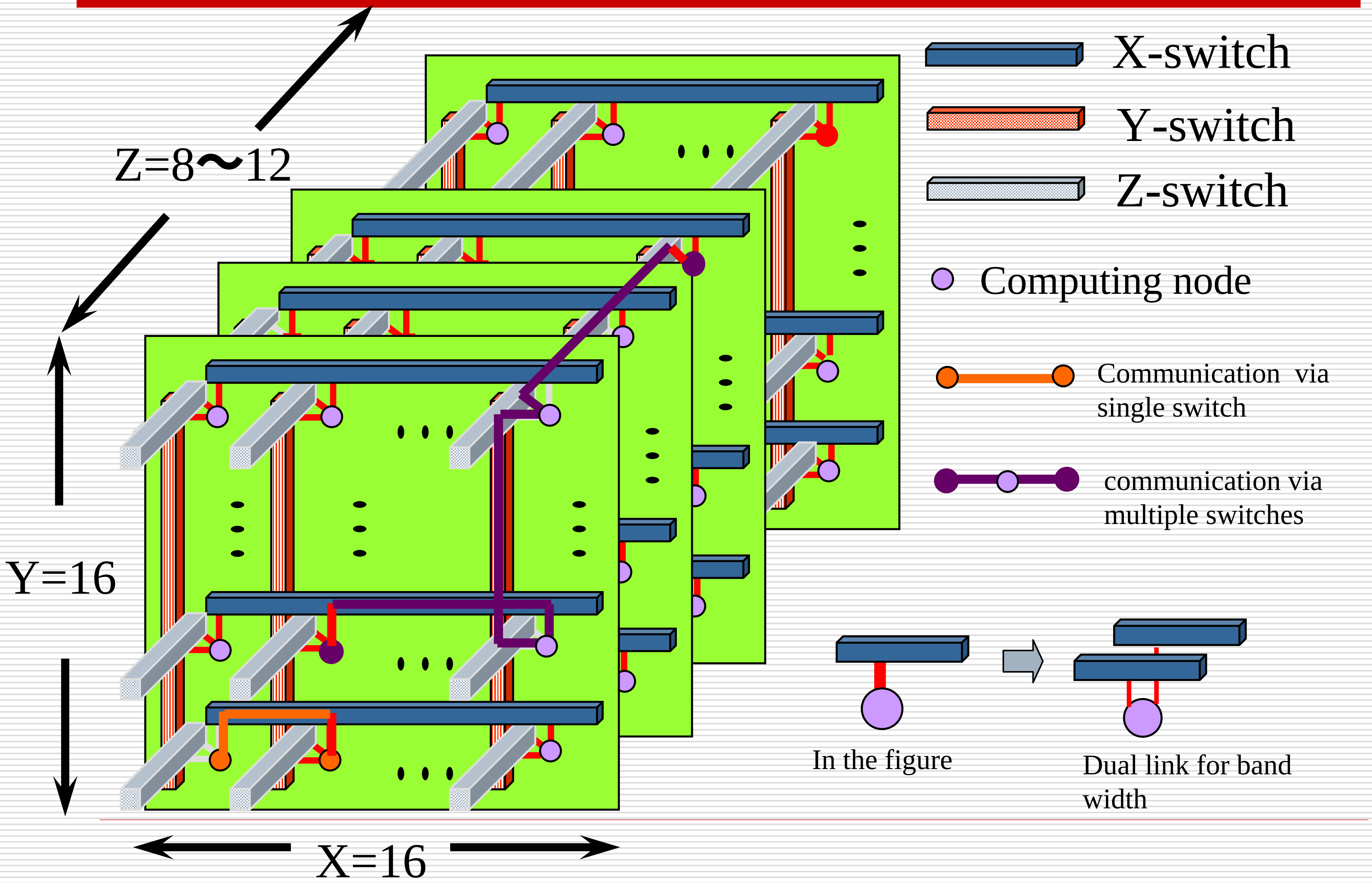


x0, x1: X-direction dual link

y0, y1: Y-direction dual link

z0, z1: Z-direction dual link

3-dimensional hypercrossbar network



Rack layout and network connections

Connections for CPU([0-F],[0-F],[0-1]) (1/6th =512 CPU) are shown

Y1-Y2,Z CPU rack(x=0〜F,y=Y1〜Y2,z=Z)

X-connection (1 line =128cables)

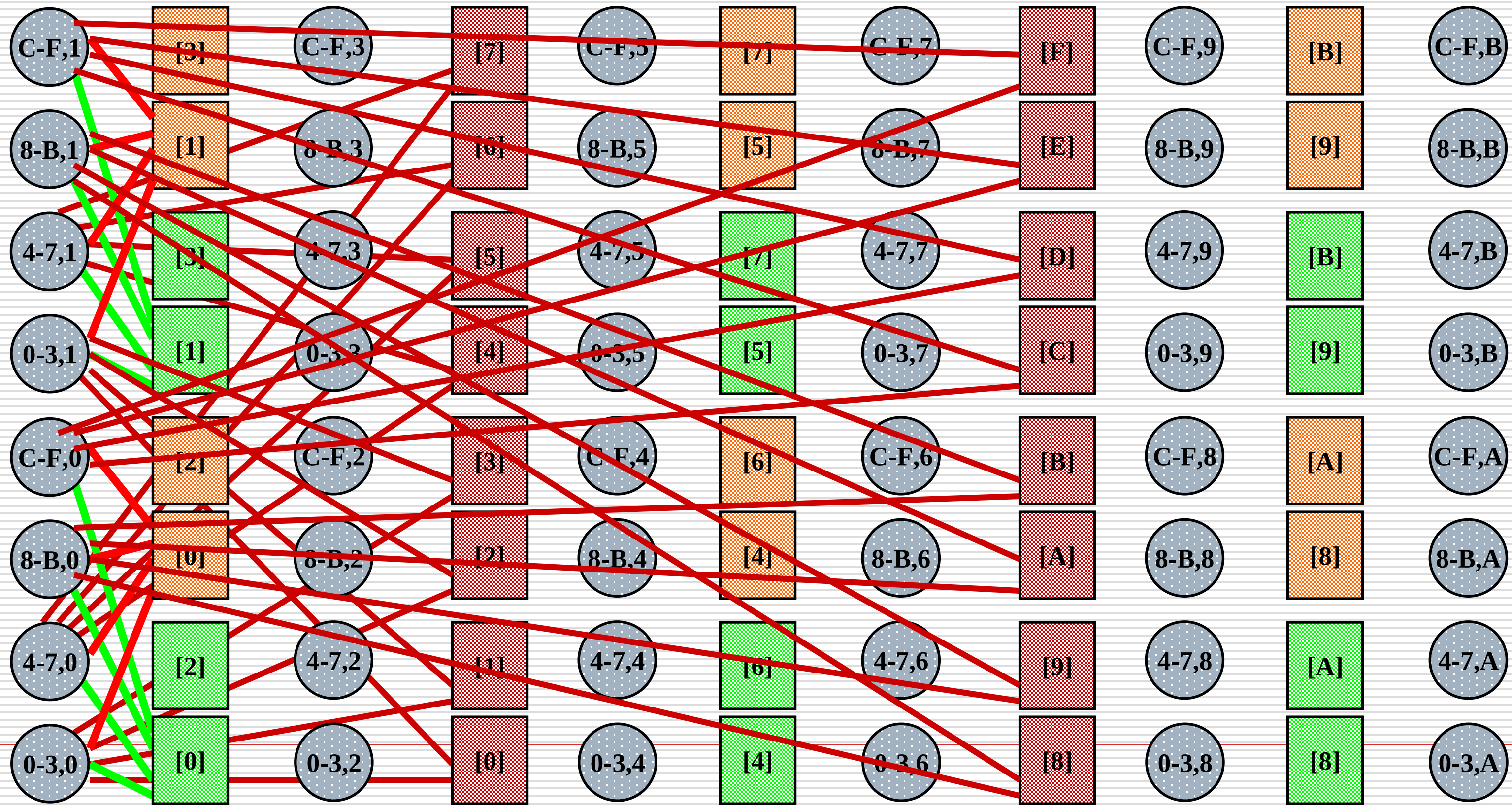
Y-connection (1 line = 128cables)

Z-connection (1 line = 32 cables)

[Z] X-switch rack(y=0〜15, z=Z)

[Z] Y-switch rack(x=0〜15, z=Z)

[Y] Z-switch rack(x=0〜15, y=Y)



Physics possibilities on the new machines

- Wilson-clover Nf=2+1 simulation
  - Runs with existing machines
    - Three lattice spacings $a^2 \approx 0.015\, fm^2$, $0.01\, fm^2$, $0.005\, fm^2$
    - But only down to $\frac{m_\pi}{m_\rho} \approx 0.6 \quad i.e., \quad \frac{m_{ud}}{m_s} \approx 0.5$
  - *Wish to go down to* $\frac{m_\pi}{m_\rho} \approx 0.4 \quad i.e., \quad \frac{m_{ud}}{m_s} \approx 0.2$ *or less…*
- Domain wall/overlap
  - Under discussion
- Staggered
  - …………

Performance Benchmark estimate (I)

- Wilson-clover Nf=2+1 PHMC code
- Parameters

pi/rho	mq(GeV)
0.7	0.079
0.6	0.044
0.5	0.0255
0.4	0.0145
0.3	0.0075
0.2	0.0025

*

lattice size	$N_s^3 \times N_t$
BiCGStab inversions	$N_{inv} = 31 + \frac{10.7}{m_q a}$ *
polynomial order	N_{poly}
HMC step size	$d\tau = \left(0.223 m_q(GeV) - 0.620 m_q(GeV)^2\right) \times \frac{24}{N_s}$ *
node array	$n_x \times n_y \times n_z$

*Nf=2 empirical fit ; need revision for Nf=2+1

- Flop/node/trajectory

$$\#\,flops = \left(130938 + 8808 N_{inv} + 3666 N_{poly}\right) \times \frac{N_s^3 \times N_t}{n_x n_y n_z} \frac{1}{d\tau}$$

- Data sent/node/trajectory

$$\#\,data\ sent = \left(10 + 8N_{inv} + \frac{7}{2} N_{poly}\right) \times 192 \times \left(\frac{N_t}{2} + 1\right) \times 2 \times \left(\frac{N_s}{n_x}\frac{N_s}{n_y} + perm\right) \frac{1}{d\tau} \quad Byte$$

Performance Benchmark estimate (II)

- Target runs

lattice size	$24^3 \times 48$ at $a \approx 0.1\,fm$, $32^3 \times 64$ at $a \approx \sqrt{\frac{1}{2}}\,0.1\,fm$
#trajectories	10000
polynomial order	300

- Machine assumptions

system size	$2048\,CPU$
job partition	$8^3 = 512\,CPU \times 4$
CPU performance	$2\,Gflops$
network performance	$0.2\,GB/s/link$ (3 *directions overlapped*)
network latency	$20\,\mu s$

Performance Benchmark estimate (III)

				Standard HMC					
1/a	lattice size		pi/rho	Ninv	1/dt	time/ traj (hr)			10000traj
(GeV)	N s	N t				calc	comm	total	(days)
2	24x48		**0.6**	517	116	0.12	0.13	0.25	**26**
			0.5	870	189	0.30	0.32	0.62	**65**
			0.4	1507	322	0.84	0.89	1.73	**180**
			0.3	2884	611	2.93	3.11	6.03	**629**
			0.2	8591	1806	25.00	26.57	51.57	**5372**
2.83	32x64		**0.6**	719	155	0.67	0.46	1.13	**118**
			0.5	1218	252	1.72	1.19	2.91	**303**
			0.4	2118	430	4.86	3.39	8.25	**860**
			0.3	4066	814	17.16	11.99	29.14	**3036**
			0.2	12143	2408	148.23	103.66	251.89	**26238**

Rather dismal number of days .....

Acceleration possibilities (I)

- Hasenbusch acceleration — M. Hasenbusch Hep-lat/0107019

$$\det M^{+}M = \int d\phi_1^{+} d\phi_1 \exp\left(-\phi_1^{+}\frac{1}{W^{+}W}\phi_1\right)\cdot\int d\phi_2^{+} d\phi_2 \exp\left(-\phi_2^{+}\frac{1}{\tilde{M}^{+}\tilde{M}}\phi_1\right)$$

$$M = 1-\kappa^2 M_{oe}M_{eo}$$

$$W = M + \rho$$

$$\tilde{M} = W^{-1}M$$

  - Factor 2 larger step size without fine tuning of ρ
  - Implementation in progress

- Domain-decomposition techniques — M. Luescher Hep-lat/0409106

Acceleration possibilities (II)

- Domain decomposition acceleration

$$\det D^{+}D=\int d\phi_{\Omega}^{+}d\phi_{\Omega}\exp\left(-\phi_{\Omega}^{+}\frac{1}{D_{\Omega}{}^{+}D_{\Omega}}\phi_{\Omega}\right)\cdot\int d\phi_{\Omega*}^{+}d\phi_{\Omega*}\exp\left(-\phi_{\Omega*}^{+}\frac{1}{D_{\Omega*}{}^{+}D_{\Omega*}}\phi_{\Omega*}\right)\cdot\int d\phi_{R}^{+}d\phi_{R}\exp\left(-\phi_{R}^{+}\frac{1}{R^{+}R}\phi_{R}\right)$$

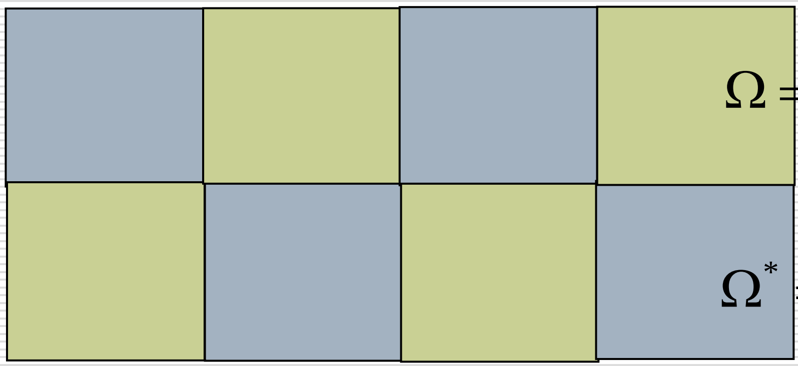
$$D_{\Omega}=\sum_{n}D_{\Lambda_n}\quad,\quad D_{\Omega*}=\sum_{n}D_{\Lambda_n^*}$$

Dirichlet b.c. on $\partial\Omega=\sum_n\partial\Lambda_n\,,\,\partial\Omega^{*}=\sum_n\partial\Lambda_n^{*}$

$$R=1-\theta_{\partial\Omega*}D_{\Omega}^{-1}D_{\partial\Omega}D_{\Omega*}^{-1}D_{\partial\Omega*}$$

Values only on the boundary of domains

$$R^{-1}=1-\theta_{\partial\Omega*}D^{-1}D_{\partial\Omega*}$$



$$\Omega=\prod_n\Lambda_n$$

$$\Omega^{*}=\prod_n\Lambda_n^{*}$$

Acceleration possibilities (III)

- Crucial observation(Luescher)

F_0 force due to gauge action

F_1 force due to D_Ω and D_{Ω^*}

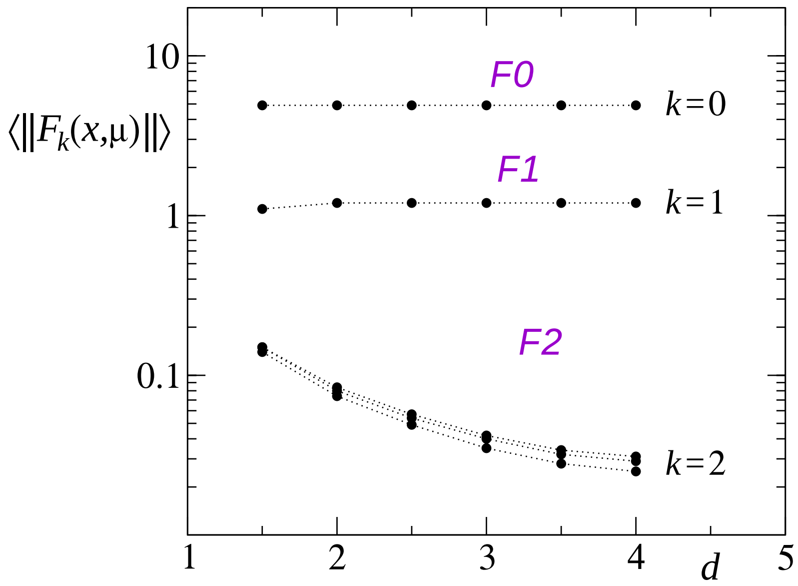
F_2 force due to R

Numerically at $\frac{m_\pi}{m_\rho} \approx 0.7 - 0.4$ and $a^{-1} \approx 2.4\mathrm{GeV}$

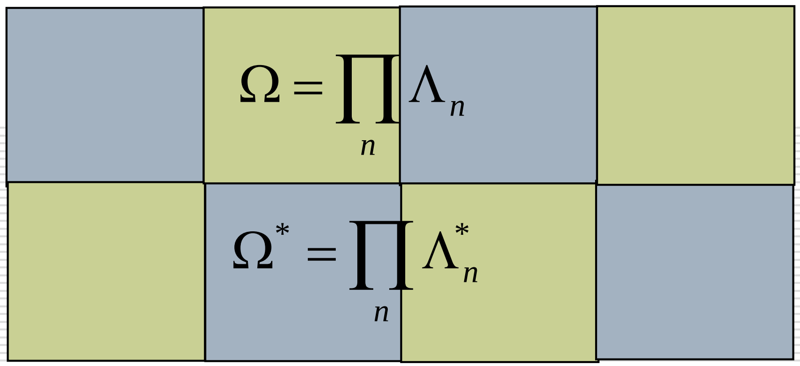
$$F_0 : F_1 : F_2 \approx 5:1:0.2 = 25:5:1$$

Use larger step sizes for quarks

$$d\tau_0 : d\tau_1 : d\tau_2 \approx 1:5:25$$



From M. Luescher Hep-lat/0409106



Advantages with domain-decomposed HMC

- $D_{\Lambda_n} \to D_{\Lambda_n}^{-1} \qquad D_{\Lambda_n^*} \to D_{\Lambda_n^*}^{-1}$
  - Inversion in each domain once every 5 MD steps or so
  - Dirichlet boundary condition, hence
    - easier to invert
    - No inter-node communication if the domain is within the node
- $R \to R^{-1} = 1 - \theta_{\partial\Omega^*} D^{-1} D_{\partial\Omega^*}$
  - Full inversion once every 25 MD steps or so
- *Both Floating and Communication Requirements are reduced......*

Acceleration possibilities (IV)

				standard HMC	domain-decomposed HMC							
1/a	lattice size		pi/rho	10000traj	#steps			time/traj(hr)			10000traj	acceleration
(GeV)	Ns	Nt		(days)	N0	N1	N2	calc	comm	total	(days)	
2	24x48		**0.6**	**26**	4	5	5	0.031	0.005	0.037	**4**	7
			0.5	**65**	4	5	6	0.058	0.010	0.068	**7**	9
			0.4	**180**	4	5	7	0.110	0.019	0.129	**13**	13
			0.3	**629**	4	5	8	0.230	0.041	0.271	**28**	22
			0.2	**5372**	4	5	9	0.747	0.13[illegible]	0.880	**92**	59
2.83	32x64		**0.6**	**118**	5	6	6	0.181	0.018	0.199	**21**	6
			0.5	**303**	5	6	7	0.333	0.033	0.366	**38**	8
			0.4	**860**	5	6	9	0.713	0.071	0.784	**82**	11
			0.3	**3036**	5	6	10	1.475	0.147	1.622	**169**	18
			0.2	**26238**	5	6	11	4.739	0.473	5.213	**543**	48

- Only a paper estimate, but more than encouraging .....
- Implementation in progress

Physics issues – standard targets (I)

- Light hadron spectrum within a few % accuracy
  - Nucleon mass
  - rho meson as a resonance
    - Finite-size techniques
    - Wilson chiral perturbation theory

- Light quark masses and strong coupling constant
  - Non-perturbative RG running and Z factors
    - Schroedinger functional methods
- Hadron physics
  - eta' meson and topology
  - exotics (glueballs, multi-quark states)

Physics issues – standard targets (II)

Interface with the electro-weak sector

- K meson system
  - BK
    - Non-perturbative Z factor
  - Kl3 form factor

German-Japan Workshop
S. Hashimoto

- Charm and bottom quark systems
  - Onium and heavy-light spectra
  - Decay constants and B parameters
  - weak form factors
    - Relativistic heavy quark action (Aoki-Kuramashi-Tominaga)

German-Japan Workshop
Y. Kuramashi

Physics issues – standard targets (III)

- Finite-temperature behavior
  - Physical Nf=2+1 Phase diagram
  - Critical temperature
  - Equation of state
  - Finite-temperature effects in the hadron spectra
    - Issue of parity-broken phase
- Finite-density behavior
  - Small chemical potential
  - Large chemical potential?

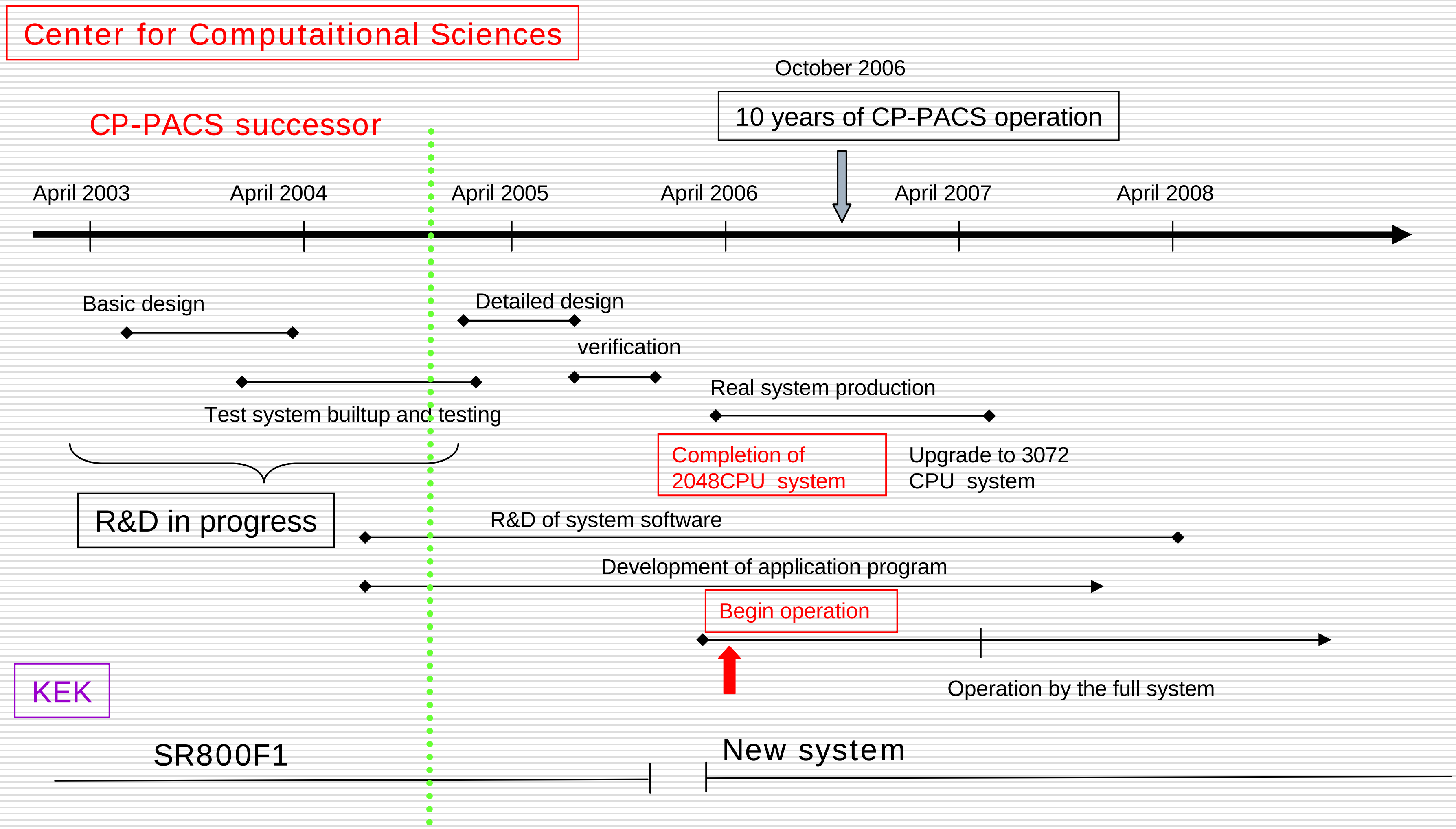
Physics issues – some challenges

- K->pipi decays and CP violation
  - Calcuation with two pions in the final state
    - Already some success in the I=2 channel
    - Much advance in the related pipi 2-body calculations
    - Finite-size theoretical setup by Lellouch-Luescher

- Hadron scattering and nulcear binding
  - Nucleon-nucleon potential
    - Dependence on quark mass is a very interesting issue
      e.g., deuteron may not bound if quarks are heavier….(?)
  - Kaon-nucleon potential
    - Relevant for penta-quark

- ……

Schedule


Center for Computaitional Sciences
October 2006
10 years of CP-PACS operation
CP-PACS successor
April 2003
April 2004
April 2005
April 2006
April 2007
April 2008
Basic design
Detailed design
verification
Real system production
Test system builtup and testing
Completion of 2048CPU system
Upgrade to 3072 CPU system
R&D in progress
R&D of system software
Development of application program
Begin operation
Operation by the full system
KEK
SR800F1
New system