



P.O. BOX 64 (Gustaf Hällströmin katu 2)
FIN-00014 UNIVERSITY OF HELSINKI
Tel: +(358-9) 191 50521
Fax: +(358-9) 191 50522

Fysiikan tutkimuslaitoksen tietotekniikkastrategia

2004-2008

1. Fysiikan tutkimuslaitos

Fysiikan tutkimuslaitos on valtakunnallista tehtävää hoitava Helsingin yliopiston, Jyväskylän yliopiston ja Helsingin teknillisen korkeakoulun yhteinen tutkimuslaitos. Laitos huolehtii Suomen yhteistyöstä Euroopan ydinfysiikan tutkimusjärjestön CERNin kanssa. Tätä tehtävää varten laitos saa Opetusministeriöltä erillismäärärahan. Fysiikan tutkimuslaitoksen tehtävänä on harjoittaa ja edistää fysiikan perustutkimusta ja soveltavaa tutkimusta, hiukkaskiihdyttimiin liittyvää soveltavaa teknologista kehitystyötä sekä osallistua alan tutkijakoulutukseen.

Fysiikan tutkimuslaitoksen tutkimustyön pääosa liittyy kokeelliseen suurenergiafysiikkaan, jonka tietojenkäsittelytarve on erittäin suuri. Fysiikan tutkimuslaitoksen teoreettisen fysiikan tutkimusohjelman projektit tarvitsevat myös suurteholaskentaa. Näin ollen suurteholaskennan kehittäminen ja tietotekniikan eri sovellukset kuuluvat Fysiikan tutkimuslaitoksen ydintoimintaan.

2. Kansainvälinen tietotekniikkastrategia

Koska Fysiikan tutkimuslaitoksen tietojenkäsittelytarve ylittää Suomen valtakunnallisten tietokoneressurssien tarjoaman laskentakyvyn, laitoksen tietkonestrategia perustuu kansainväliseen yhteistyöhön. Laitos osallistuu Pohjoismaiden yhteiseen sekä Euroopan unionin tietohilaverkoston kehittämiseen. Kehitystyö suoritetaan yhteistyössä Tieteellisen laskennan keskuksen (CSC) kanssa. Laitos osallistuu pohjoismaiseen Nordic Data Grid Facility -projektiin, ja edustaa Suomea CERNin Large Hadron Collider Computing Grid projektissa sekä Euroopan Unionin European Data Grid ja Enabling Grids for E-science in Europe (EGEE) hankkeissa. Laitos toimii näin ollen dataintensiivisen suurteholaskennan pilottilaitoksena Suomessa.

Fysiikan tutkimuslaitoksen kansainvälistä tietotekniikkatoimintaa varten laitos ylläpitää teknologiaohjelmaa, jonka päätavoitteena on tietohilan teknologian sovellusten kehittäminen.

3. Fysiikan tutkimuslaitoksen atk-voimavarat; nykytila

Laitoksen käytössä on 70 työasemaa (desktop). Lisäksi on 34 kannettavaa miktoa. Suurin osa koneista on yksikön omassa hallinnassa ja ylläpidossa. Tutkimusryhmän tai vastaavan hallinnassa ja ylläpidossa on noin 40 konetta. Koneisiin asennetaan normaalisti sekä Windows että Red Hat pohjainen Linux (dualboot). Työasemien varmistus on henkilökunnan omalla vastuulla. Hallintoryhmän jäsenillä on mahdollisuus käyttää hallintopalvelimen varmuuskopiointia. Tietotekniikkaosaston hallinnassa ja ylläpidossa on 5 CompaqPro XP1000 palvelinkonetta.

Varmuuskopiointi tehdään KuKan (tietotekniikkaosaston Kumpulankampuksen kampuksyksikkö) nauharobotin avulla. Yksikön omassa hallinnassa ja ylläpidossa on 4 palvelinkonetta. Yksi niistä toimii www-, nimi- ja postipalvelimenä. Käyttöjärjestelmänä on Red Hat Linux -pohjainen Fedora Core 1. Päivityksiä suoritetaan yum-ohjelmalla, joka hakee automaattisesti päivityspaketteja yliopiston linux-asennuspalvelimelta. Varmuuskopiointi tehdään kerran viikossa palvelimen omalla nauha-asemalla.

Toinen on Windows 2000 Server, joka on hallintoryhmän käytössä. Siinä on tärkeitä laitoksen asiakirjoja ja taloushallinnon ohjelmistot. Päivityksiä tekee HIPin tukihenkilö. Varmuuskopiointi tehdään iltaisin palvelimen omalla nauha-asemalla.

Kaksi palvelinkonetta ovat Silicon Graphics koneita, joissa on IRIX käyttöjärjestelmä. Ne ovat tutkimusryhmien käytössä. IRIXin maksullisia päivitysversioita saadaan suoraan laitetoimittajalta. Varmuuskopiointi tehdään kerran viikossa omalla nauha-asemalla.

Fysiikan tutkimuslaitoksen vuoden 2003 henkilötövuosien määrä oli 96. Tutkimuslaitoksen tehtäviin kuuluu tutkimus ja tutkijakoulutus. Laitoksen henkilöstöön kuuluu 6,5 hallinnosta vastaavaa henkilöä ja 2 atk-tukea antavaa henkilöä: vanhempi suunnittelija ja laboratorioinsinööri.

Uusia työasemia hankitaan ”standardimikro” -konseptia hyväksi käyttäen yliopiston puitesopimus-toimittajilta.

Käytöstä poistuvien laitteiden hävittäminen tapahtuu yliopiston kierrätyskeskuksen ohjeiden mukaan ja sen kautta.

Laitos hoitaa pääasiallisesti laitehankintansa itse ja vastaa työasemien ja tulostimien ylläpidosta.

Tietotekniikkaosaston tuottamista atk-palveluista laitokselle tärkeimmät ovat verkko- ja sähköpostipalvelut, posterien tulostus sekä muutamat matemaattiset ohjelmistot kuten Matlab ja Mathematica.

CERNin IT-osaston palvelut ovat erityisesti HIPin CERNissä työskentelevien hiukkas- ja ydinfysiikoiden käytössä. Näihin kuuluu mm. laitehankinta, järjestelmätuki, verkkopalvelut, sähköviestit, tulosteet, tietokanta ja CPU-palvelut. CERNillä on myös suuri joukko ohjelmistolisenssejä, jotka ovat voimassa CERNin kanssa yhteistyötä tekevissä ryhmissä ja voidaan asentaa myös kotilaboratorion koneisiin.

4. Analyysi tutkimusaloittain; kokeellinen hiukkasfysiikka

4.1 Taustaa

Kokeelliselle hiukkasfysiikalle on luonteenomaista:

- kalliit investoinnit: kiihdyttimet miljardeja ja kokeet satoja miljoonia (€)
- suuret tutkimuskeskukset, kuten CERN: 20 jäsenmaata Euroopasta
- suuret yhteistyökollaboraatiot, esim. CMS-koe lähes 2000 henkilöä
- ilmaisimista saatavat suuret datamäärät: tuhansia teratavuja vuosittain
- mittavat tietojenkäsittelyhaasteet (ohjelmistot, prosessointi- ja tiedontallennuskapasiteetit, atk-yhteistyön toteutuminen kollaboraatioissa j.n.e.)

Mainitut piirteet asettavat erityiset reunaehdot tietojenkäsittelyn strategioille ja menetelmille kokeellisessa hiukkasfysiikassa.

HIP on mukana hiukkastutkimuksessa CERNin ja muiden tutkimuskeskusten kanssa huomattavalla panoksella. Tähän tutkimukseen liittyvä HIPin atk-strategia on vahvasti kytköksissä erityisesti CERNissä ja suurissa yhteistyökollaboraatioissa luotuihin strategioihin. Mittavimmat HIPin atk-hankkeet hiukkasfysiikassa liittyvät CMS-kokeen datan prosessointiin ja analysointiin.

4.2 Nykytila

Kokeellisen hiukkasfysiikan tutkijoiden käytössä on henkilökohtaisia työasemia (desktop) Kumpulassa toimintayksikössä yhteensä 21, Otaniemessä 1 ja CERNin toiminta-yksikössä 15. Lisäksi on kannettavia (laptop) yhteensä 10. Henkilökohtaisia tietokoneita on käytössä siis yhteensä viitisenkymmentä. Henkilökohtaisten työasemien lisäksi on laitoksen ilmaisineläboratoriassa Kumpulassa 5 työasemaa koelaitteistojen oheiskoneena. CERNin testisuihkualueella on yksi ilmainen-laitteiston tiedonkeruutehtävissä toimiva työasema sekä toinen ilmaisineläboratoriassa laitetestauksessa toimiva työasema.

Laitteista suuri osa on tutkimusryhmän ja laitoksen hallinnassa ja ylläpidossa. Kumpulassa kampusyksikön hallinnassa on yksi työasema ja CERNiin hankittuja työasemia hallinnoi CERNin atk-tukiryhmä. Laitoksen osuus hallintatyössä on 0,5 henkilötyövuotta, tutkimusryhmän osuus 0,2 työvuotta ja CERNin tukiryhmän osuus 0,1 työvuotta.

Kumpulassa on hiukkaskokeiden tiedontallennusta varten rakennettu 1,4 TB levypalvelin Silo, jota ryhmä hallinnoi. Asennettuna on myös 66 CPU:ta käsittävä Linux-ryväs Mill, jota hallinnoi yhteisesti HIP ja Fysikaalisten tieteiden laitos. Laitteisto on käyttövalmis. Ryväs liitetään NorduGridiin, johon on tällä hetkellä liitetty eri pohjoismaista yli 1300 solmua. Tällä hetkellä on jo olemassa neljän koneen testiryväs, joka on liitetty NorduGridiin.

Interaktiivinen analyysityö hoidetaan tällä hetkellä tutkijoiden omissa työasemissa.

CMS Monte Carlo tuotantoajoissa tuotettua dataa on varmistettu CSCn nauha-robotille, jonka nykyinen kapasiteetti on 60 TB, koska riittävän isoa nauharobottia ei ole Kumpulassa.

LHC-kokeiden analyysiryhmissä käytetään pääasiassa CERNin omaa Red Hat Linux jakelupakettia. CERNin oma Red Hat-jakelu on LHC-projektin tärkein sovellusalusta. Sun Solaris on toinen tuettu alusta, jolla varmistetaan ohjelmistojen toimivuus ja siirrettävyys, mutta muuten sen merkitys on vähenemään päin. CERNin käyttöjärjestelmä perustuu Red Hat Linux versioon 7.3. Koska Red Hatin käyttöjärjestelmät tulivat maksullisiksi, tullaan jatkossa siirtymään lähdekoodista käännettyyn Enterprise Linux-käyttöjärjestelmään, johon liitetään CERNin omat sovelluspaketit. CERNiä lähellä olevat grid-projektit, LHC Computing Grid projekti (LCG) ja European Data Grid (EDG) sekä sen seuraajaprojekti, tukevat samoja alustoja kuin CERN. NorduGrid projektin päälusta on myös Red Hat Linux. Red Hat Linux on myös suosittu suurenergiafysiikassa maailmanlaajuisesti.

Kokeellisen fysiikan sovellusohjelmistojen asennuksen tekevät tällä hetkellä tutkijat itse. CMS-ohjelmistot on saatavilla CDROM/DVD-levyllä, joka helpottaa ohjelmistojen asentamista tutkijoiden työasemiin. Sekä CERNissä että Fermilabissa on selvä pyrkimys käyttää pelkästään avoimen lähdekoodin ohjelmistoja, joten lisenssiongelmat ovat vähenemässä. Kaikkien suurien hiukkaskokeiden Monte Carlo-tuotantoa ajetaan atk-keskuksissa eri puolilla maailmaa. Tuotannossa tarvittavat ohjelmistot ovat saatavilla keskitetysti CERNistä ja Fermilabista verkon kautta.

Tutkijoiden työasemien varmistus on tutkijoiden omalla vastuulla. Silo levypalvelimen systeemilevy on KuKan nauharobotin varmistama ja se on kytketty KuKan konehuoneen UPSiin (Uninterruptable Power Supply) sähkökatkokkien haittojen minimoimiseksi. Mill-ryppään edustakone varmistetaan samalla tavalla kuin Silo, mutta laskentanoodien levyjärjestelmiä ei ole tarkoituksenmukaista kytkeä UPSiin eikä ottaa systeemilevyistä varmistuksia nauhalle.

Työasemien turvapäivityksissä HIPin tukihenkilöstö avustaa, mutta työasemien tietoturvasessa olisi parantamisen varaa, sillä tietoturvapäivityksien aloitteen tekeminen jää usein tutkijoille itselleen. CERNin Red Hat Linux päivitetään nopeasti, jos laajoja turva-aukkoja ilmaantuu ja CERNissä on käytössä automaattinen verkon yli toimiva käyttöjärjestelmän päivitysjärjestelmä, joten tuki CERNin päässä on hyvin hoidettu.

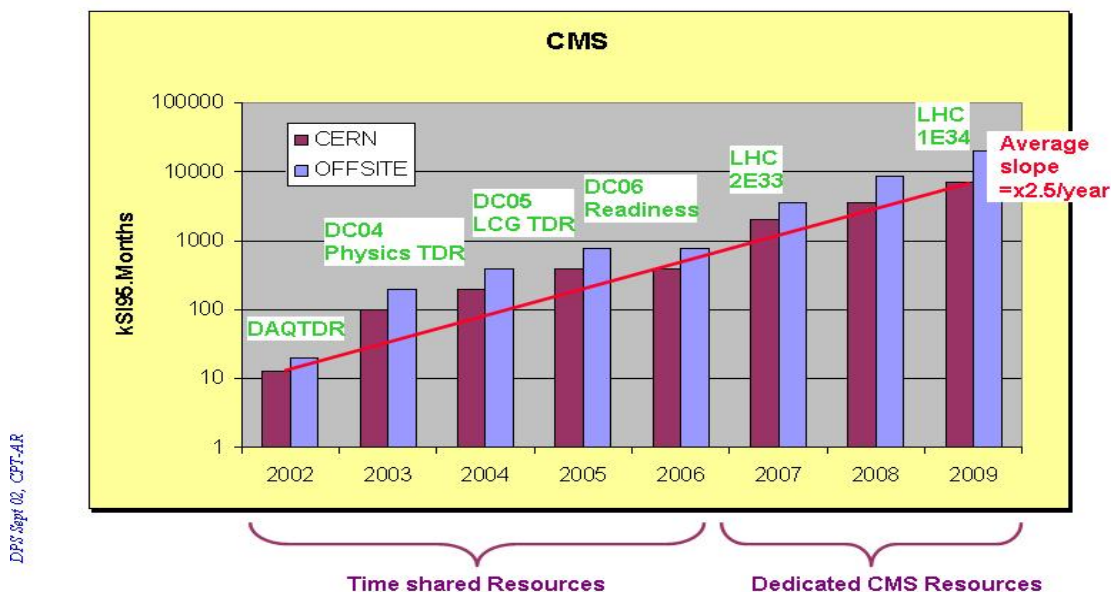
Ryhmän käytössä on videoneuvotteluohjelmisto VRVS, jota käytetään mm. CERNin yhteyksissä.

4.3 Tavoitetilä

Tavoitteena on, että työasemien hankinta tehdään mahdollisuuksien mukaan tietotekniikkaosaston hankintayksikön kautta.

Kokeellisen hiukkasfysiikan tutkijoilla on riittävästi eräajo- ja interaktiivista laskentakapasiteettia, joka mahdollistaa haastavan, mielenkiintoisen ja merkittävän fysiikan analyysin teon sekä LHC-että CDF-datalla. Tärkeää on, että tämä kapasiteetti on käytettävissä koti-instituutissa eli HIP:n tapauksessa Kumpulassa ja Jyväskylän yliopistossa. Käytettävän laskenta-alustan on oltava yhteensopiva sekä CERNissä ja LHC Computing Gridissä että Fermilabissa tuettujen alustojen kanssa. CPU-kapasiteetin lisäksi levy- ja nauhatilaa on riittävästi CPU-tehoon nähden. Levy- ja nauhatila tulisi liittää saumattomasti CERNissä yleisesti käytettyihin massamuisti-ohjelmistoihin ja niiden käyttö tulisi sujua helposti ja vaivattomasti grid-työkaluja käyttäen sekä paikallisverkosta (LAN) että internetistä (WAN) käsin. Yliopiston runkoverkossa ja palomuurin suorituskyvyssä sekä FUNETissa on huomioitu huomattavat datamäärien siirrot CERNiin ja muihin hiukkaskokeiden aluekeskuksiin grid-ohjelmistojen avulla HIPin tutkijoiden käyttämistä Linux-ryppäistä ja levypalvelimista.

HIP osallistuu ja tukee keskeisiä suurenergiafysiikan grid-hankkeita. HIPissä on hyvä grid-ohjelmistojen tietotaito ja tutkijat voivat käyttää verkon yli NorduGrid- ja LCG-resursseja tieteelliseen laskentaan.



HIPin osuus CMS-kokeen datan prosessoinnista on noin 2 prosenttia. Kuvassa on laskettu CPU-tehon kokonaistarve CMS-kokeessa vuosittain. Skaalattuna koko CMS-kokeen resurssitarpeisiin on HIPin ryhmän tarve CPU- ja levykapasiteetin osalta muutaman vuoden tähtämällä seuraava:

- levytilan tarve: v. 03/04/05/06/07 1.0/2.5/6/15/40 TB
- CPU-tehon tarve: v. 03/04/05/06/07 1.3/3.2/8/20/50 kSpecInt95

Tyypillinen työasema v. 2003 vastaa n. 100 SpecInt95 eli CPU-tarve v. 2007 tyydyttäisiin n. 500 nyky-CPUlla (Pentium IV).

CDF:n data-analyysia varten on alustavasti arvioitu, että vuoteen 2006 mennessä tarvitaan noin 5 TB:a levytilaa ja saman verran robotti-kapasiteettia. CPU-tehon tarve arvioidaan olevan 5-10 kSpecInt95 luokkaa. Nämä luvut riippuvat lähinnä siitä miten paljon luminositeettia Tevatron kiihdytin pystyy antamaan CDF-kokeelle ja siitä miten monipuolisesti CDF:n fysiikan analysointia tullaan tekemään Kumpulassa.

LHC:n TOTEM-koe tulee olemaan hyvin sidottu CMS-kokeeseen sekä online- että offline-mielessä, joten on luonnollista, että sen sovellusohjelmisto ja data tulee olemaan hyvinkin paljolti integroitu CMS:n vastaavien kanssa. Kuitenkin, johtuen siitä, että datastreamit ja MC tuotanto tulevat olemaan hyvin erilaisia kuin HIPn CMS-ryhmän vastaavat, on odotettavissa, että yllä olevat CMS-tarpeet joudutaan skaalamaan tekijällä 1,5, jotta yhteisen CMS/TOTEM-kokeen HIPin ryhmien tarpeet datan prosessoinnissa ja MC tuotannossa voidaan tyydyttää.

Koska ongelmat ovat samankaltaiset kaikissa hiukkaskokeissa on nähtävissä, että voidaan saada suuria synergiaetuja, jos suurempia ryppäitä ja levypalvelimia suunniteltaessa huomioidaan kaikkien hiukkaskokeiden tarpeet ja ohjelmistojen yhteensopivuus. On pyrittävä siihen, että CSC voisi tarjota merkittävän osan kaikkien HIPin ryhmien tarvitsemista laskentapalveluista.

Kumpulan kampukselle keskittyvät laskentaintensiivisten tieteiden alat tuovat mahdollisuuden rakentaa yhteisiä laskentajärjestelmiä, joiden laskentateho on suurempi, kuin mitä yhdellä laitoksella on varaa toteuttaa. HIPin ja FTL:n välistä vankkaa yhteistyötä on jatkettava ja kehitettävä sekä yhteistyömahdollisuuksia on edelleen kartoitettava muidenkin Kumpulan

yksiköiden kanssa, joilla on suuria laskentatarpeita. Tällaisia ovat erityisesti kemian ja Ilmatieteen laitokset. HIPin haastavat laskentaongelmat tarjoavat hyviä opinnäytetöiden aiheita.

Kokeellisen hiukkasfysiikan tutkijoiden koneiden käyttöjärjestelmät seuraavat CERNin Red Hat Linux jakelua ja niitä ylläpidetään ammattimaisesti automaattisin menetelmin, mikä korottaa tietoturvasaon. Tutkijoiden kotihakemistoista otetaan automaattiset varmistukset.

Verkon kasvava tiedonsiirtokapasiteetti antaa mahdollisuuden parannettuihin videokonferenssi-sovellutuksiin, kuten VRVSn jatkokehittely Access Grid.

4.4 Toimenpiteet, resurssit ja aikataulu tavoitetilan saavuttamiseksi

Yhteistyötä Fysikaalisten tieteiden laitoksen kanssa kehitetään edelleen. HIP osallistuu yhdessä ::n kanssa konsortiohankkeeseen, johon on saatu rahaa Suomen Akatemialta Linux-ryppäiden hankkimiseksi vuoden 2004 aikana. Saaduilla varoilla rakennetaan Kumpulaan noin 200 CPU:n laitteisto.

Osa Akatemian rahasta pyritään käyttämään erityisesti kokeellisen hiukkasfysiikan tarvitseman levykapasiteetin lisäämiseen vuoden 2004 aikana. Tavoitteena on myös yhdessä FTL:n kanssa hankkia nauharobotti vuoden 2005 aikana, jos siihen saadaan tarvittavat varat.

Yliopiston huomattavaa työasemien "vapaata" laskentakapasiteettia öisin ja viikonloppuisin tulisi yrittää hyödyntää kustannuksien vähentämiseksi. OpenMosix-ryppään moonshine kokeilun tulokset olivat lupaavia ja tälle projektille tulisi kehittää jatkoa. Tulossa olevat Linux-työasemat mahdollistanevat ainakin periaatteessa Condor-G tyyppiset ratkaisut.

Yhteydenpitoa CSC:n ja muiden Kumpulan kampuksen laitoksien kanssa tulisi parantaa entisestään. CSC:n, Suomen Akatemian ja Opetusministeriön kanssa on neuvoteltava kokeellisen hiukkasfysiikan huomattavien laskentatarpeiden rahoittamisesta.

5. Analyysi tutkimusaloittain; teoreettinen ja laskennallinen fysiikka

5.1 Taustaa

Teoreettisen ja laskennallisen fysiikan saralla Fysiikan tutkimuslaitoksen laskennalliset haasteet koskevat ennen kaikkea sen teoriaohjelman laskennalliseen mallintamiseen keskittyvien ryhmien erityisongelmia. Karkeasti ottaen laskennalliset haasteet voidaan jakaa kahteen kokonaisuuteen:

- a. Huomattavan suurta muistia tai erityisen suurta laskennallista kapasiteettia edellyttävät hankkeet. Osaltaan tähän voidaan laskea mukaan myös laskennallinen tutkimus, joka edellyttää sovellusresursseja, joiden hankkiminen yksittäisten tutkijoiden käyttöön ei ole taloudellisista syistä mielekäästä. Kaikissa näissä tilanteissa on perusteltua toteuttaa palvelut keskitetysti esim. kansallisessa laskennallisessa yksikössä, joka tarjoaa mittavat laskennalliset palvelut mm. supertietokoneiden, sovellusten ja niiden hyödyntämiseen ja tukeen liittyvän asiantuntijapalvelun muodossa. Suomessa tätä palvelua tarjoaa tällä hetkellä CSC.

- b. Hankkeet, joiden palveleminen voidaan toteuttaa myös paikallisesti ilman massiivista laskennallista kapasiteettia tai laaja-alaista palveluorganisaatiota. Ohjenuorana olisi ajatella, että se mikä on toteutettavissa paikallisesti, tulisi näin myös tehdä.

Kohta b) on erityisen perusteltu lukuisista syistä johtuen. CSC:n roolina on keskittyä niihin laskennallisen tieteen ongelmiin, joiden toteuttaminen paikallisesti ei olisi mielekästä joko kulujen tai niiden vaatiman erityisosaamisen takia. Monet laskennallisen tieteen ongelmat voidaan kuitenkin hyvin selvittää pienillä PowerPC-tyyppisillä koneilla, sillä vaikka niiden laskennallinen työ voi viedä useita CPU-päiviä, ne eivät välttämättä vaadi huomattavaa määrää muistia tai parallelisointia. Lisäksi oleellisesti kaikkien mallisysteemien laskennalliseen tutkimukseen liittyy mallin testausvaihe, joka voi olla hyvin työläs ja aikaavievä. Tämä on järkevintä ja ajallisesti tehokkainta toteuttaa paikallisesti pienillä pöytäkoneilla.

5.2 Nykytila

Fysiikan tutkimuslaitoksen teoreettiset tutkimusryhmät käyttävät tällä hetkellä huomattavia määriä CSC:n palveluita ja tietokoneresursseja hyväkseen. Niiden paikalliset tietokoneresurssit (Fysiikan tutkimuslaitoksen ja/tai Kumpulan atk-yksikön kautta) ovat hyvin rajalliset.

Fysiikan tutkimuslaitoksen tämänhetkinen laskennallinen kapasiteetti koostuu oleellisesti viidestä DEC Alpha -arkkitehtuuriin pohjautuvasta laskentayksiköstä (hipxp-sarja), jotka on hankittu vuonna 1999. Näiden lisäksi Fysiikan tutkimuslaitoksen käytössä on muutamia Kumpulan kampusyksikön alaisuudessa olevia palvelimia (esimerkkinä heavy), jotka samanaikaisesti ovat myös muun Helsingin yliopiston Fysiikaalisten tieteiden laitoksen ja henkilökunnan käytössä. Muita paikallisia laskennallisia resursseja ei oleellisesti ole.

Muutamia Fysiikan tutkimuslaitoksen teoriaohjelman ryhmät ovat myös osa muuta kuin Kumpulan kampusaluetta. Näistä biologisen fysiikan ryhmällä Teknillisessä korkeakoulussa on paikallisella tasolla käytössään muutamia palvelimia, jotka tosin ovat myös muiden fysiikan laboratorion (TKK) tutkijoiden käytössä. Mittavaan laskennalliseen tutkimukseen näitä palvelimia ei voi käyttää. Tämänhetkinen tilanne biologisen fysiikan ryhmässä on se, että työn alla olevien mallien testaaminen ja muu pienimuotoinen simulaatiotyö tehdään usein kannettavilla tietokoneilla ja pöytäkoneilla sekä Kumpulassa olevalla kapasiteetilla. Jyväskylässä URHIC-projekti on toistaiseksi saanut hyödyntää Jyväskylän yliopiston fysiikan laitoksen laitteita ja muita resursseja. Tämä järjestely on toiminut hyvin ja sen toivotaan säilyvän ennallaan. URHIC-projektin Helsingin ryhmän tilannetta koskevat ajatukset kiteytyvät osana yllä kuvattua koostetta.

Ylläpito ja varmistus on Kumpulan kampuksen osalta toteutettu edellä kuvatulla tavalla. Siten käytännön ylläpidon kannalta Fysiikan tutkimuslaitoksen Kumpulassa toimivat laboratorioinsinöörit sekä TKK:ssa ja Jyväskylässä toimivien ryhmien atk-henkilöstö antavat teoriaprojekteille niiden tarvitseman atk-tuen. Teknillisessä korkeakoulussa toimivan biologisen fysiikan ryhmän käytössä olevien pöytäkoneiden ja levypalvelimien ylläpidosta ja varmistuksesta huolehtii fysiikan laboratorion atk-henkilöstö. Kannettavien tietokoneiden osalta ylläpito on pääosin atk-henkilöstön toteuttamaa, kun taas varmistuksesta huolehtivat käyttäjät itse. Jyväskylässä ylläpito ja tuki on vastaavasti toteutettu paikallisen fysiikan laitoksen atk-henkilöstön toimesta, ja he vastaavat myös Unix-verkon varmuuskopioinnista. PC/Linux-puolella varmuuskopioinnin toteuttaminen on sen sijaan pääasiassa käyttäjien omalla vastuulla. Suurimittainen data-varastointi on yleisesti käyttäjien omalla vastuulla CSC:n palveluja ja resursseja hyödyntäen, ja tätä täydennetään paikallisesti kussakin ryhmässä CD- ja DVD-tallennusmenetelmiä käyttäen.

Laitehankinnat tehdään pääasiallisesti keskitetysti ja yhteistyössä Helsingin yliopiston Fysikaalisten tieteiden laitoksen kanssa. Tätä täydentävät muissa yliopistoissa olevien ryhmien keskitetyt hankinnat.

5.3 Tavoitetila

Teoriaryhmien kannalta nykyinen tilanne on siis ylläpidon ja varmistusten kannalta suhteellisen hyvä. Olennaisin ongelma liittyy laskennallisen kapasiteetin vähyyteen.

Fysiikan tutkimuslaitoksen teoriaryhmien laskennallinen tutkimus lepää pääosin CSC:n tarjoamien palvelujen ja laskentakapasiteetin varassa. Tehokasta paikallista laskentakapasiteettia, mikä olisi ensiarvoisen tärkeää lukuisista syistä johtuen, ei oleellisesti ole. Tärkeimpänä tavoitteena on siten kehittää laskennallisia resursseja siten, että Fysiikan tutkimuslaitoksen teoriaryhmillä olisi käytössään merkittävä paikallinen laskennallinen kapasiteetti CSC:n antamien asiantuntijapalvelujen ja resurssien lisäksi. Tämä on erityisen tärkeää juuri nyt, kun olemassaoleva kapasiteetti (hipxp-sarja) on piakkoin saavuttamassa elinkaarensa päätepisteen. Tähän hankkeeseen on tavoitteena kytkeä riittävät resurssit data-varastointiin ja etäylläpitoon sekä varmennukseen.

CSC:n rooli on laskennallisessa tutkimuksessa yleisesti hyvin tärkeä ja yhteistyötä tällä saralla on syytä kehittää edelleen. grid-laskenta ja sen kehittäminen on tästä yhteistyöstä yksi tärkeä esimerkki.

5.4 Tavoitetilan saavuttaminen

Tavoitetilan saavuttamiseksi Fysiikan tutkimuslaitos on osallistunut kansalliseen materiaalitutkimuksen grid-yhteistyöhankkeeseen, jossa tavoitteena on pystyttää useassa suomalaisessa yliopistossa ja tutkimuslaitoksessa oleva merkittävä, hajautettu laskennallinen kapasiteetti, joka tukee laskennallista materiaalitutkimusta. Hankkeen tarkoituksena on rakentaa usealla eri paikkakunnalla oleva tietokoneverkosto, jossa eri partnereiden osat voidaan kytkeä toisiinsa grid-teknologiaa käyttäen. Hankkeen tieteelliset tutkimusalueet keskittyvät materiaali-, nanorakenne- ja biologisen fysiikan tutkimukseen, mutta laajasti katsoen kyse on teoreettisen ja laskennallisen fysiikan vuorovaikutuksesta ajankohtaisilla tutkimusalueilla.

Yhteistyöhankkeen myötä Fysiikan tutkimuslaitoksen teoriaryhmille avautuu mahdollisuus hyödyntää erittäin tehokkaita paikallisia laskennallisia resursseja. grid-yhteistyöhankkeen puitteissa tämä luo myös mahdollisuuden hyödyntää muiden partnereiden resursseja soveltuvilta osin. Hankkeeseen kytkettävä etäylläpito ja varmennus sekä data-hallinta edelleen luovat tästä kokonaisuuden, jonka puitteissa toivottu tavoitetila voidaan saavuttaa ainakin seuraavan muutaman vuoden ajaksi. Toteutettavassa hankkeessa on CSC:llä tärkeä, ellei jopa keskeinen rooli.

6. Analyysi tieteenaloittain; teknologiaohjelma

6.1 Nykytila

Teknologiaohjelma toimii fyysisesti sekä Otaniemessä että CERNissä. Ohjelma tutkii hajautettuun laskentaan, ICT- ja tietohilaututkimukseen liittyviä kysymyksiä ja ylläpitää omaa laskentarypystä, joka vaatii hyvää asiantuntemusta sekä atk-laitteisto että -ohjelmistotasolla. Ohjelman toiminnalle on tärkeintä riittävän monipuolisen ohjelmistokehitysympäristön tarjoaminen jokaiselle tutkijalleen,

sekä mahdollisuus kokeilla ja kehittää ohjelmisto- ja laiteratkaisuja eristetyssä testiympäristössä, mikä ei saa häiritä muuta toimintaa. Tarkasteltaessa tarjollaolevia atk-palveluja kahdessa toimipisteessämme, CERN tarjoaa oman atk-tukiorganisaationsa kaikkien jäsenten ulottuville, kun taas Otaniemessä lähimmät tukipisteet ovat Kumpulassa sekä Suomen yliopistoverkko-organisaatiossa Funet'ssa, joka toimii CSC:n alaisuudessa Otaniemessä. Kehitettäessä HIPin omaa atk-strategiaa, on järkevää rajoittaa käsittelemään ohjelman tarpeita Otaniemessä osana HY:n hajautettua kampusta. CERN toimii tässä kuitenkin hyvänä vertailukohtana atk-palveluita kehitettäessä.

Koska tutkijoidemme tutkimusalue liittyy kiinteästi tietojenkäsittelyyn, ohjelma on pystynyt toimimaan varsin omavaraisesti erilaisissa ongelmatapauksissa. Jokainen tutkija on vastuussa vähintään oman työasemakoneensa ylläpidosta. Yleensä hankalimmat tilanteet liittyvät verkkopalveluissa ilmeneviin häiriöihin, joihin vaaditaan muiden osapuolten väliintuloa. Ohjelma on hankkinut tarvittavia laitteita ja ohjelmistoja itsenäisesti ja tämä laitekanta on pääsääntöisesti Linux-pohjaisia pc-koneita. Toinen, vaikkakin harvemmin käytetty käyttöjärjestelmä on MS Windows. Ohjelma suosii vapaan lähdekoodin sovellusohjelmia esim. toimistosovelluksissa OpenOffice ja pääsääntöisesti kaikki tietohila/laskentarypäs -kehityksessä käytetyt järjestelmätason ohjelmistot ovat vapaassa levityksessä. Ainoat lisenssipohjaiset kaupalliset ohjelmistot ovat liittyneet verkkoliikenteen turvaamiseen (palomuuriohjelmisto), sekä laskentaryppäiden loppukäyttäjien tarvitsemiin työkaluohjelmistoihin (esim. kääntäjät, simulointi ja matemaattiset laskentaohjelmistot kuten Matlab). Ohjelman Otaniemen toimipiste on kytketty yliopiston runkoverkkoon. Näiden palvelujen lisäksi ulkopuolelta on saatu tallennevarmistus-palvelu CSC:n tarjoamana.

6.2 Tavoitetila

Teknologiaohjelman tarpeita varten tulee varmistaa peruspalveluiden luotettava toiminta. Tämä pitää sisällään verkkoyhteydet, tietoturvan ja sähköpostipalvelut. Keskitettyjä palveluja koskien verkkotunnuksia ja niiden hallintaa, yhteisiä lisenssejä yleisimpiin ohjelmistoihin, jaettua verkkolevykapasiteettia ja ohjelmistopäivityksistä tiedottamista on syytä lisätä. Laitoksen laskentainstrumenttien yhteiskäyttöä olisi tehostettava mm. NorduGrid -yhteistyön avulla.

Koska ohjelma on fyysisesti erillään pääkampuksesta ja joutuu näin ollen olemaan omavarainen laitteiston ylläpidon ja teknisen tuen suhteen, mekanismeja keventää ylläpidolle aiheuttamaa kuormaa on kehitettävä. Tämä voidaan nähdä osana pitkäaikaista suunnitelmaa, jossa PC-työasemaverkoista siirrytään keveisiin verkkopäätteisiin, jotka lataavat rypäspalvelimilta standardin käyttöympäristön (thin client computing). Tallennustila tarjotaan verkkolevypalvelimelta, josta otetaan automaattisesti varmuuskopioita päivittäin, ilman että tutkijan tarvitsee huolehtia asiasta. Siirtyminen tämälämpäiseen ympäristöön voidaan tehdä asteittain siirtämällä pöytäkoneita osaksi palvelinrypästä uusien keveiden verkkopäätteiden (esim. bare bone pc) tai salkkumikrojen hankinnan yhteydessä.

Toinen merkittävä seikka teknologiaohjelman tarpeiden osalta on tutkijoiden liikkuvuuden tukeminen osana atk-strategiaa. Uuden työaseman tai salkkumikron lyhytaikaisen käytön mahdollistaminen esim. langattoman verkon osana (WLAN) ja verkkoidentiteettipalvelujen muodossa tulisi olla mahdollista molemmissa toimipisteissä. Valmiita ratkaisuja, kuten koko Otaniemen kampuksen käsittävä MediaPoli (www.otaverkko.fi) liittymä ja langaton lähiverkko, olisi järkevää käyttää hyväksi.

Vaikka teknologiaohjelma rekrytoi aktiivisesti Teknillisen korkeakoulun tekniikan ylioppilaita ja työllistää heitä kesäohjelmien kautta, tiiviimpi tutkimusyhteistyö saattaisi mahdollistaa muidenkin

kuin CSC:n kautta saatavien atk-peruspalvelujen tuomisen teknologiaohjelman ulottuville. Eräiden tällaisten palvelujen tuottaminen Otaniemen toimipisteessä olisi todennäköisesti kokonaisuuden kannalta edullisempaa järjestää isäntäyliopistona toimivan TKK:n kuin HY:n Kumpulankampuksen kautta.

Tekniset apuvälineet hajautettujen projektien suorittamiseen ja ohjaamiseen ovat myös varsin tärkeitä. Teknologiaohjelma otti vuonna 2003 käyttöön laadukkaat Polycom Viavideo II –laitteistot molemmissa toimipisteissään, jotka olivat edellytys neljän erillisen tutkimusryhmän yhteisten Internet-pohjaisten videopalaverien järjestämiseen osana Tekes-hanketta. Tähän mennessä laitteiston käytöstä saavutetut kokemukset ovat olleet rohkaisevia. Huomattakoon lisäksi, että yksittäisten tutkijoiden osallistuminen kokouksiin mahdollistuu jo yksinkertaisilla Web-kameroilla ja Vrvs/Access Grid –ohjelmistoilla.

Vuonna 2004 CERNin tekninen osasto on suunnittelemassa asteittaista siirtymistä verkkopuheluihin (Voice over IP, Voip). Tämä on toinen yksityisellä sektorilla voimakkaasti kasvava tekninen innovaatio, joka tuo merkittäviä kustannussäästöjä yritysten sisäisessä ja konsernien toimipisteiden välisessä kommunikaatiossa. Uusien laitepohjaisten voip-puhelimien tarjonnan myötä käyttökokemus on täsmälleen sama kuin perinteisellä lankapuhelimella. Teknologiaohjelma pystyisi toteuttamaan instituutin sisäisenä pilotti-projektina kartoituksen tarvittavien laitteistojen ja ohjelmistojen integroinnista sekä mallitoteutuksen mikäli tälle haettaisiin ulkopuolista rahoitusta. HIP kokonaisuudessaan voisi saavuttaa merkittäviä kustannussäästöjä ottamalla verkkopuhelut käyttöön eri toimipisteissään, jopa ennen kuin CERN on valmis toteuttamaan tämän laajassa mittakaavassa.

6.3 Tavoitetilan saavuttaminen

Teknologiaohjelman tulevien vuosien laskentatarpeet tulevat turvattua materiaalfysiikan grid-rahoituksen turvin, johon muutkin ohjelmat ovat aiemmin viitanneet. Hanke tulee kehittämään myös yhteistyötä CSC:n kanssa, mallissa, jossa nämä tuottavat keskitettyjä ylläpitopalveluja hankkeeseen osallistuville tahoille. Teknologiaohjelman atk-tarpeet HY:n etäyksikkönä Otaniemessä tulee huomioida tasapuolisesti ja kirjata selkeästi jomman kumman isäntäyliopiston (TKK, HY) vastuulle. Ohjelman toimintaa kehitettäessä, havaittuihin puutteisiin puututaan tarvittaessa ohjelman sisäisillä hankkeilla, jolla järjeistetään eritoten järjestelmien ylläpitoon liittyvät menettelytavat. ja turvataan tutkijoiden tehokas työskentely ja perustutkimukseen käytettävien resurssien maksimointi. Tähän suuntaan tähtäviä hankkeita on ollut vireillä myös CERNissä osana tulevia OpenLab yhteishankkeita, jotka liittyvät tietokoneavusteiseen ryhmätyöskentelyn ja opetuksen kehittämiseen (Computer Supported Collaborative Working and Learning).