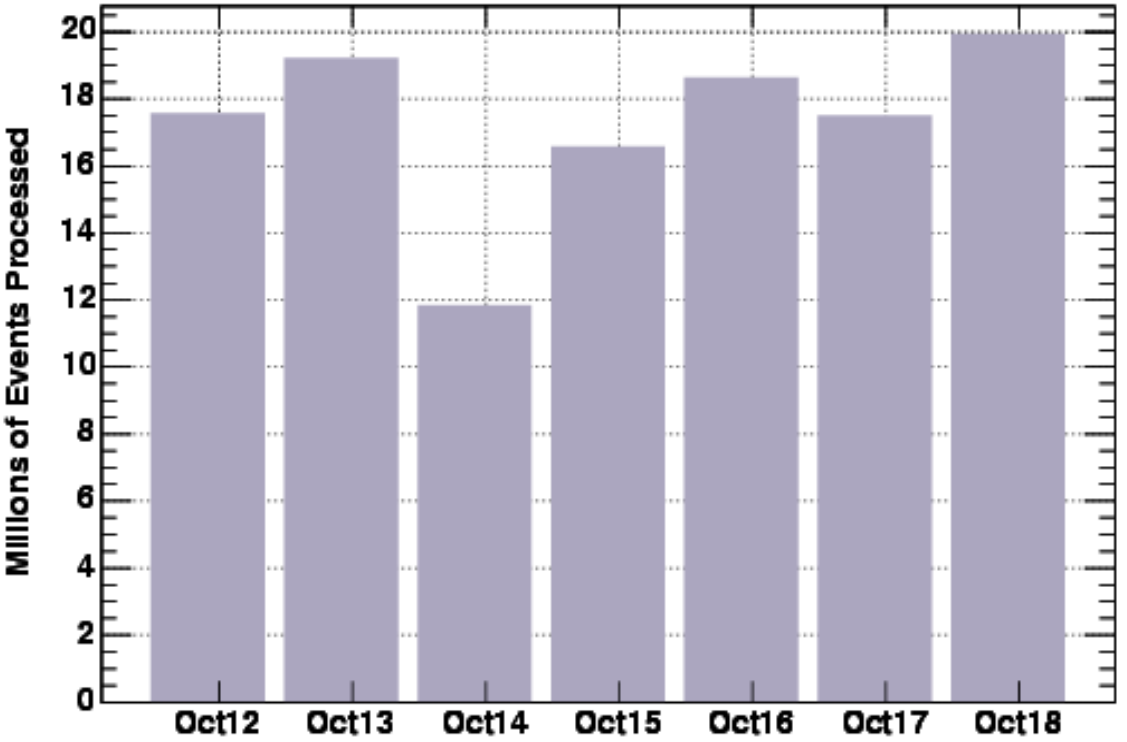
CDF Computing in 2006

- Outline
  - Operations
  - Software
  - Computing strategy
    - Transition to the GRID
    - Optimization of the resource usage
- Summary

Operations: data processing

- **'single-pass processing' scheme a continuing success:**
  - Offline reconstruction follows data taking within the 2 months
  - All the data taken up to Sep 01 2006 are fully calibrated
  - **> 95% the data processed, expect completion of processing by October 31, on schedule**



- **Smooth operations:**
  - total of 215/pb (350M events)
  - peak 24M events/day
  - Crash rate : ~10^{-8}
  -
  -

- **The physics groups already started updating the analysis datasets**

Operations: ntuple production

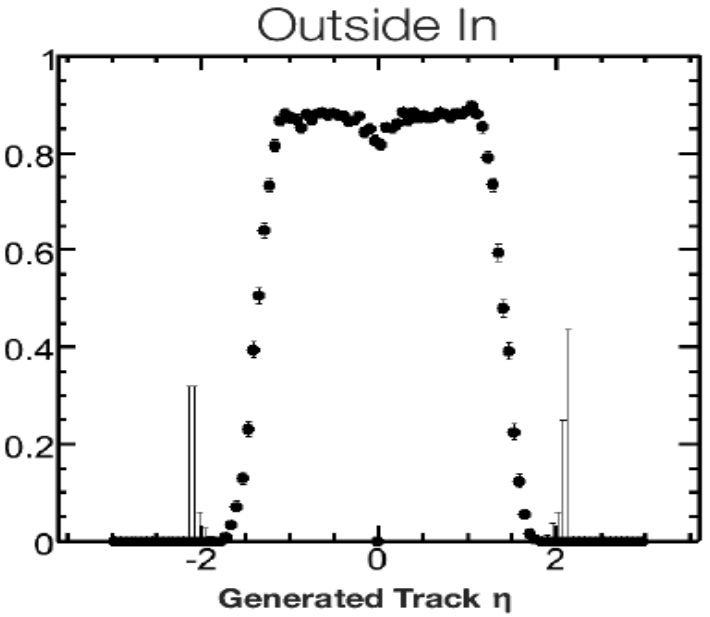
- **Spring'06:** Ntuple making code reached production-level stability
- production of 3 major ntuples (used by >90% of all the CDF physicists) incorporated into the offline reconstruction framework
- all the infrastructure (book-keeping, error recovery etc) shared with the data reconstruction
- **Ntuple Production Task Force: several months of successful operations**
- **Established unified production cycle for data reconstruction and ntuples**
  - Ntuples produced in parallel with the data reconstruction
  - Robust, stable and maintainable operations
  - Shared, growing expertise in the collaboration

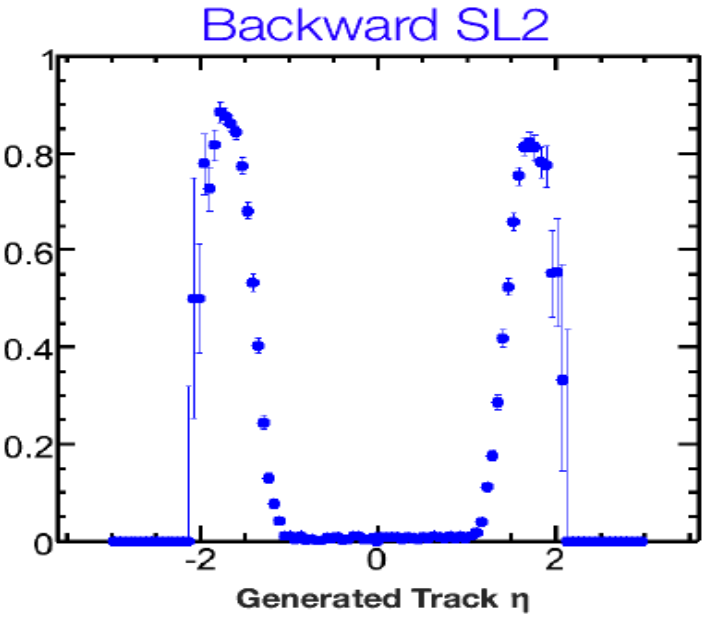
Operations: automating the calibrations

- calibrations for offline reconstruction: the most labor-intensive part of the offline operations
- Ongoing project to automate production of the calibration constants
  - Reduce effort to support 1-pass production
  - convert expertise of the current calibrators into a 'smart' software infrastructure, preparing for running production in 2010
  - Reduce latency of the calibration process

Software: the last production release

- **Incorporate all the desired improvements in the physics performance of the reconstruction algorithms**
  - Increased acceptance for the forward tracks
  - Significant gains for the high p_T leptons, b-tagging
  - increased robustness of the tracking at high luminosity

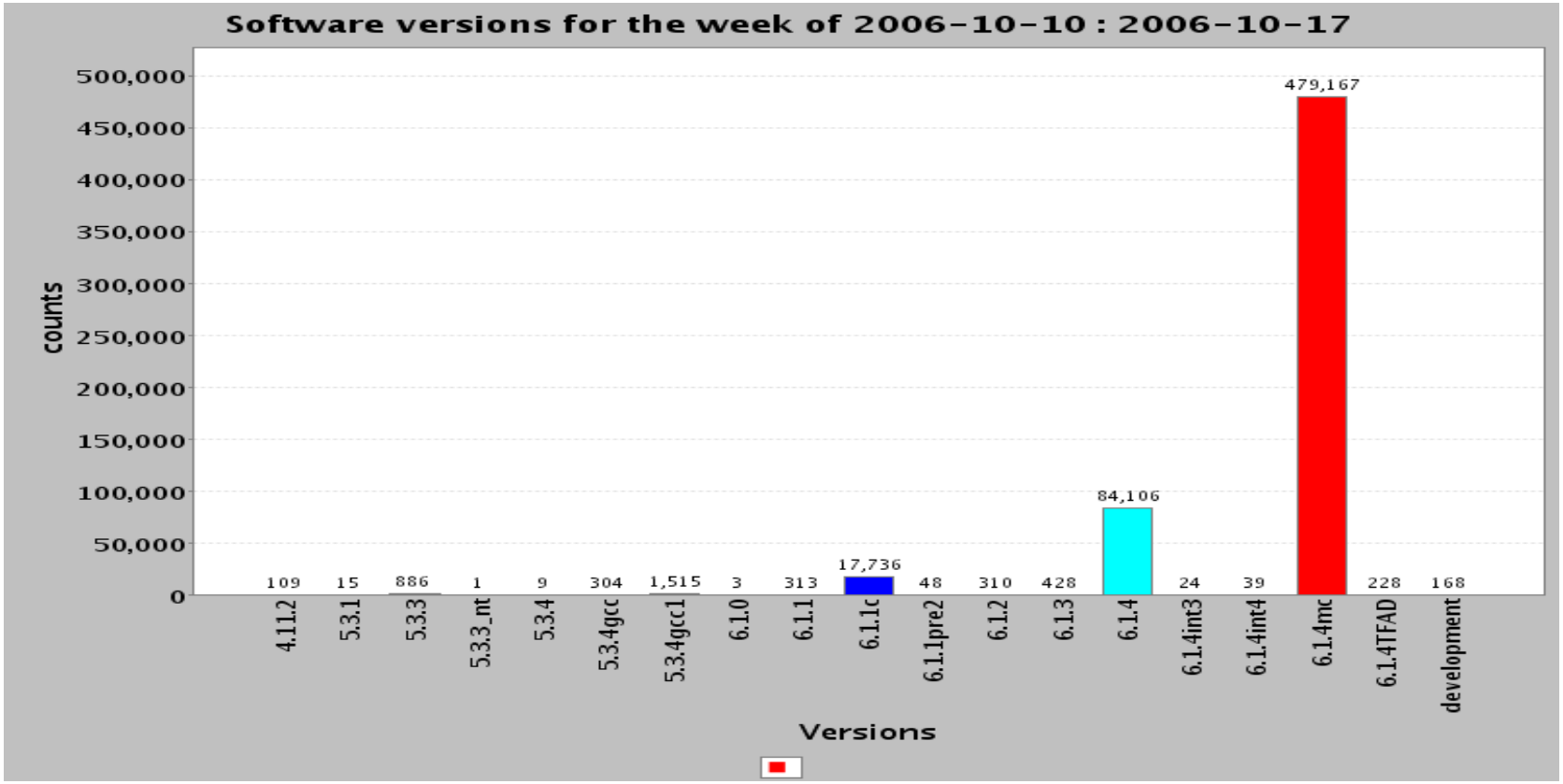




- **Deployment of the final production release: 1st half of 2007**

Software: reducing the maintenance load

- consolidated on using a very small number of the different versions of the CDF offline software (6.1.4*) , reducing overhead of the software maintenance



MC production

- Centrally supported infrastructure shared by everybody in the collaboration
- Efficiently using GRID resources in opportunistic mode
  - no CDF-specific requirements to the hardware/software of the GRID computing elements
    - **Can run on any GRID site which lets CDF jobs in**
  - In 2006 MC production executable runs 1.6 times faster due to compiler optimizations
- **2 years of stable operations**

Computing: optimizing resource usage

- **Strategy motivated by physics**

- Emphasize reconstruction and analysis on Fermilab-based CPU, exploit closeness to the data storage (tape and disk)

  - optimize performance of the reconstruction and analysis algorithms as needed

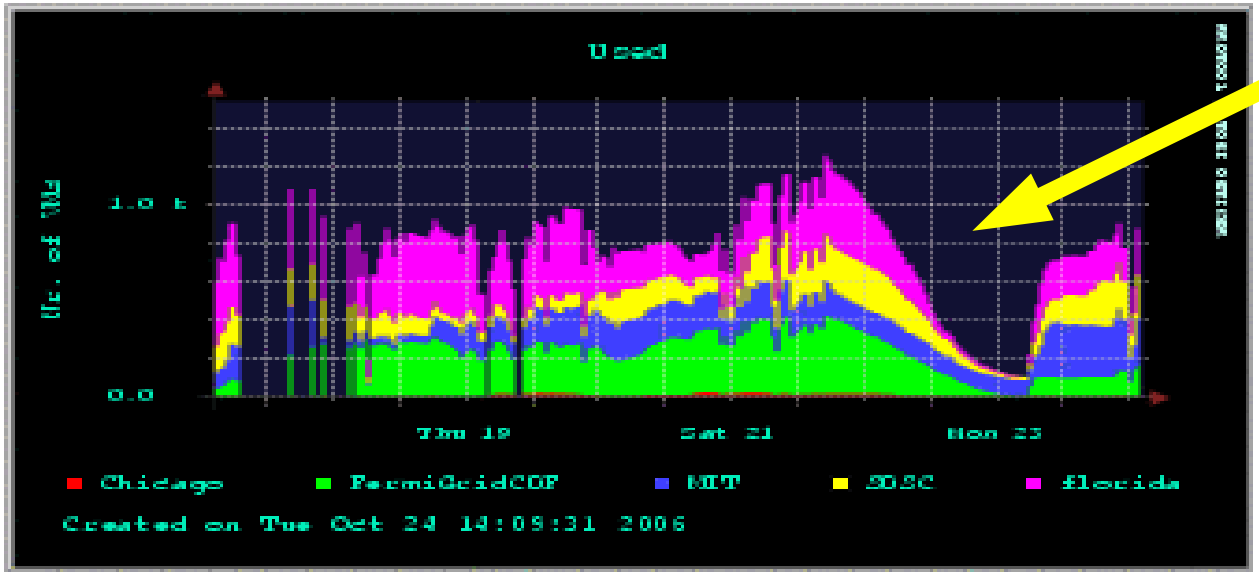
- Perform bulk of MC generation off-site

Moving to GRID-based computing infrastructure

- CDF established a unified infrastructure of the batch computing farm: CAF (Central Analysis Facility, based on the **Condor** GRID toolkit)
- Computing trends:
  - CDF-dedicated off-site CPU facilities (dCAF's) becoming generic GRID pools
- GRID tools improve efficiency of the computing resource usage
  - move all the on-site CPUs into one GRID pool – Fermigrid
- New technologies taking CDF to the GRID:
  - North American CAF (NAmCAF)
  - LCGCAF (for European GRID sites)

GRID computing: NAmCAF

- **NamCAF:** single job submission point for CDF users, to Open-Science-Grid (OSG) sites across North America (Igor Sfiligoi, INFN/Frascati)
  - Looks & feels like the CAF (uses OSG tools underneath)
  - Except no CDF-specific hardware or software
- Accesses GRID sites at **MIT, Fermilab, UCSD, Florida** and **Chicago**
- **Purdue, Toronto, Wisconsin, McGill** in the queue to be added

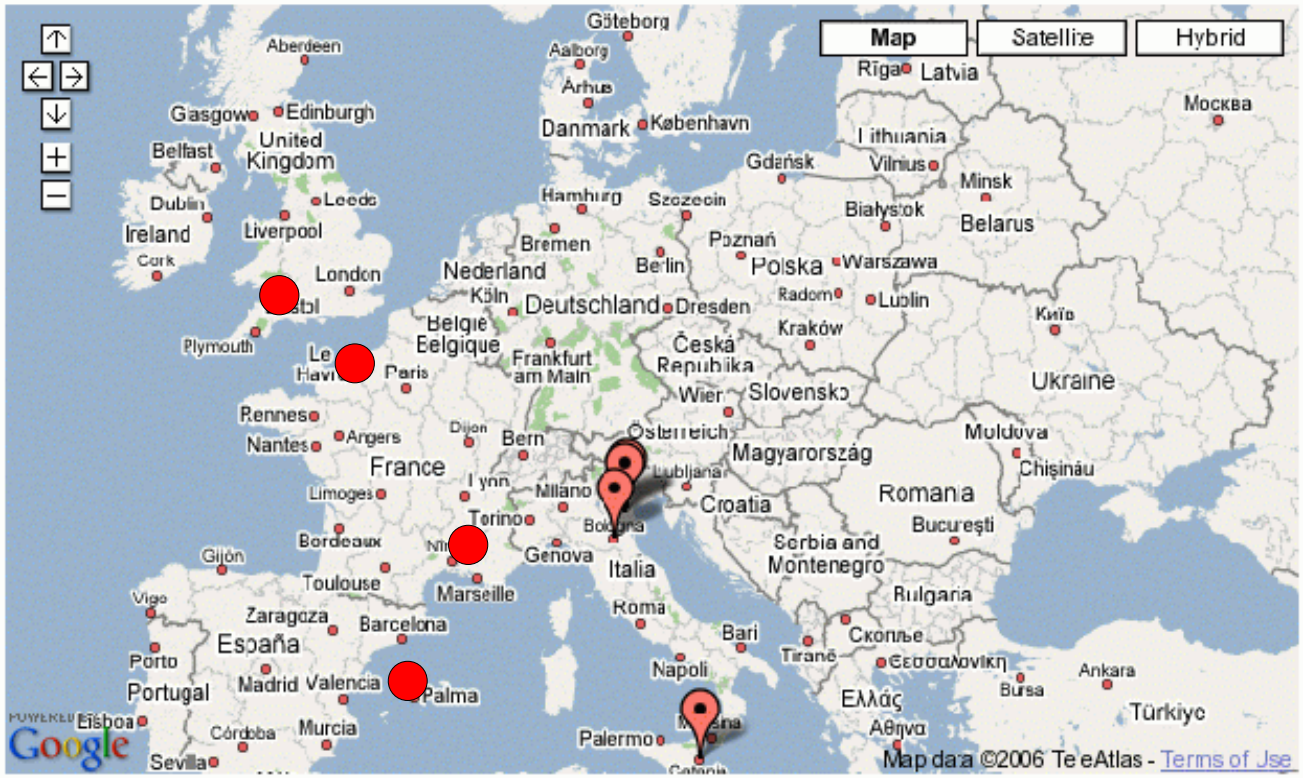


no CDF jobs waiting in the queue

- **Provides up to 1000 job slots already (CAF has 1800 slots) in opportunistic mode**

GRID computing: LCGCAF

- Provides a single job submission point for CDF users, to LHC-Computing-Grid sites across Europe
  - Looks & feels like the CAF (but uses IGEE middleware underneath)
  - Developed by CDF Italian group led by D.Lucchesi
- Currently accesses 4 INFN sites, 2 in Spain, 2 in UK & France



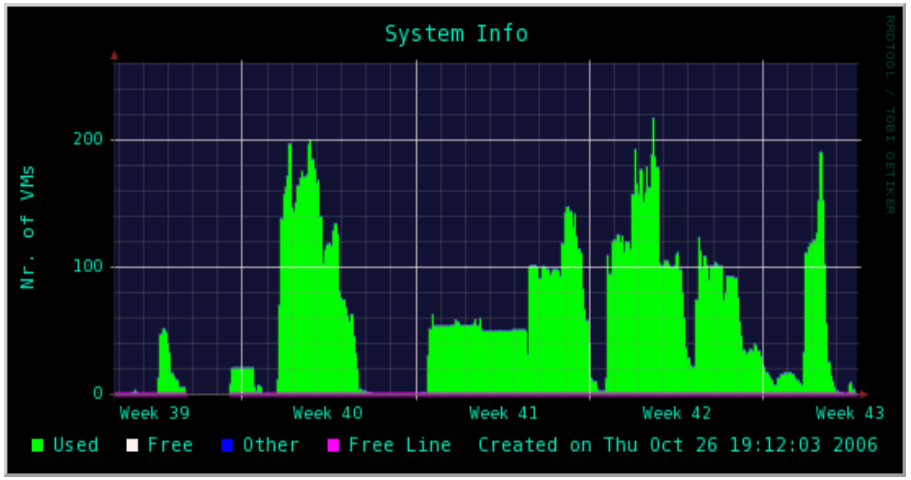
~2700 CPU's in total

Both LCGCAF and NamCAF have large potential for expansion

- more sites
- More CPU/site

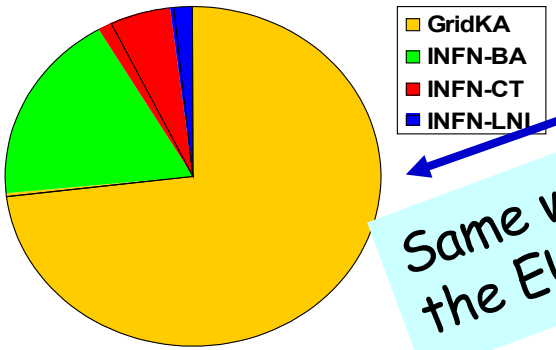
LcgCAF Usage (from D.Lucchesi)

- in production since fall'2006,generic users just start to use it
- Already provides simultaneous access to 200 CPU's

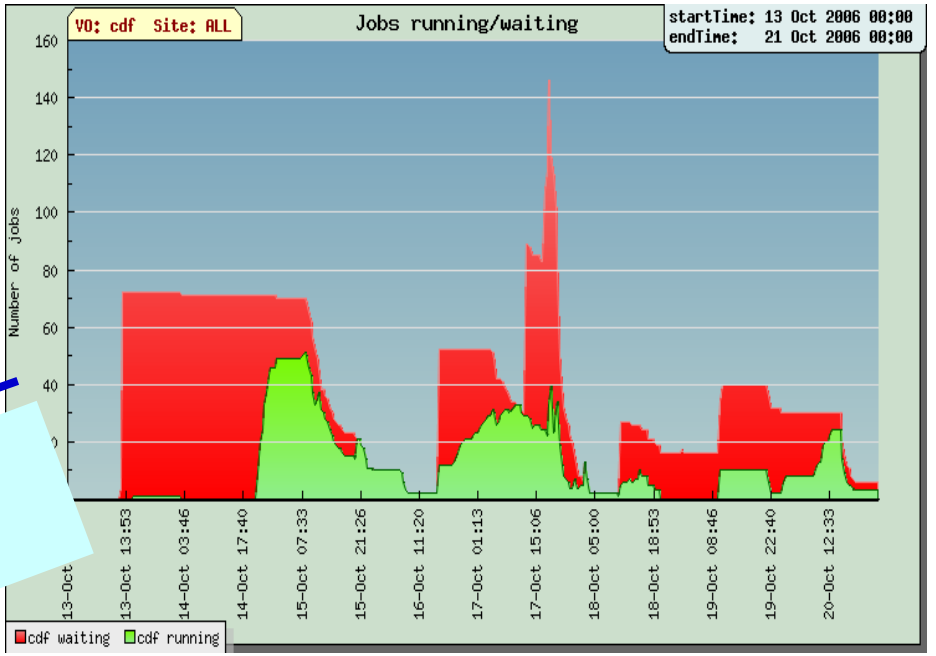


Running sections in a week as seen by GridIce monitor on Italian resources

Running jobs in the last month as seen by our monitor



Same week on the EU GRID



GRID computing: plan for the next years

- CDF successfully uses OSG and LCG resources accessing up to 1200 CPU's simultaneously in opportunistic mode

The plan:

- **maintain limited number of GRID "entry points":**
  - FermiGRID: "local" GRID pool, access to CDF-dedicated resources
  - One GRID entry point per continent providing CPU resources in opportunistic mode (NamCAF, LCGCAF,...)
- **Explore large potential of the GRID**

Summary

- Very smooth operations of the CDF computing in 2006:
  - 1-pass reconstruction scheme: **1 year of smooth running**
  - Implementation of the automated calibration scheme in progress
  - Production of major ntuples integrated into 1-pass scheme
  - MC production: stable running
- Final improvements to the reconstruction algorithms to be deployed in early 2007
- CDF is efficiently using GRID resources in opportunistic mode
  - Access up to 1200 CPU's
  - pool of accessible CPUs expanding
- Development efforts ramping down
- **focus on stability and efficiency, moving towards 100% operations mode**