Flow Modeling on Massive Grids

Laura Toma, Rajiv Wickremesinghe
with
Lars Arge, Jeff Chase, Jeff Vitter

in collaboration with

NICHOLAS School of the ENVIRONMENT DUKE

Pat Halpin, Dean Urban

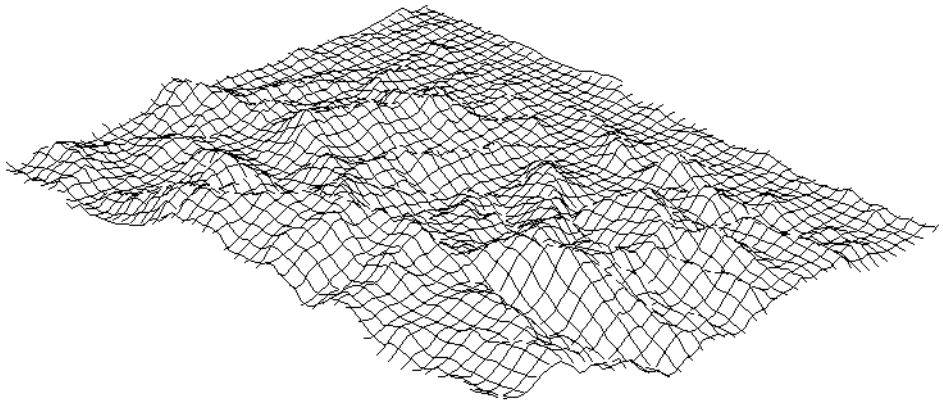
DUKE Computer Science

Applications

- Automatic estimation of terrain parameters
  - watersheds
  - drainage networks
  - topographic index
- Surface saturation
- Soil water content
- Erosion, Deposition
- Forest structure
- Species diversity
- Sediment transport

Modeling Flow

- Terrain modeled by grid of elevations



- Flow direction
  - direction water would flow from each cell
  - what does it mean on flat areas?

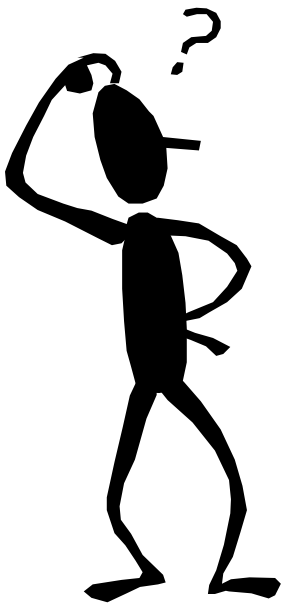
Massive Data

- Remote sensing data available today
  - NASA-SRTM (whole Earth: 5TB)
  - USGS (entire US at 10m resolution)
- Typical grid
  - East coast USA
  - 11,000 x 25,000 grid
  - 0.5 GB

DUKE
Computer Science

Scalability

- Current GIS software minimizes CPU time
- I/O is bottleneck
- Total time does not scale

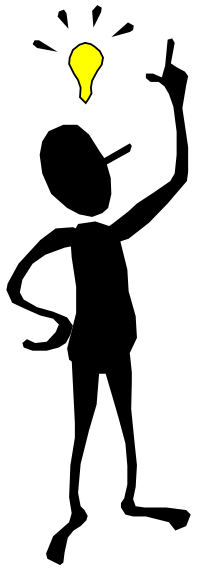


- ArcInfo
  - more than 3 days on East coast USA grid
- TARDEM
  - 2 days on 1/4 East coast USA grid

DUKE
Computer Science

Theoretical Results

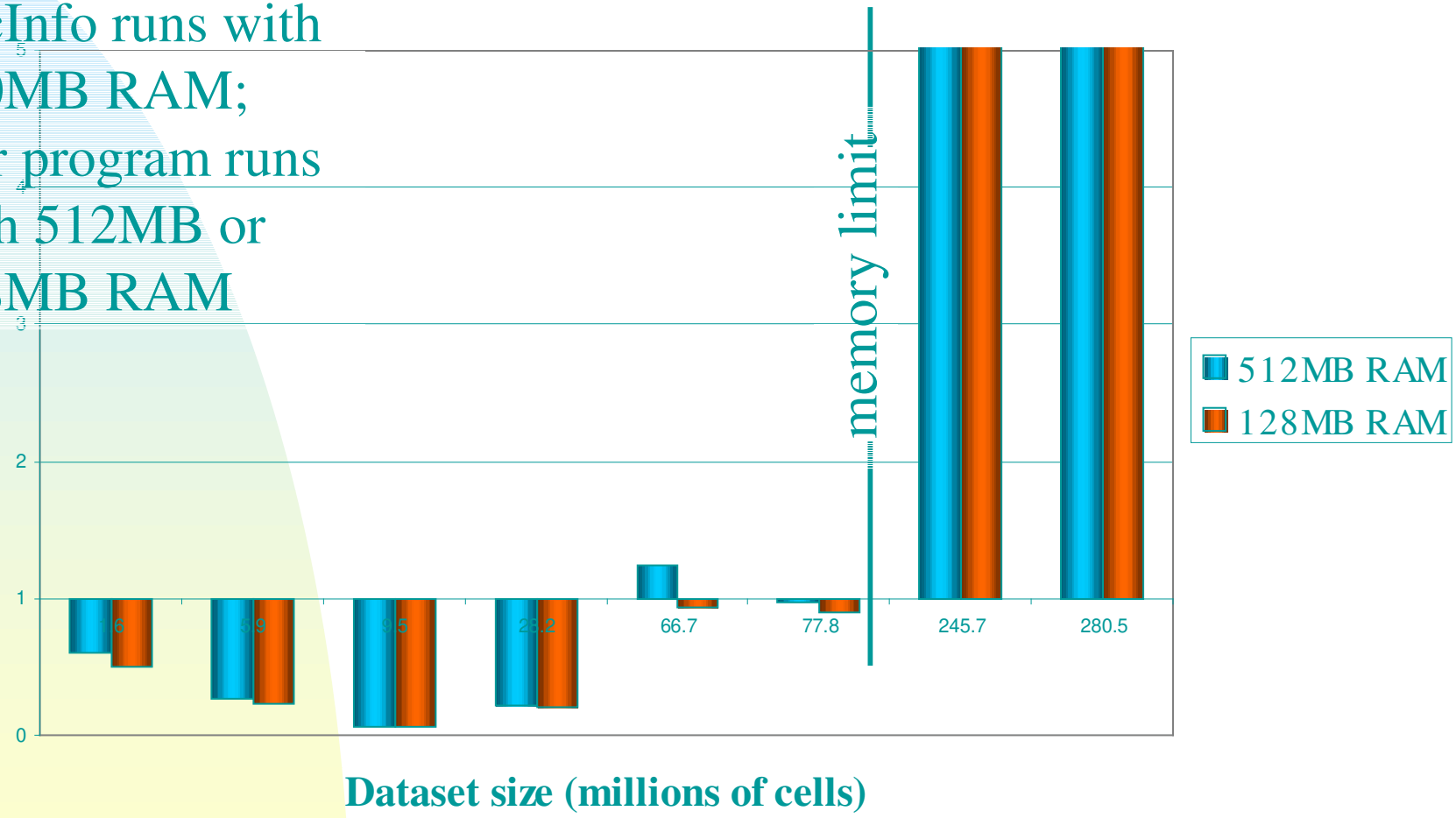
- I/O Model
  - disk much slower than CPU
  - count number of I/O operations (rather than CPU time)



- I/O Efficient Algorithms
  - identifying flat areas
  - assigning flow directions on plateaus
  - identifying watersheds
  - filling depressions/sinks
  - assigning flow directions on terrains
  - flow accumulation

Speedup over ArcInfo

ArcInfo runs with 750MB RAM; Our program runs with 512MB or 128MB RAM



DUKE Computer Science