A Problem-Based Learning Approach to GPU Computing

Robert Geist
Joshua A. Levine
James Westall

School of Computing
Clemson University

November 16, 2015
GPUs

- Since 2001, increasing importance in numerical computation due to high performance, low energy consumption
GPUs

- Since 2001, increasing importance in numerical computation due to high performance, low energy consumption

- NVIDIA Titan X GPU theoretical peak: 7 teraflops; idle: 9W; Battlefield4 MP: 258W
GPUs

- Since 2001, increasing importance in numerical computation due to high performance, low energy consumption
- NVIDIA Titan X GPU theoretical peak: 7 teraflops; idle: 9W; Battlefield4 MP: 258W
- Foundational results in computational science will require simulation of physical systems that are orders of magnitude larger than current systems
- Run on CPU-based systems?
GPUs

- Since 2001, increasing importance in numerical computation due to high performance, low energy consumption
- NVIDIA Titan X GPU theoretical peak: 7 teraflops; idle: 9W; Battlefield4 MP: 258W
- Foundational results in computational science will require simulation of physical systems that are orders of magnitude larger than current systems
- Run on CPU-based systems? We’d need a dedicated nuclear reactor!
GPUs

- Since 2001, increasing importance in numerical computation due to high performance, low energy consumption

- NVIDIA Titan X GPU theoretical peak: 7 teraflops; idle: 9W; Battlefield4 MP: 258W

- Foundational results in computational science will require simulation of physical systems that are orders of magnitude larger than current systems

- Run on CPU-based systems? We’d need a dedicated nuclear reactor!

- Titan Cray XK7 - Oak Ridge
GPUs

Teaching GPU programming is challenging ...
- Parallel algorithms, but also ...

GPUs

Teaching GPU programming is challenging ...

- Parallel algorithms, but also ...
- Control flow patterns
- Memory access patterns
- Memory hierarchies
- Staging techniques
- Synchronization primitives
τέχνη

- τέχνη is the Greek word for *art*.
- It shares its root with τεχνολογία, the Greek word for *technology*.
τέχνη

- τέχνη is the Greek word for art.
- It shares its root with τεχνολογία, the Greek word for technology.
- Method in development since 2004 (See Proc. SIGCSE’04, ’07, ’11, ’14)
- Foundation is cognitive constructivism (Piaget, Dewey, Rousseau)
τέχνη method

Pillars, upon which every course should be designed:
πέχνη method

Pillars, upon which every course should be designed:

- **problem-based learning** - Carefully designed problems demand that learners acquire self-directed strategies and critical knowledge; new: size and scope of the problem.
Pillars, upon which every course should be designed:

- **problem-based learning** - Carefully designed problems demand that learners acquire self-directed strategies and critical knowledge; new: size and scope of the problem.

- **visual domain** - Problems connect with computer graphics, image processing, or visualization.
Pillars, upon which every course should be designed:

- **Problem-based learning** - Carefully designed problems demand that learners acquire self-directed strategies and critical knowledge; new: size and scope of the problem.

- **Visual domain** - Problems connect with computer graphics, image processing, or visualization.

- **Cognitive apprenticeship** - Transfer master-apprentice relationship from physical skills arena to cognitive skills arena. Cognitive demands on the master must be authentic.
CPSC 4780/6780 at Clemson

- Taught every year since 2010
- Increasing enrollment each year
CPSC 4780/6780 at Clemson

- Taught every year since 2010
- Increasing enrollment each year
- Semester-long focus on a single problem class
- Real problem, not toy problems
CPSC 4780/6780 at Clemson

- Taught every year since 2010
- Increasing enrollment each year
- Semester-long focus on a single problem class
- Real problem, not toy problems
- Sacrifice some breadth of coverage for depth
- Focus on numerical, HPC, rather than accelerating graphics algorithms
CPSC 4780/6780 at Clemson

- Taught every year since 2010
- Increasing enrollment each year
- Semester-long focus on a single problem class
- Real problem, not toy problems
- Sacrifice some breadth of coverage for depth
- Focus on numerical, HPC, rather than accelerating graphics algorithms
- Problem choice is important!
Current Problem Choice

In modeling any of

- photon transport through participating media (clouds, leaves, water)
- cloud formation (water vapor - thermal energy)
- waves on water surfaces
- general classes of PDEs
Current Problem Choice

In modeling any of

- photon transport through participating media (clouds, leaves, water)
- cloud formation (water vapor - thermal energy)
- waves on water surfaces
- general classes of PDEs

coupling \textit{lattice-Boltzmann} (LB) methods with GPUs provides an interesting, highly parallel solution paradigm.
Lattice-Boltzmann Methods

- a class of *cellular automata* (CA), whose origins probably date to Conway’s *Game of Life*: 
Lattice-Boltzmann Methods

- a class of *cellular automata* (CA), whose origins probably date to Conway’s *Game of Life*:
  - each site on a 2D lattice is marked “populated” or “not populated”
  - each site follows only local rules, based on nearest-neighbor states, in updating itself
  - synchronous updates
  - global system behavior emerges from local rules
Lattice-Boltzmann Methods

- a class of *cellular automata* (CA), whose origins probably date to Conway’s *Game of Life*:
  - each site on a 2D lattice is marked “populated” or “not populated”
  - each site follows only local rules, based on nearest-neighbor states, in updating itself
  - synchronous updates
  - global system behavior emerges from local rules
- the most interesting CA have global behaviors with provable characteristics
Lattice-Boltzmann Methods

- replace discrete populations with continuous ones
Lattice-Boltzmann Methods

- replace discrete populations with continuous ones
- provable characteristic: solves a PDE!
Lattice-Boltzmann Methods

- replace discrete populations with continuous ones
- provable characteristic: solves a PDE!
- methods regarded as alternatives to FEMs
- comparable in speed, stability, efficiency
Lattice-Boltzmann Methods

- replace discrete populations with continuous ones
- provable characteristic: solves a PDE!
- methods regarded as alternatives to FEMs
- comparable in speed, stability, efficiency
- advantages (over FEM):
  - ease of implementation
  - parallelization
  - handling complex boundary conditions
Lattice-Boltzmann Methods

- replace discrete populations with continuous ones
- provable characteristic: solves a PDE!
- methods regarded as alternatives to FEMs
- comparable in speed, stability, efficiency
- advantages (over FEM):
  - ease of implementation
  - parallelization
  - handling complex boundary conditions
- principal drawback: counter-intuitive derivation; computational update (local) $\Rightarrow$ PDE (global)
Consider a flow of some density on a 1-D lattice (e.g., flow of heat along a 1-D wire). Assume:
A 1-D Transport Model

Consider a flow of some density on a 1-D lattice (e.g., flow of heat along a 1-D wire). Assume:

- lattice spacing $\lambda$
A 1-D Transport Model

Consider a flow of some density on a 1-D lattice (e.g., flow of heat along a 1-D wire). Assume:

- lattice spacing $\lambda$
- synchronous updates with time step $\tau$
A 1-D Transport Model

Consider a flow of some density on a 1-D lattice (e.g., flow of heat along a 1-D wire). Assume:

- lattice spacing $\lambda$
- synchronous updates with time step $\tau$
- $f_\pm(x, t) = \text{density at site } x, \text{ time } t, \text{ flowing in dir. } \pm 1$
A 1-D Transport Model

Consider a flow of some density on a 1-D lattice (e.g., flow of heat along a 1-D wire). Assume:

- lattice spacing $\lambda$
- synchronous updates with time step $\tau$
- $f_{\pm}(x, t) =$ density at site $x$, time $t$, flowing in dir. $\pm 1$
- at each time step, at each site, fraction $\sigma$ of $f_{\pm}(x, t)$ continues in current dir.; remainder reverses dir.
A 1-D Transport Model

Consider a flow of some density on a 1-D lattice (e.g., flow of heat along a 1-D wire). Assume:

- lattice spacing $\lambda$
- synchronous updates with time step $\tau$
- $f_{\pm}(x, t) = \text{density at site } x, \text{ time } t, \text{ flowing in dir. } \pm 1$
- at each time step, at each site, fraction $\sigma$ of $f_{\pm}(x, t)$ continues in current dir.; remainder reverses dir.

Obtain fundamental update:

\[
\begin{pmatrix}
    f_{+}(x + \lambda, t + \tau) \\
    f_{-}(x - \lambda, t + \tau)
\end{pmatrix}
= \begin{pmatrix}
    \sigma & 1 - \sigma \\
    1 - \sigma & \sigma
\end{pmatrix}
\begin{pmatrix}
    f_{+}(x, t) \\
    f_{-}(x, t)
\end{pmatrix}
\]
A 1-D Transport Model

Incremental form:

\[
\begin{pmatrix}
    f_+(x + \lambda, t + \tau) - f_+(x, t) \\
    f_-(x - \lambda, t + \tau) - f_-(x, t)
\end{pmatrix}
= \Omega
\begin{pmatrix}
    f_+(x, t) \\
    f_-(x, t)
\end{pmatrix}
\]

where \( \Omega = \begin{pmatrix}
    \sigma - 1 & 1 - \sigma \\
    1 - \sigma & \sigma - 1
\end{pmatrix} \).
A 1-D Transport Model

Incremental form:

\[
\begin{pmatrix}
  f_+(x + \lambda, t + \tau) - f_+(x, t) \\
  f_-(x - \lambda, t + \tau) - f_-(x, t)
\end{pmatrix}
= \Omega
\begin{pmatrix}
  f_+(x, t) \\
  f_-(x, t)
\end{pmatrix}
\]

where \( \Omega = \begin{pmatrix}
  \sigma - 1 & 1 - \sigma \\
  1 - \sigma & \sigma - 1
\end{pmatrix} \). Taylor expand left side:
A 1-D Transport Model

Incremental form:

\[
\begin{pmatrix}
  f_+ (x + \lambda, t + \tau) - f_+ (x, t) \\
  f_- (x - \lambda, t + \tau) - f_- (x, t)
\end{pmatrix}
= \Omega
\begin{pmatrix}
  f_+ (x, t) \\
  f_- (x, t)
\end{pmatrix}
\]

where \( \Omega = \begin{pmatrix} \sigma - 1 & 1 - \sigma \\ 1 - \sigma & \sigma - 1 \end{pmatrix} \). Taylor expand left side:

\[
\begin{pmatrix}
  \lambda \frac{\partial f_+}{\partial x} + \tau \frac{\partial f_+}{\partial t} + \left( \frac{\lambda^2}{2} \right) \frac{\partial^2 f_+}{\partial x^2} + \ldots \\
  -\lambda \frac{\partial f_-}{\partial x} + \tau \frac{\partial f_-}{\partial t} + \left( \frac{\lambda^2}{2} \right) \frac{\partial^2 f_-}{\partial x^2} + \ldots
\end{pmatrix}
= \Omega
\begin{pmatrix}
  f_+ \\
  f_-
\end{pmatrix}
\]

(1)
Seek expression describing evolution of site density, \( \rho(x, t) = f_+(x, t) + f_-(x, t) \), as both \( \lambda, \tau \to 0 \).
Seek expression describing evolution of site \textit{density},
\[ \rho(x,t) = f_+(x,t) + f_-(x,t), \] as both \( \lambda, \tau \to 0. \)

Assumptions:

- \( \tau \to 0 \) faster than \( \lambda \to 0. \)
  
  Write \( \lambda = \varepsilon \lambda_0 \) and \( \tau = \varepsilon^2 \tau_0 \) for small \( \varepsilon > 0. \)
A 1-D Transport Model

Seek expression describing evolution of site density, $\rho(x,t) = f_+(x,t) + f_-(x,t)$, as both $\lambda, \tau \to 0$.

Assumptions:

- $\tau \to 0$ faster than $\lambda \to 0$.
  Write $\lambda = \epsilon \lambda_0$ and $\tau = \epsilon^2 \tau_0$ for small $\epsilon > 0$.

- (Chapman-Enskog expansion) flow can be written as a small perturbation: $f_\pm = f^{(0)}_\pm + \epsilon f^{(1)}_\pm + \epsilon^2 f^{(2)}_\pm + ...$
A 1-D Transport Model

Seek expression describing evolution of site density, \( \rho(x, t) = f_+(x, t) + f_-(x, t) \), as both \( \lambda, \tau \to 0 \).

Assumptions:

- \( \tau \to 0 \) faster than \( \lambda \to 0 \).
  Write \( \lambda = \varepsilon \lambda_0 \) and \( \tau = \varepsilon^2 \tau_0 \) for small \( \varepsilon > 0 \).

- (Chapman-Enskog expansion) flow can be written as a small perturbation: 
  \[ f_\pm = f^{(0)}_\pm + \varepsilon f^{(1)}_\pm + \varepsilon^2 f^{(2)}_\pm + \ldots \]
  where \( f^{(0)}_+ + f^{(0)}_- = \rho \).
A 1-D Transport Model

Seek expression describing evolution of site density, 
\[ \rho(x, t) = f_+(x, t) + f_-(x, t), \] as both \( \lambda, \tau \to 0. \)

Assumptions:

- \( \tau \to 0 \) faster than \( \lambda \to 0. \)
  Write \( \lambda = \varepsilon \lambda_0 \) and \( \tau = \varepsilon^2 \tau_0 \) for small \( \varepsilon > 0. \)

- (Chapman-Enskog expansion) flow can be written as 
a small perturbation: 
\[ f_\pm = f^{(0)}_\pm + \varepsilon f^{(1)}_\pm + \varepsilon^2 f^{(2)}_\pm + \ldots \]
where 
\[ f^{(0)}_+ + f^{(0)}_- = \rho. \]

Substitute these
A 1-D Transport Model

Seek expression describing evolution of site density, 
\[ \rho(x, t) = f_+(x, t) + f_-(x, t), \] 
as both \( \lambda, \tau \to 0 \).

Assumptions:

- \( \tau \to 0 \) faster than \( \lambda \to 0 \).
  Write \( \lambda = \epsilon \lambda_0 \) and \( \tau = \epsilon^2 \tau_0 \) for small \( \epsilon > 0 \).

- (Chapman-Enskog expansion) flow can be written as a small perturbation:
  \[ f_\pm = f_\pm^{(0)} + \epsilon f_\pm^{(1)} + \epsilon^2 f_\pm^{(2)} + \ldots \]
  where \( f_\pm^{(0)} + f_-^{(0)} = \rho \).

Substitute these and this into (1).
A 1-D Transport Model

Seek expression describing evolution of site density, \( \rho(x, t) = f_+(x, t) + f_-(x, t) \), as both \( \lambda, \tau \to 0 \).

Assumptions:

- \( \tau \to 0 \) faster than \( \lambda \to 0 \).
  Write \( \lambda = \varepsilon \lambda_0 \) and \( \tau = \varepsilon^2 \tau_0 \) for small \( \varepsilon > 0 \).
- (Chapman-Enskog expansion) flow can be written as a small perturbation: 
  \( f_\pm = f_\pm^{(0)} + \varepsilon f_\pm^{(1)} + \varepsilon^2 f_\pm^{(2)} + \ldots \)

where \( f_\pm^{(0)} + f_-^{(0)} = \rho \).

Substitute these and this into (1) and equate coefficients of like powers of \( \varepsilon \ldots \).
A 1-D Transport Model

Conclude: if $\rho(x,t) = f_+(x,t) + f_-(x,t)$ then
Conclude: if \( \rho(x, t) = f_+(x, t) + f_-(x, t) \) then

\[
\begin{pmatrix}
    f_+(x + \lambda, t + \tau) \\
    f_-(x - \lambda, t + \tau)
\end{pmatrix}
= \begin{pmatrix}
    \sigma & 1 - \sigma \\
    1 - \sigma & \sigma
\end{pmatrix}
\begin{pmatrix}
    f_+(x, t) \\
    f_-(x, t)
\end{pmatrix}
\]
A 1-D Transport Model

Conclude: if $\rho(x, t) = f_+(x, t) + f_-(x, t)$ then

$$
\begin{pmatrix}
  f_+(x + \lambda, t + \tau) \\
  f_-(x - \lambda, t + \tau)
\end{pmatrix} =
\begin{pmatrix}
  \sigma & 1 - \sigma \\
  1 - \sigma & \sigma
\end{pmatrix}
\begin{pmatrix}
  f_+(x, t) \\
  f_-(x, t)
\end{pmatrix}
$$

$$
\Rightarrow 
\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2}
$$
A 1-D Transport Model

Conclude: if $\rho(x,t) = f_+(x,t) + f_-(x,t)$ then

$$
\begin{pmatrix}
  f_+(x+\lambda, t+\tau) \\
  f_-(x-\lambda, t+\tau)
\end{pmatrix} =
\begin{pmatrix}
  \sigma & 1-\sigma \\
  1-\sigma & \sigma
\end{pmatrix}
\begin{pmatrix}
  f_+(x,t) \\
  f_-(x,t)
\end{pmatrix}
$$

$$
\Rightarrow \quad \frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2}
$$

where

$$
D = \left( \begin{array}{cc}
\frac{\lambda^2}{\tau} \\
\frac{\sigma}{2-2\sigma}
\end{array} \right)
$$
Conclude: if $\rho(x,t) = f_+(x,t) + f_-(x,t)$ then

$$
\begin{pmatrix}
  f_+(x+\lambda, t+\tau) \\
  f_-(x-\lambda, t+\tau)
\end{pmatrix} =
\begin{pmatrix}
  \sigma & 1-\sigma \\
  1-\sigma & \sigma
\end{pmatrix}
\begin{pmatrix}
  f_+(x,t) \\
  f_-(x,t)
\end{pmatrix}
$$

$$
\Rightarrow \quad \frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2}
$$

where

$$
D = \left( \frac{\lambda^2}{\tau} \right) \left( \frac{\sigma}{2-2\sigma} \right)
$$

Method is not restricted to diffusion processes or to 1-D!
Higher Dimensions

- 2D: The wave equation:

\[ \frac{\partial^2 h(r,t)}{\partial t^2} = c^2 \nabla^2 h(r,t) \]

- Model (and visualize) ocean waves.

- 3D: Navier-Stokes:

\[ \frac{\partial u}{\partial t} + u \cdot \nabla u = -(1/\rho) \nabla p - \nu \nabla^2 (u) \]

- Model (and visualize) air flow over simple geometries.
Visualization of Results

- 1-D: gnuplot
- 2-D: OpenGL basics covered
- 3-D: Paraview or VisIt; write in .vtk format
Animations?
Conclusions

- GPU Computing at Clemson taught since 2010
- Increasing importance, increasing enrollment
- τέχνη method; PBL
- Single problem: system for parallel solution of PDEs
- Serves as both real-world problem and vehicle for GPU topic exploration
Thanks!

- NSF Award CNS-1126344
- NSF Award IIS-1314757
- NVIDIA Corporation
  (David Luebke, Cliff Woolley, Chandra Cheij)