1.

Here is a link that can be of help.
http://en.wikipedia.org/wiki/List_of_Nvidia_graphics_processing_units

The table below captures some of the information from the above link, subject to the following constraints:
- We restrict our choices, as the question stated, to GeForce GT/GTX
- We restrict our choices to single die (i.e. single GPU in the chip). So entries that have “2x..” are not taken into account.
- We picked one GPU per year.
- If there are several entries for the same item (like memory size), we pick the largest number.

<table>
<thead>
<tr>
<th>Model</th>
<th>#cores</th>
<th>mem size (MB)</th>
<th>mem bw (GB/s)</th>
<th>year</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeForce GTX 285</td>
<td>240</td>
<td>2048</td>
<td>159.0</td>
<td>2009</td>
</tr>
<tr>
<td>GeForce GTX 480</td>
<td>480</td>
<td>1536</td>
<td>177.4</td>
<td>2010</td>
</tr>
<tr>
<td>GeForce GTX 560</td>
<td>336</td>
<td>2048</td>
<td>128.1</td>
<td>2011</td>
</tr>
<tr>
<td>GeForce GTX 680</td>
<td>1536</td>
<td>4096</td>
<td>192.256</td>
<td>2012</td>
</tr>
<tr>
<td>GeForce GTX Titan</td>
<td>2688</td>
<td>6144</td>
<td>288.4</td>
<td>2013</td>
</tr>
<tr>
<td>GeForce GTX Titan Black</td>
<td>2880</td>
<td>6144</td>
<td>336.4</td>
<td>2014</td>
</tr>
</tbody>
</table>

2.
Number of cores is steadily increasing. However, the amount of memory does not increase with the same rate, same with the memory bandwidth. This means that the main bottleneck will be the memory access. It is currently so and expected to continue.

3.
- n-body problems
- matrix operations
- neural-networks
- genetic algorithms
• the Black-Scholes (mathematical model for financial market)

The main characteristics of the above examples are:

- It involves consuming large amount of data
- Each application can be divided to similar procedures working on different data.
- No very severe dependencies among the procedures

4.

- operating system kernel
- generating Fibonacci numbers
- editor
- circuit simulator (like SPICE)
- device drivers

All the above applications have the following characteristics:

- The different operations have very high dependencies
- Not many repeated operations

5. To be a GPU friendly, an application needs to have the following characteristics:

- A lot of data to be worked on
- Repeated operations
- Low dependence among operations