

LPeg Primer for AWK People

Hartmut Henkel, Oftersheim, Germany
N 49°22.275' E 008°35.553'

2nd International ConTEXt Conference, Bohinj
N 46°16.655' E 13°51.771'

22 August 2008

LPEG Primer for *really* AWK People

Hartmut Henkel, Oftersheim, Germany

N 49°22.275' E 008°35.553'

2nd International ConTEXt Conference, Bohinj

N 46°16.655' E 13°51.771'

22 August 2008

Outline

You will learn about. . .

- ▶ LPEG — The Tool
- ▶ Basics of the *match* and *capture*
- ▶ Practical demonstration

The Tool

... what LPEG means.

The Tool

... what LPEG means.

Low

The Tool

... what LPEG means.

Low **P**ressure

The Tool

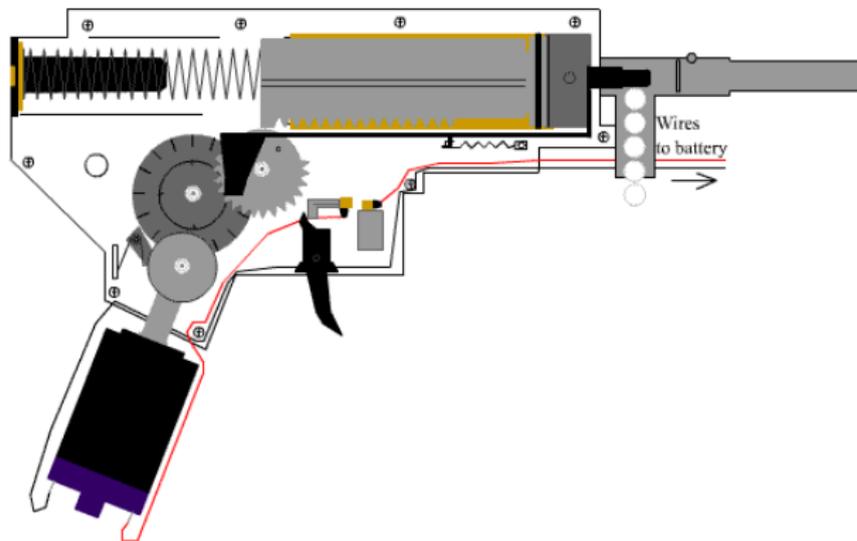
... what LPEG means.

Low **P**ressure **E**lectric

The Tool

... what LPEG means.

Low Pressure Electric Gun

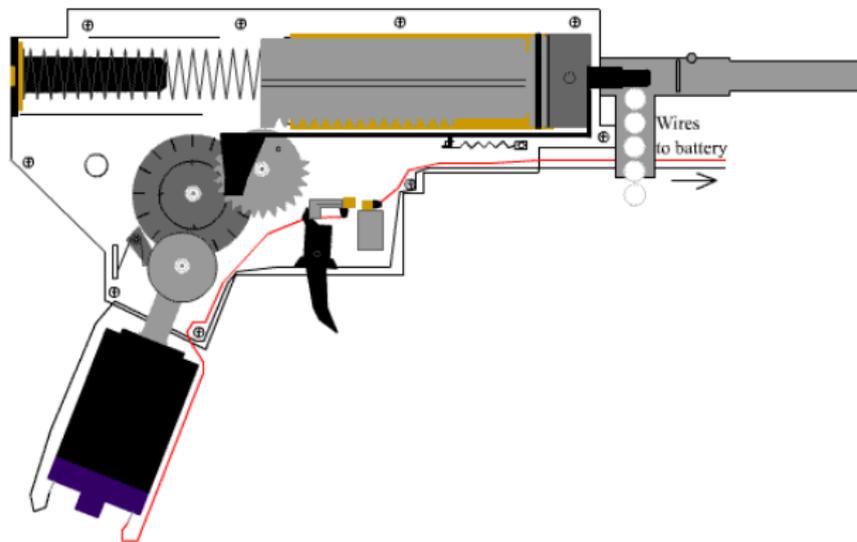


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

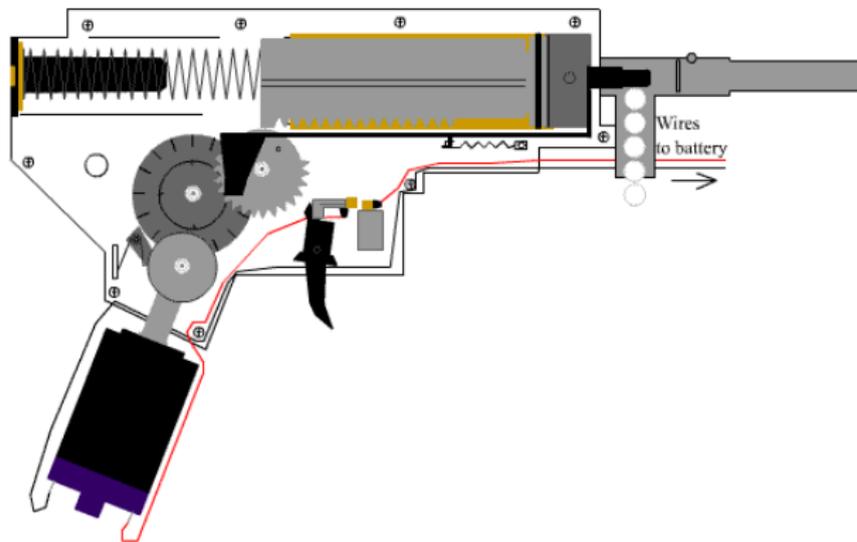


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

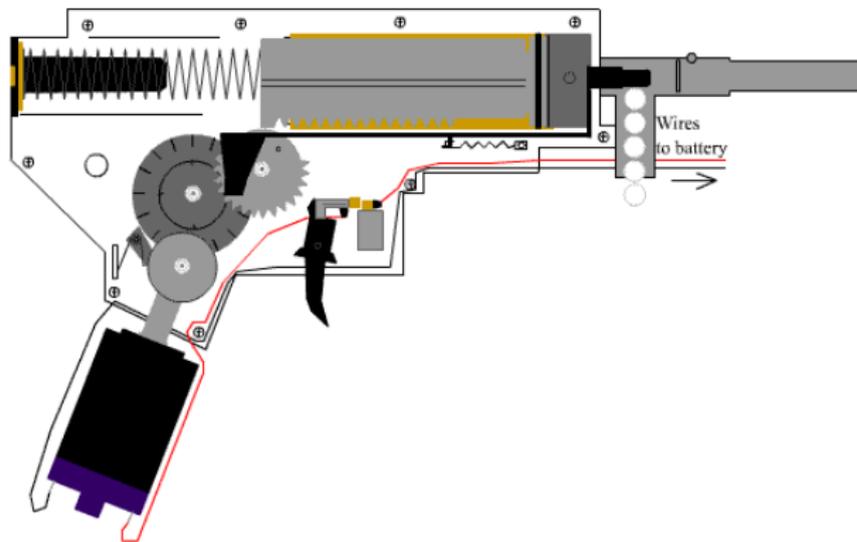


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

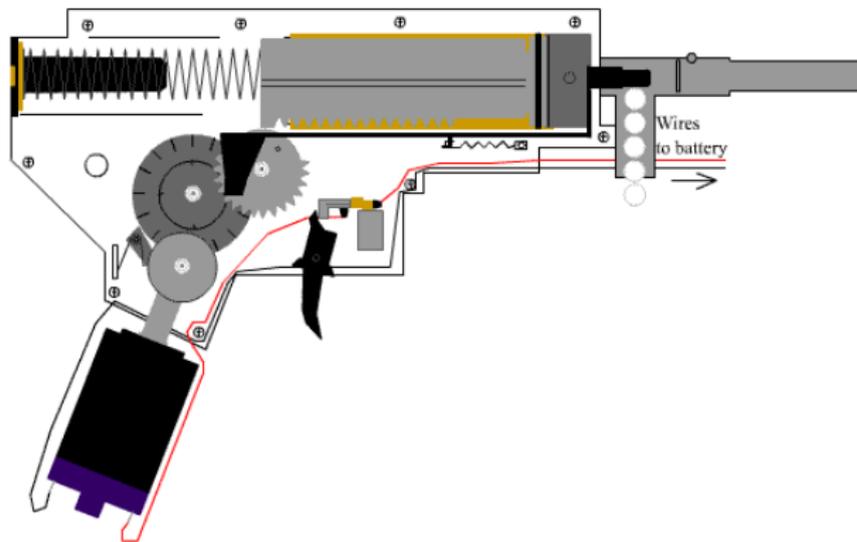


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

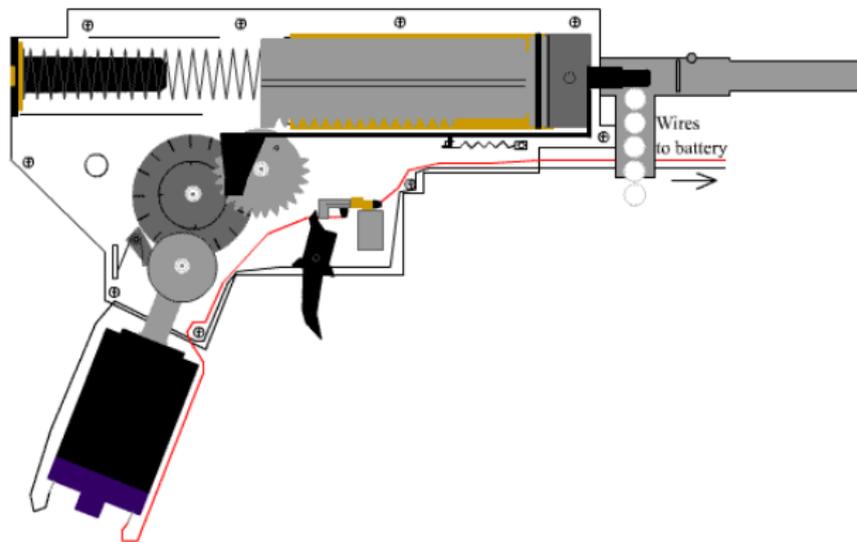


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

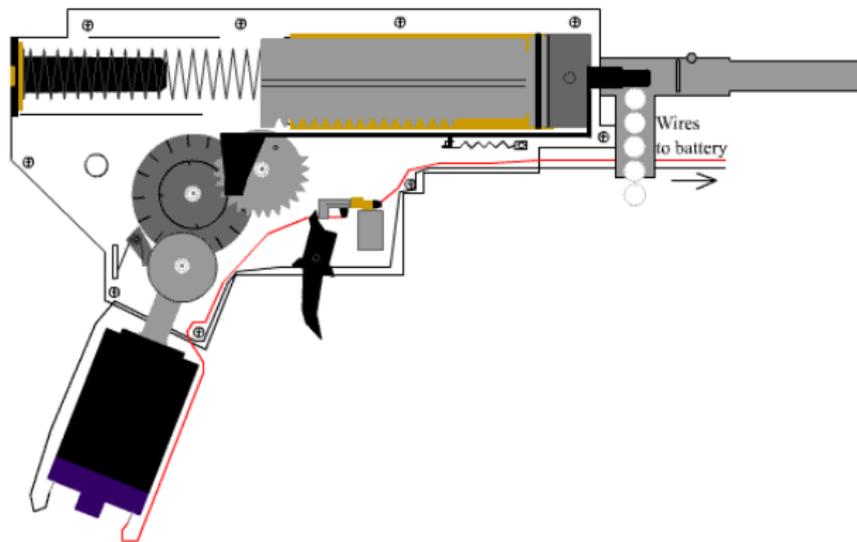


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

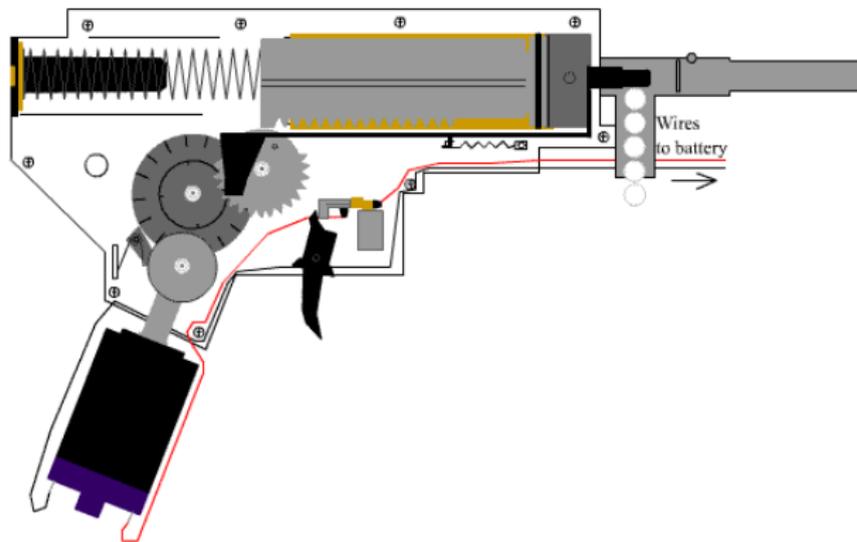


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

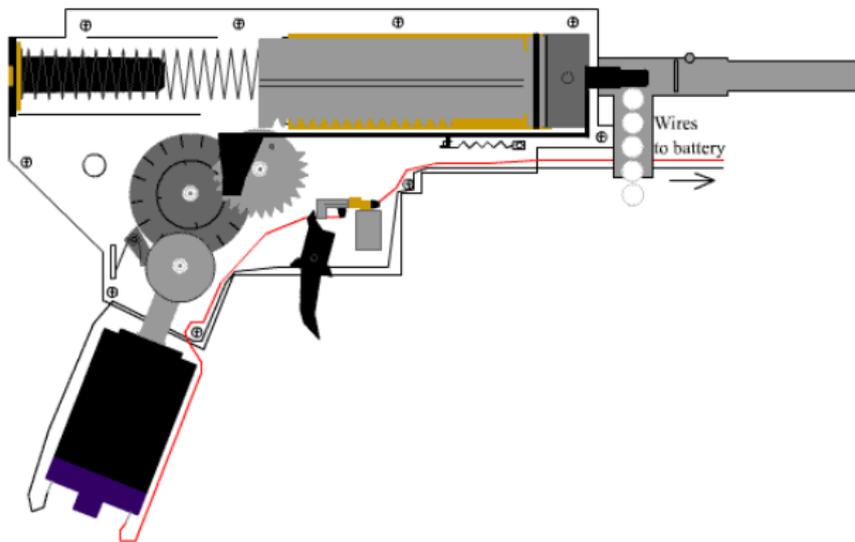


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

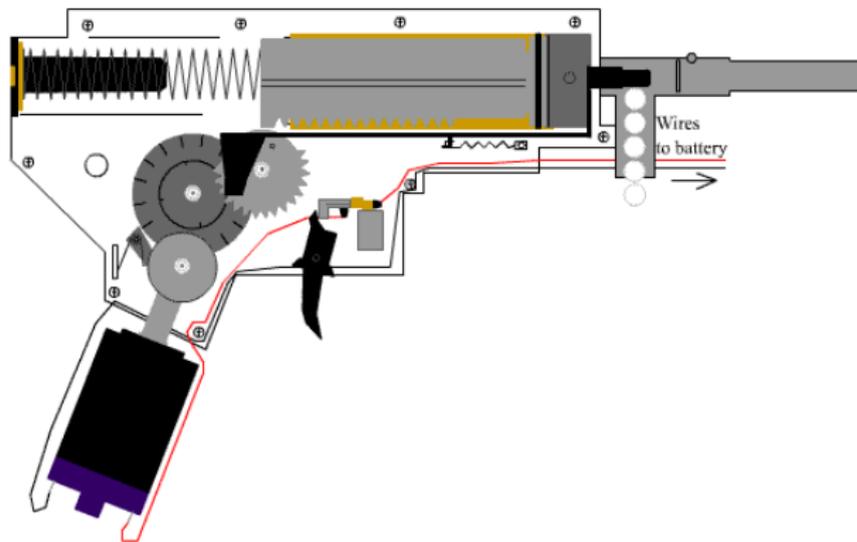


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

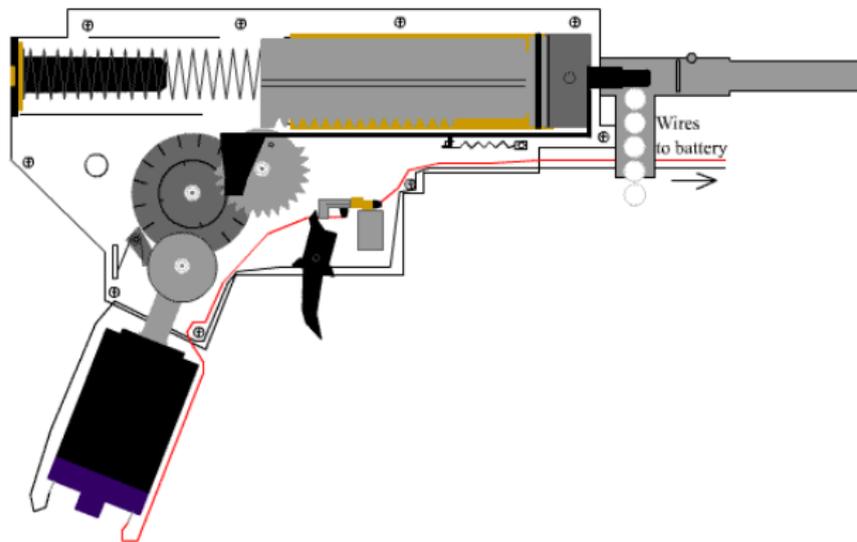


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

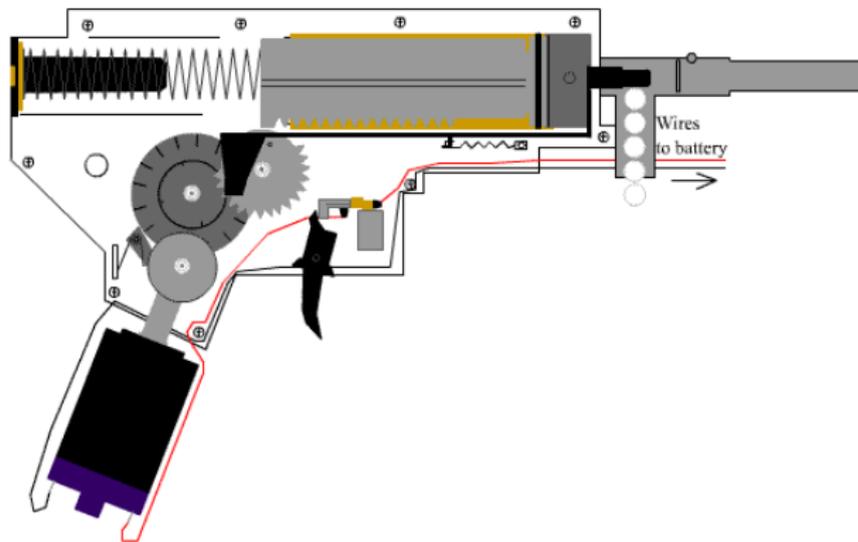


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

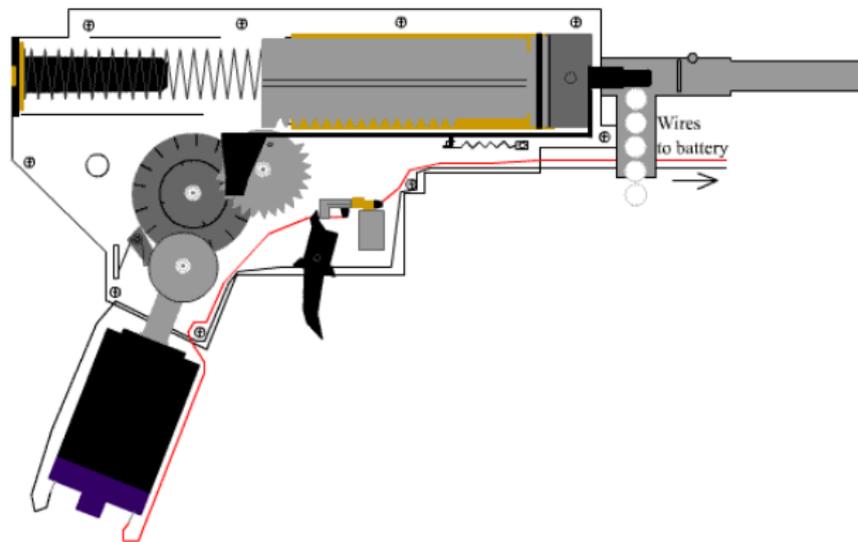


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

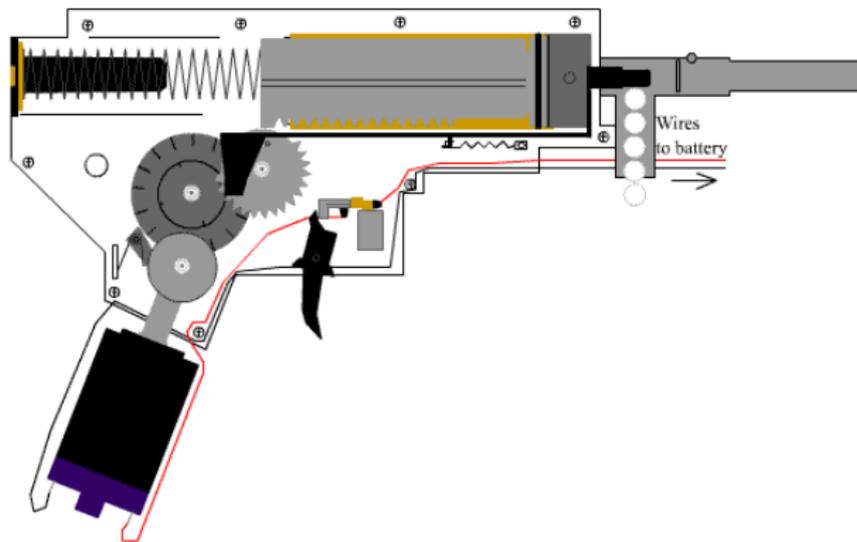


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

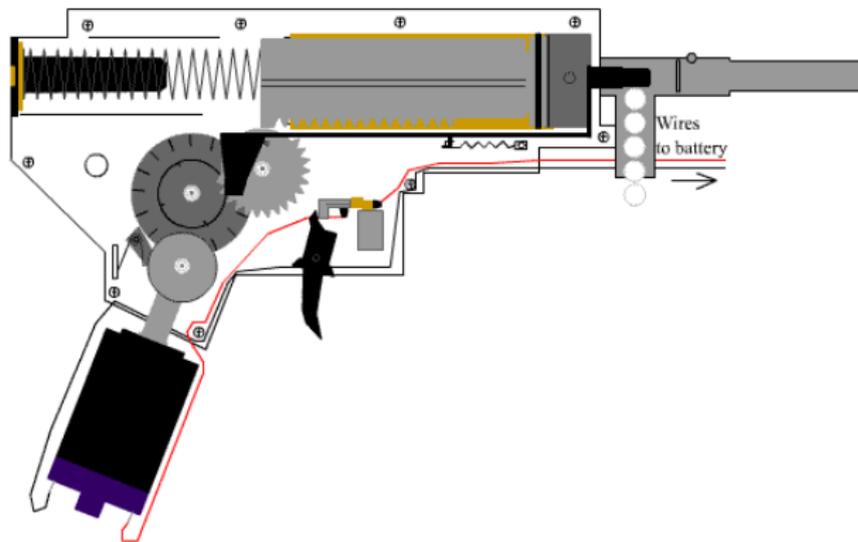


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

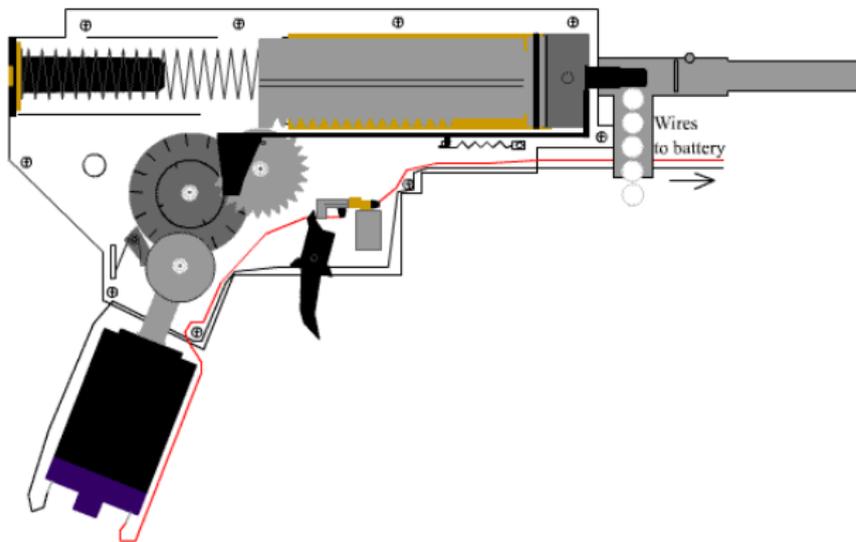


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

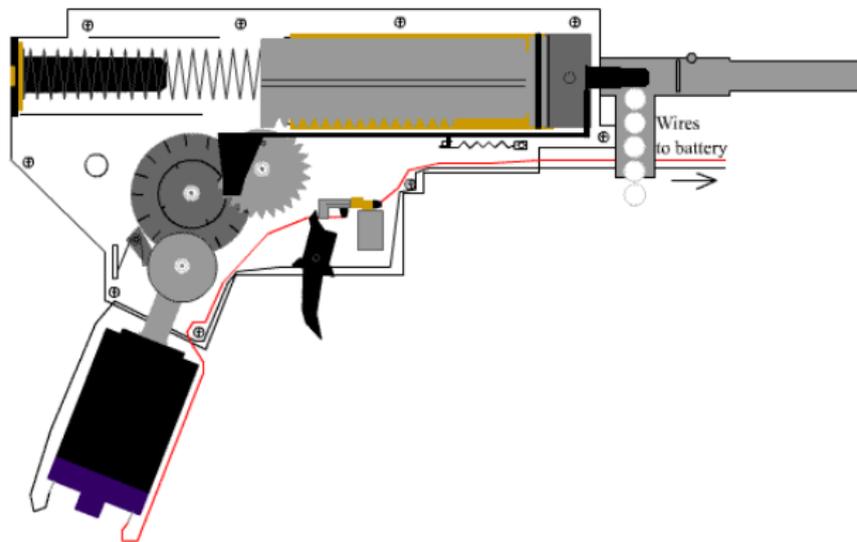


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

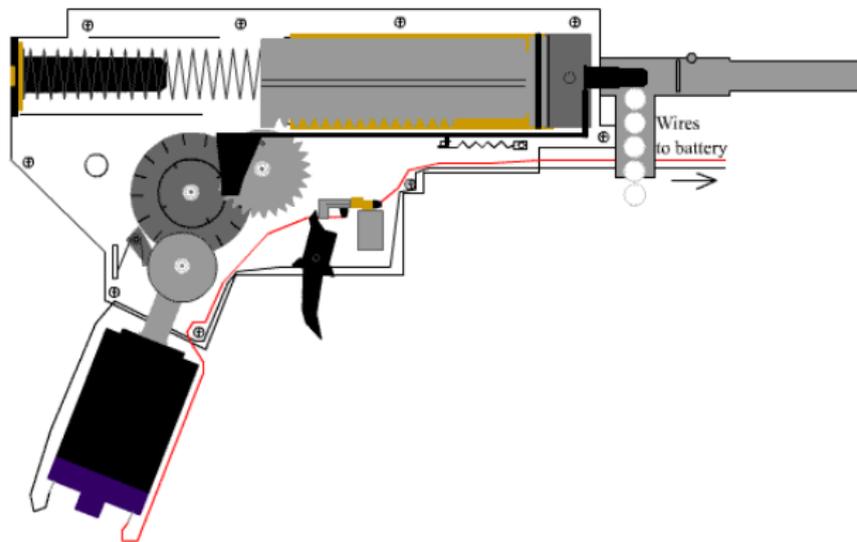


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

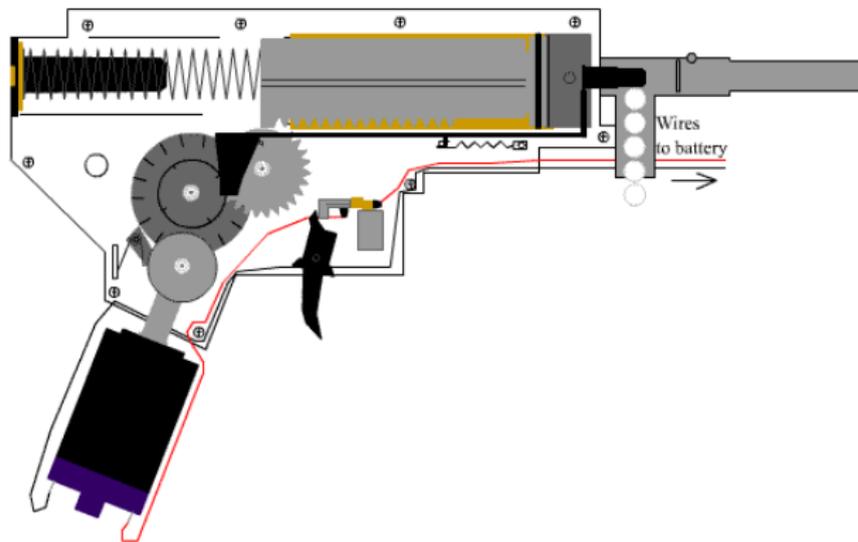


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

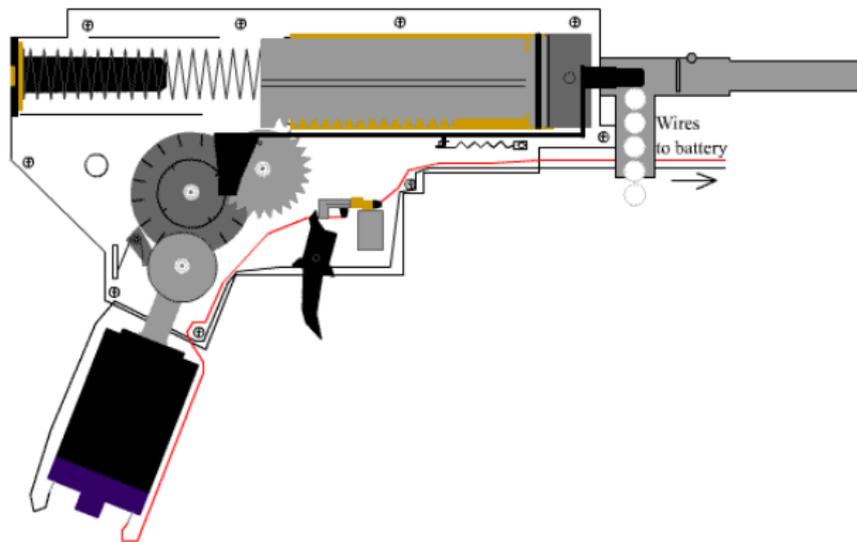


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

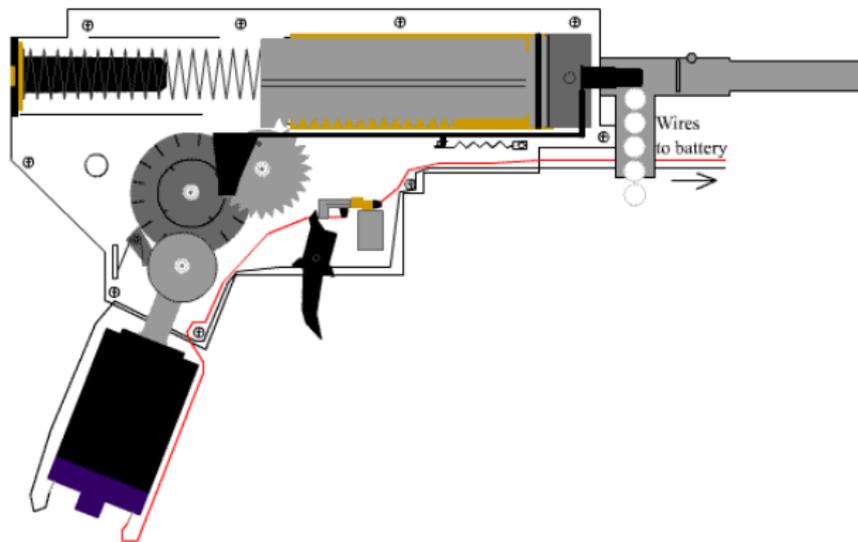


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

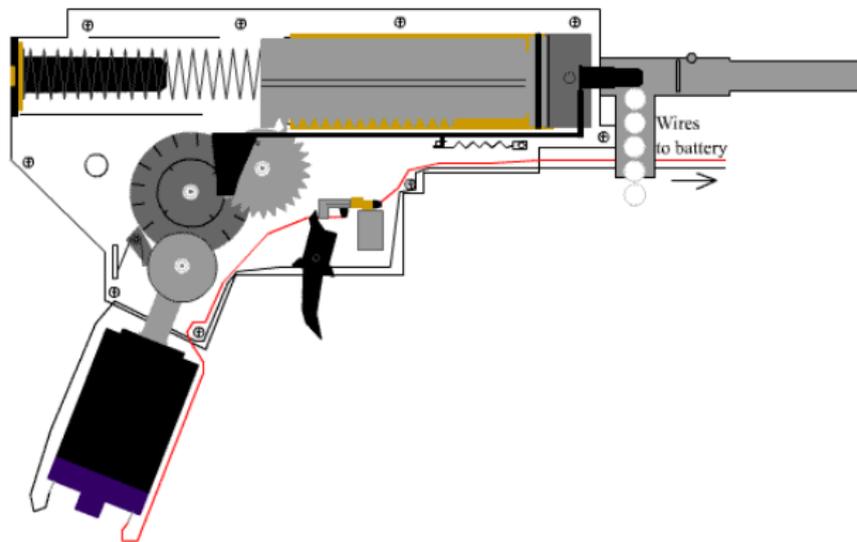


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

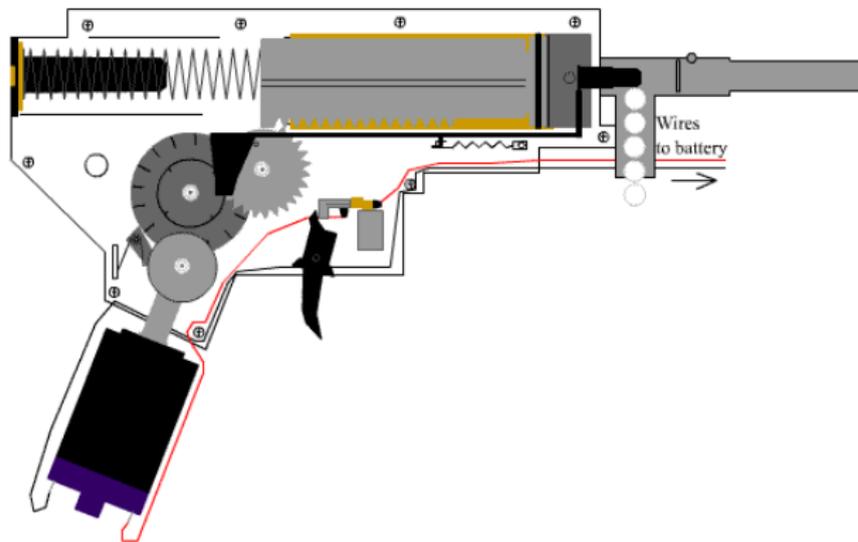


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

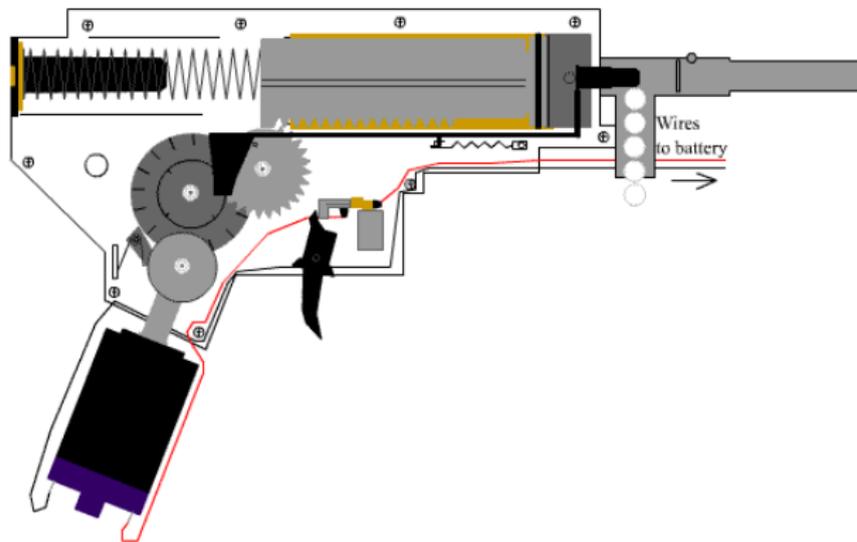


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

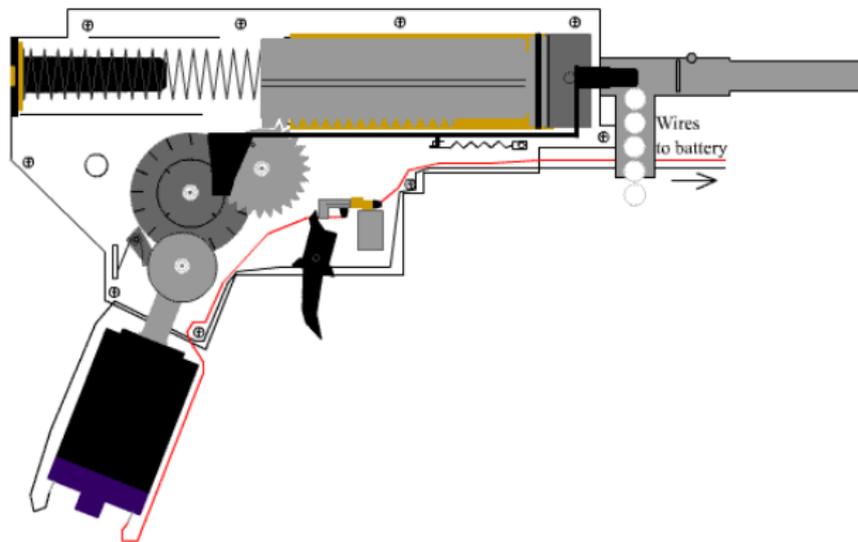


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

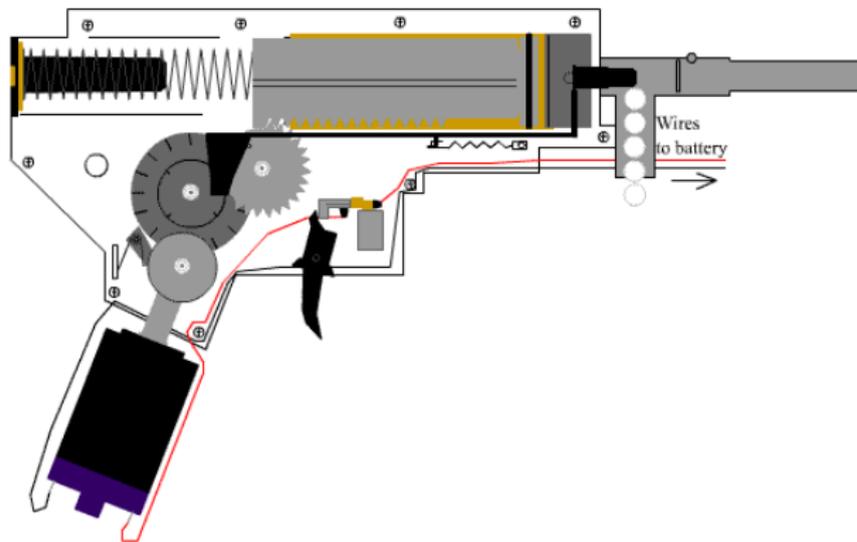


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

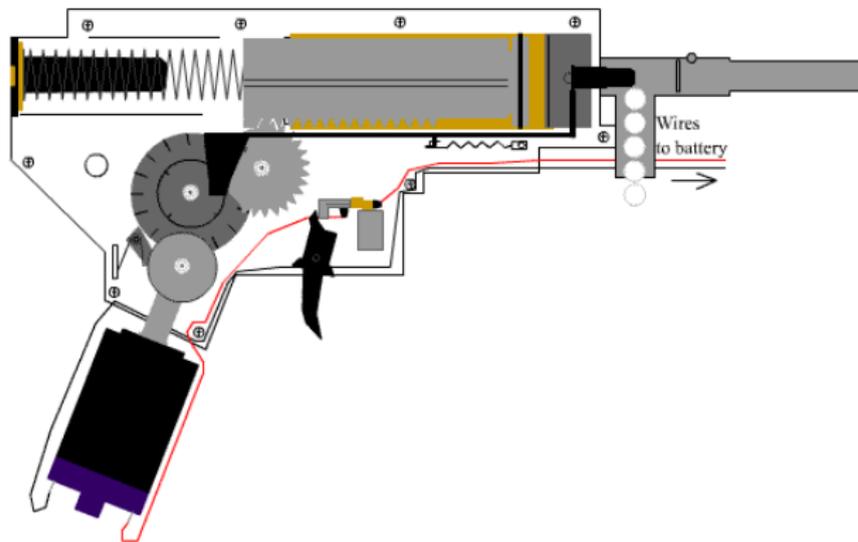


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

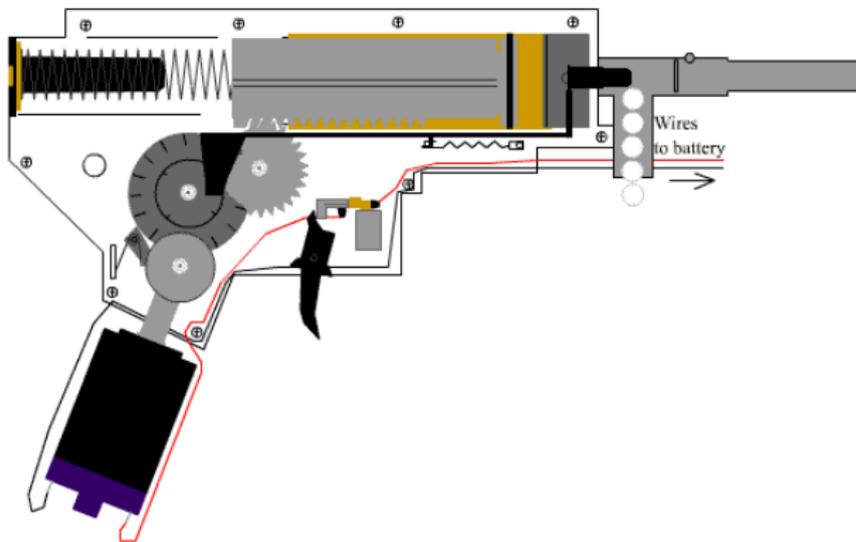


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

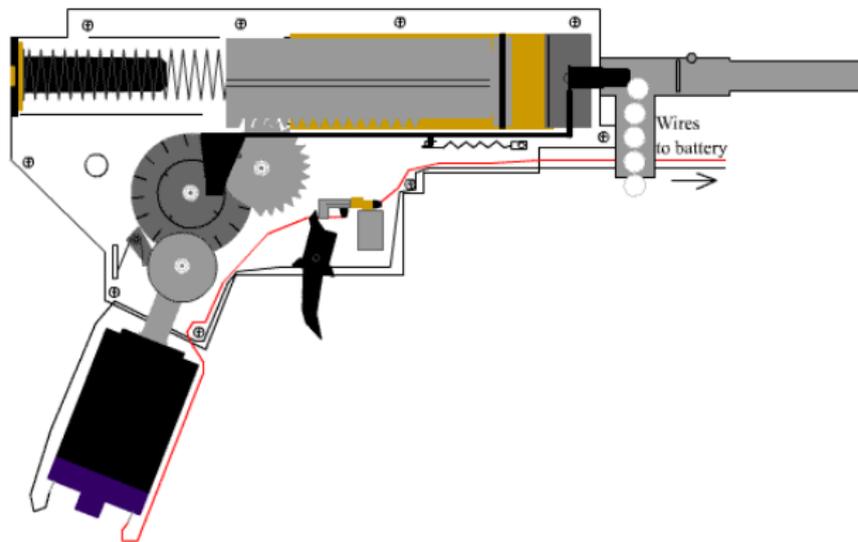


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

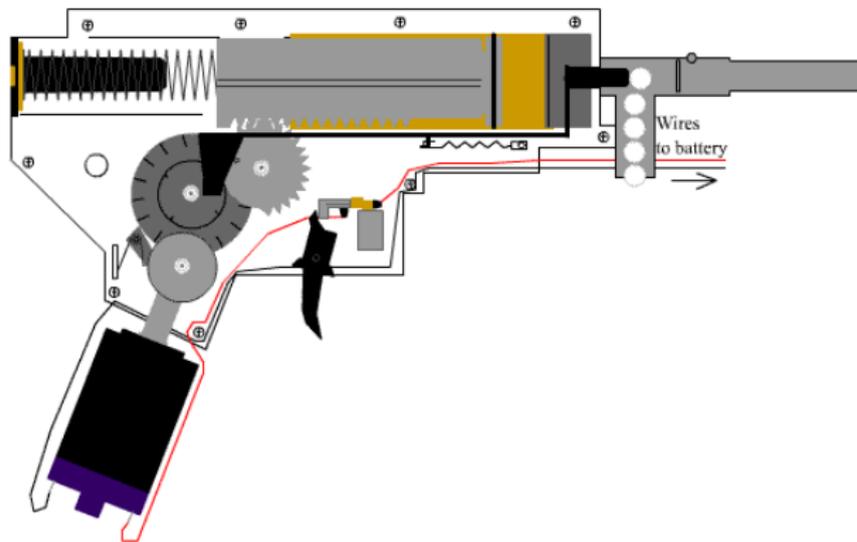


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

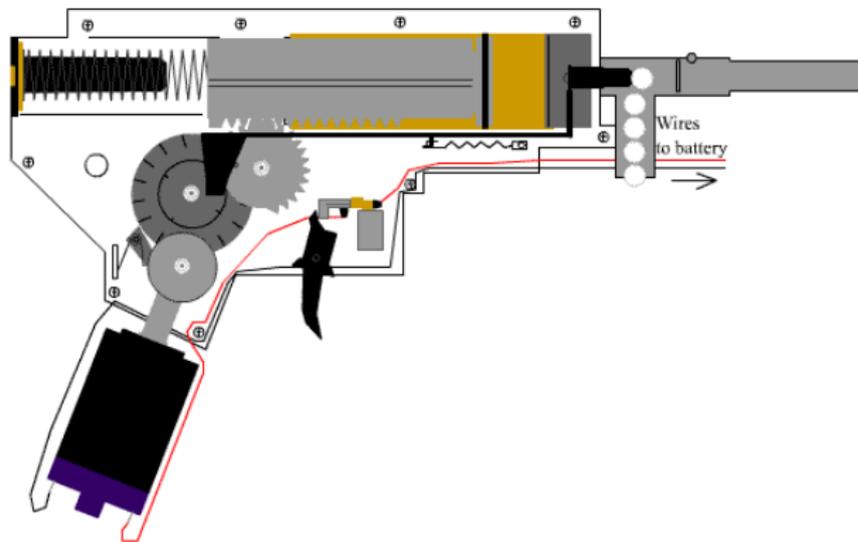


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

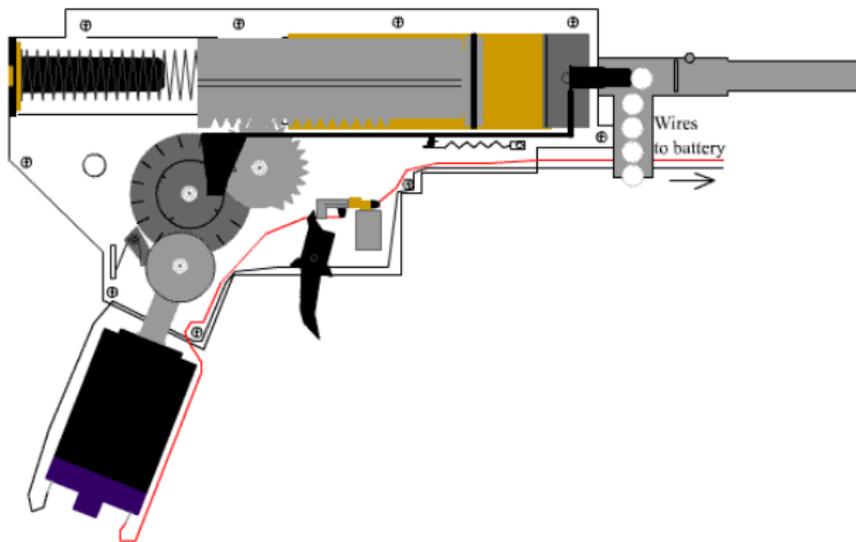


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

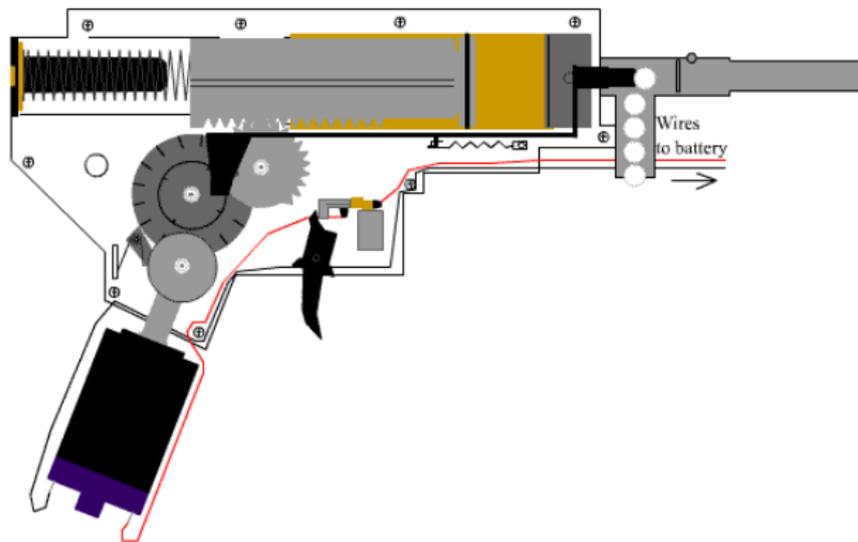


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

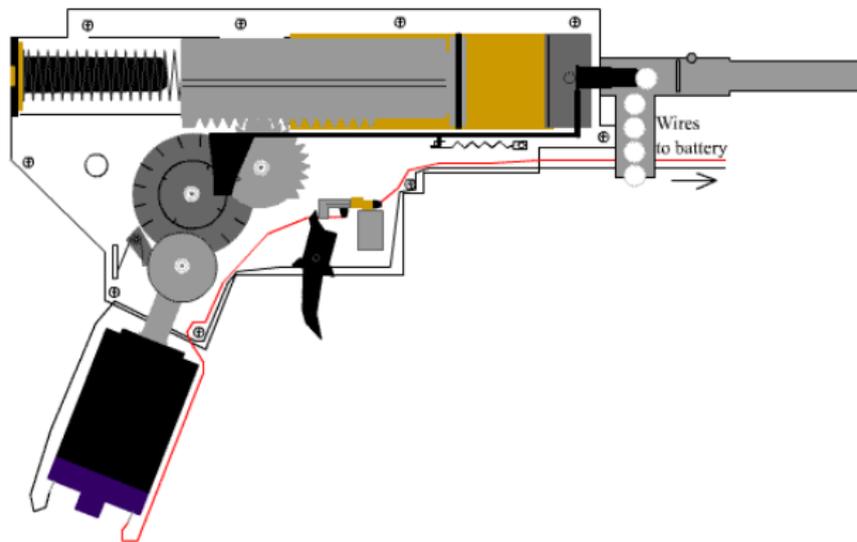


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

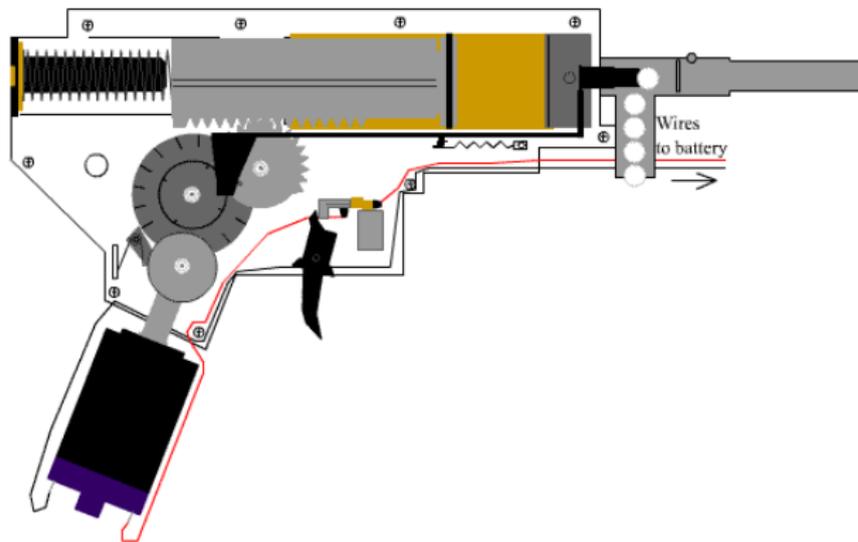


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

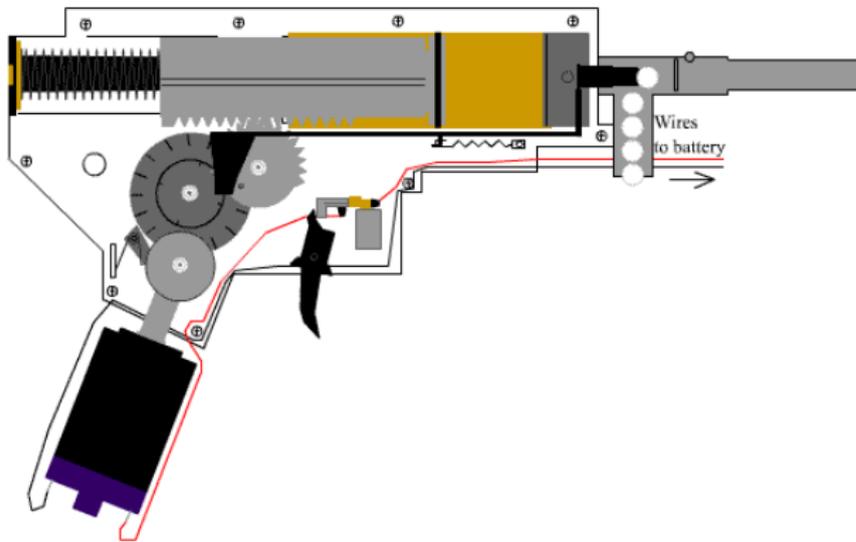


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

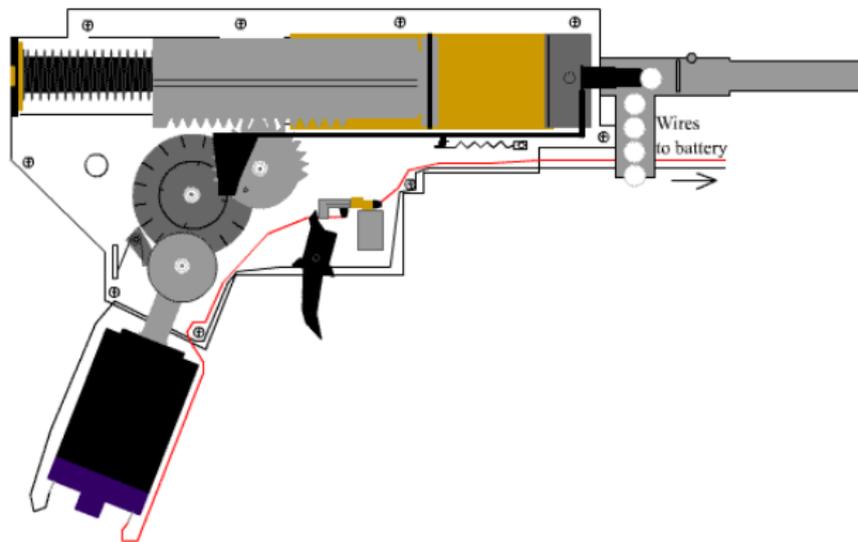


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

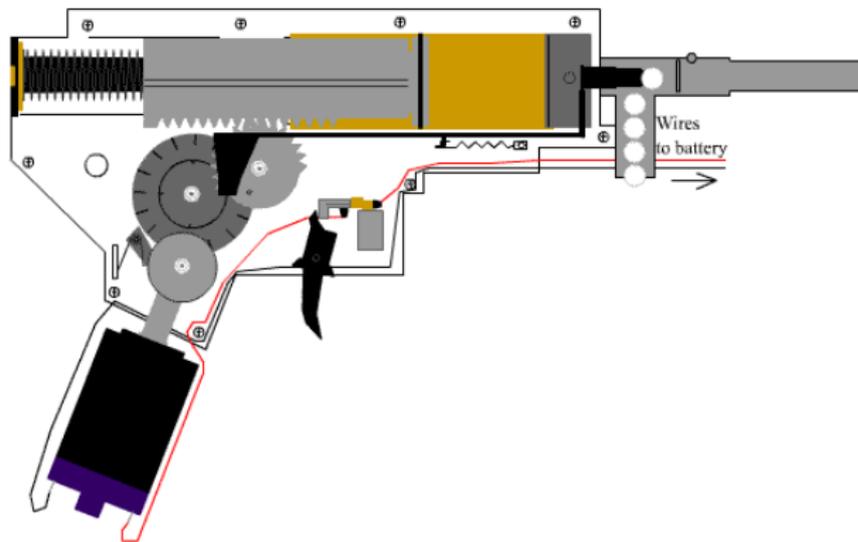


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

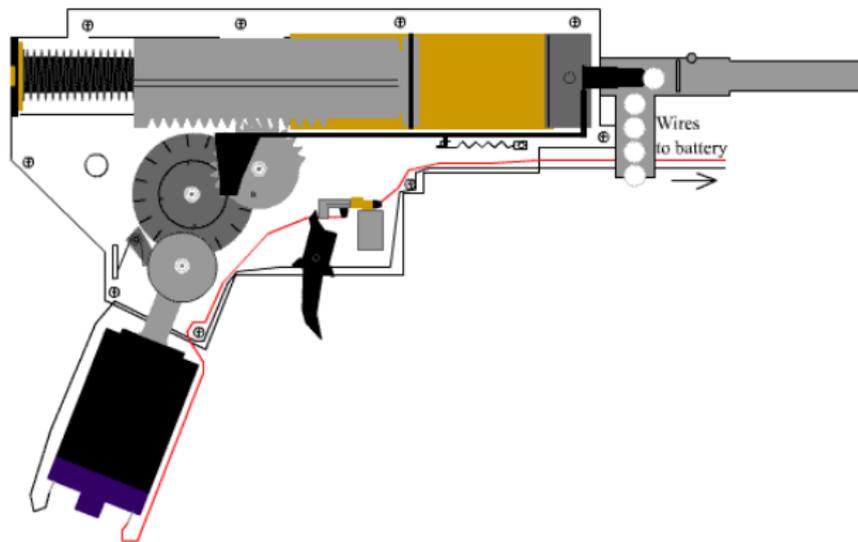


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

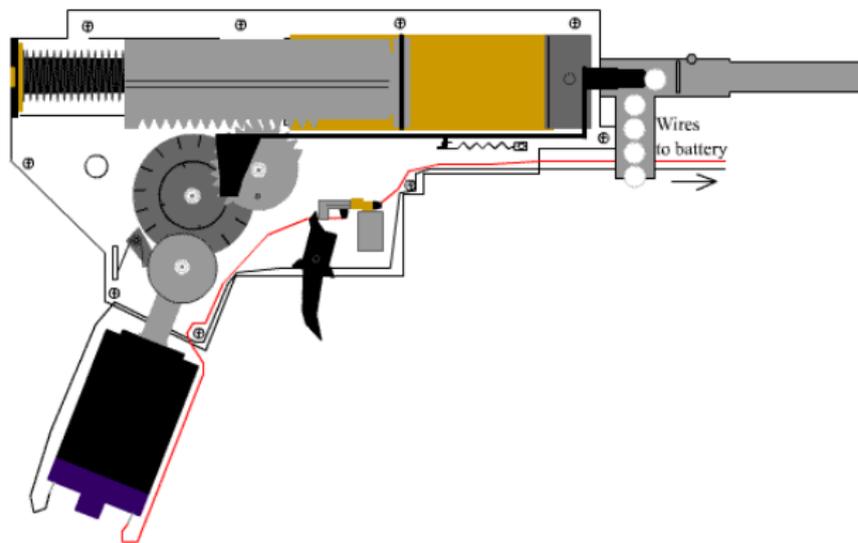


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

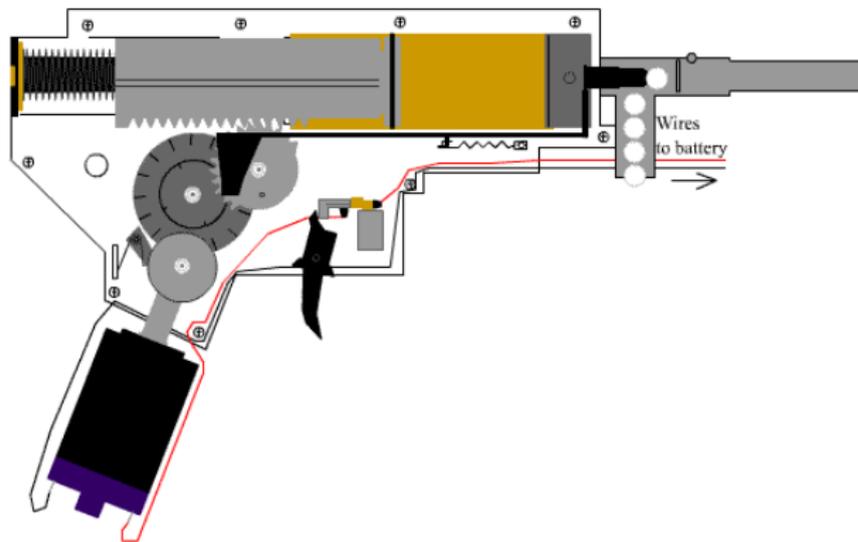


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

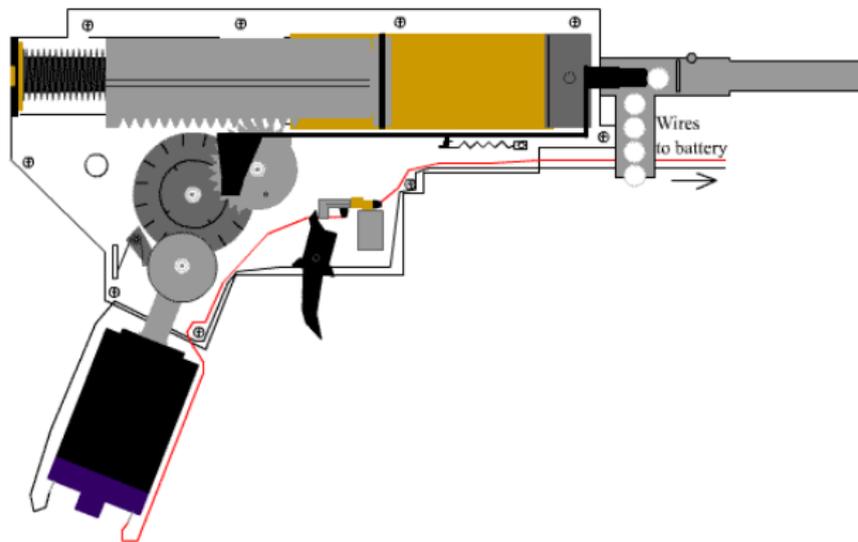


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

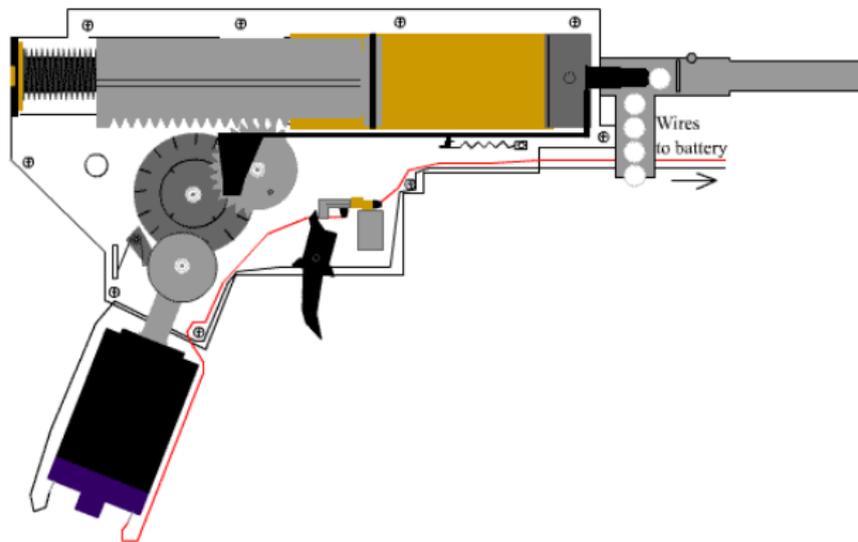


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

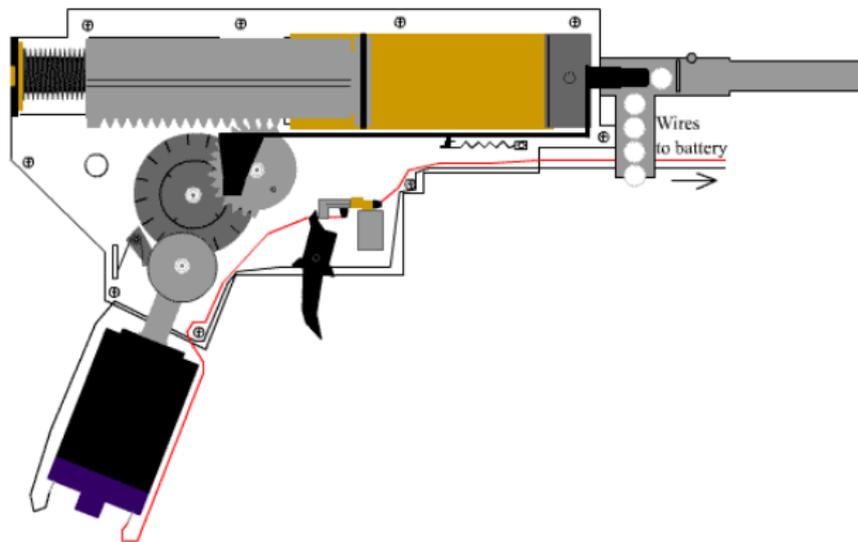


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

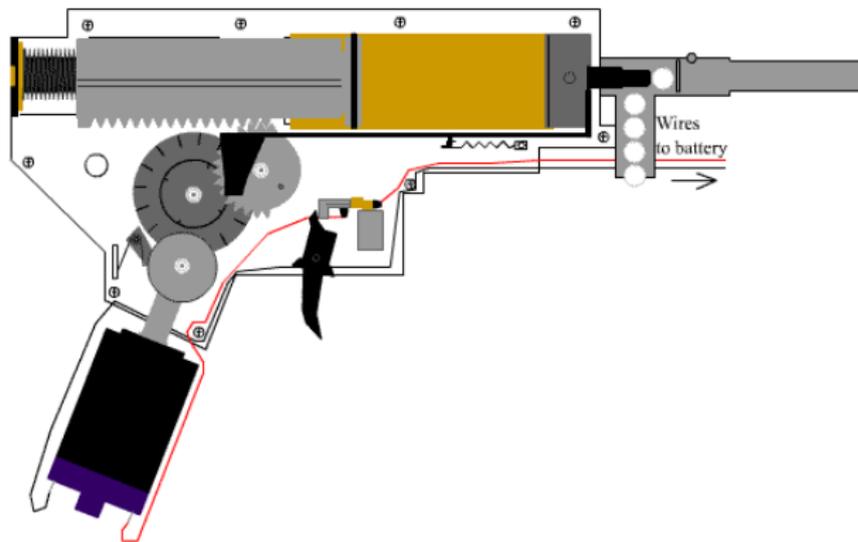


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

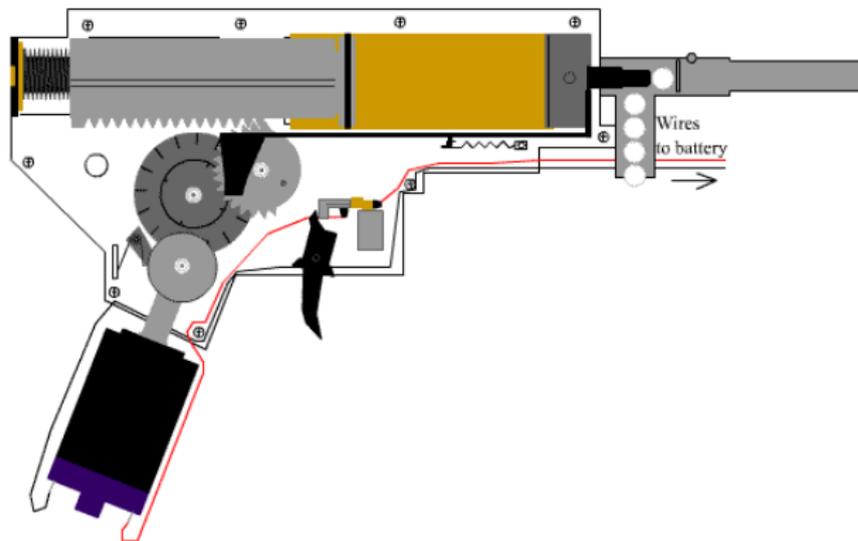


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

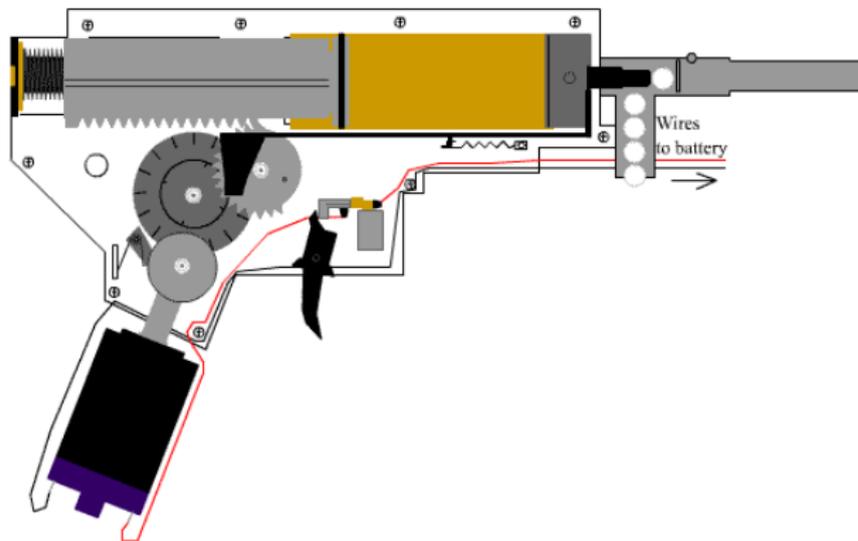


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

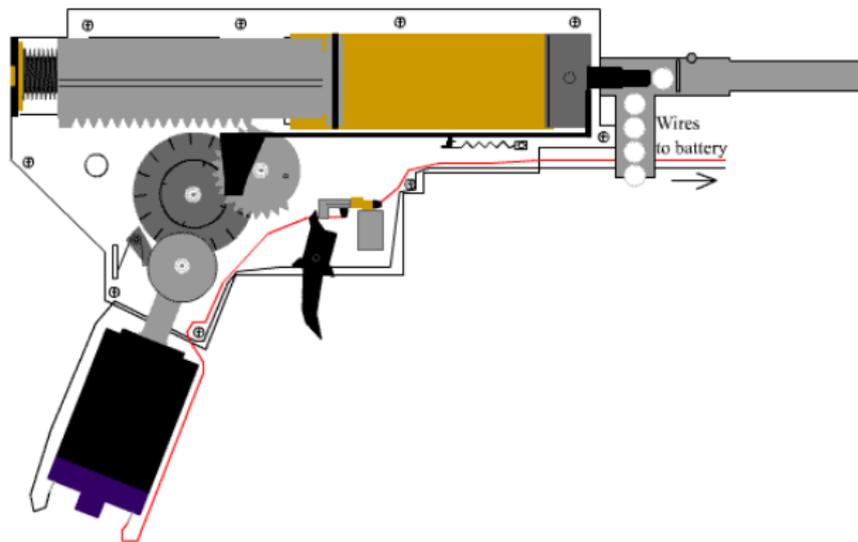


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

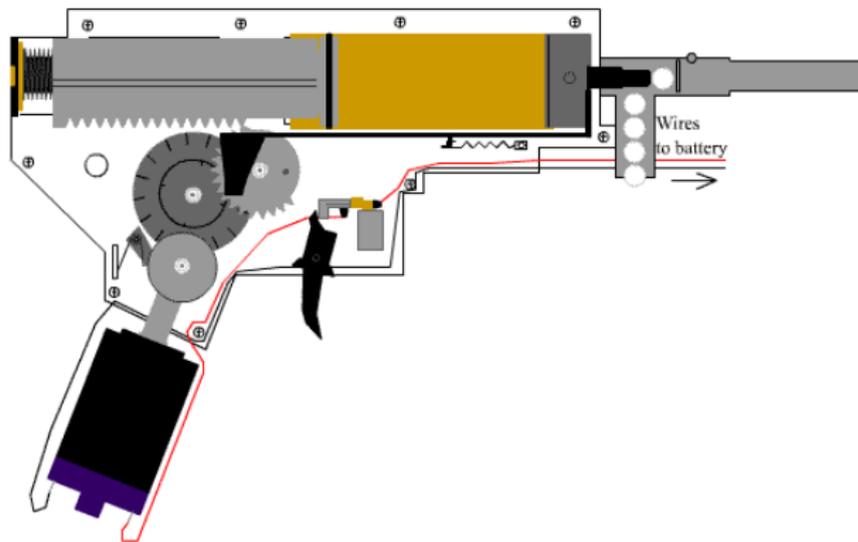


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

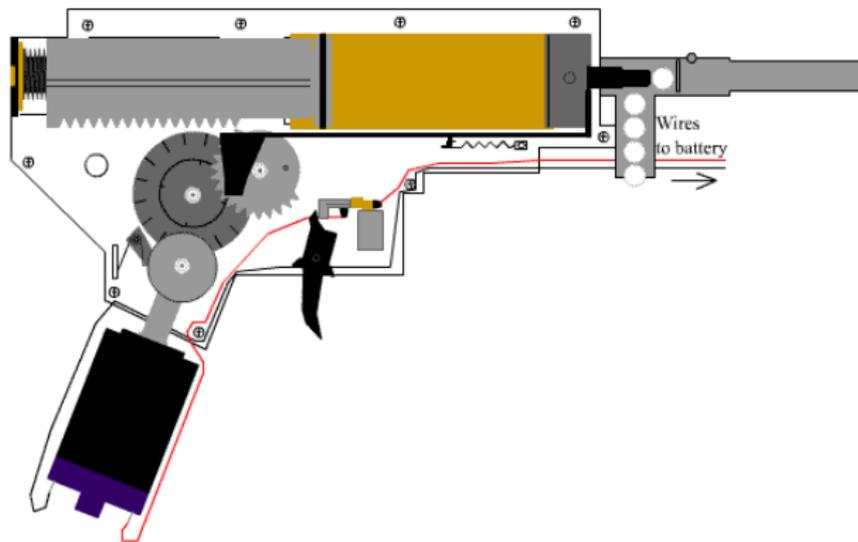


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

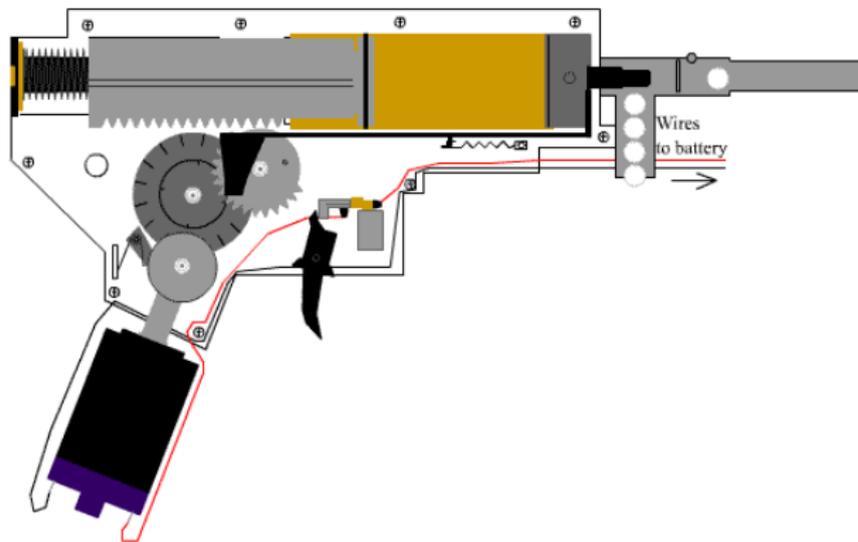


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

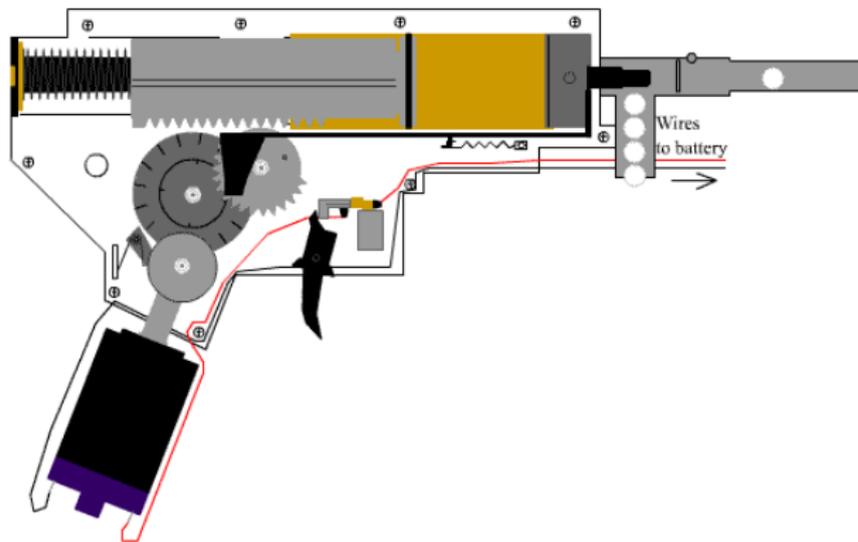


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

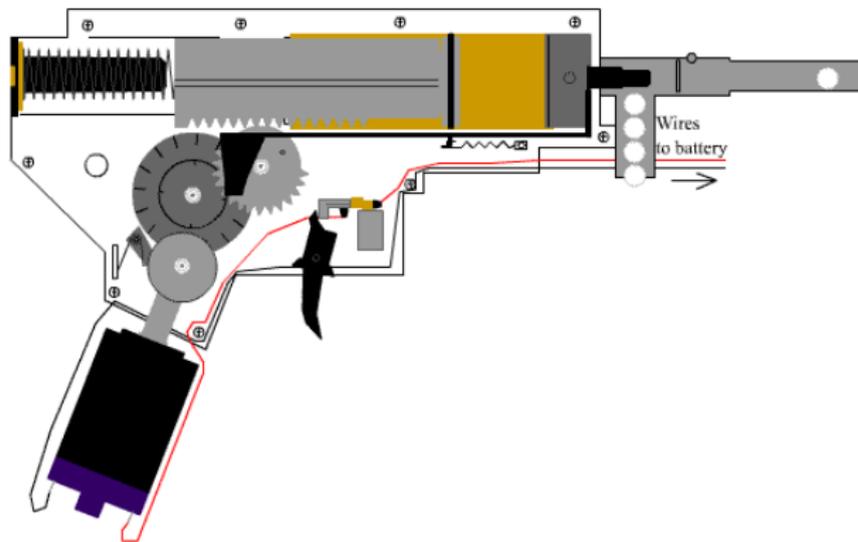


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

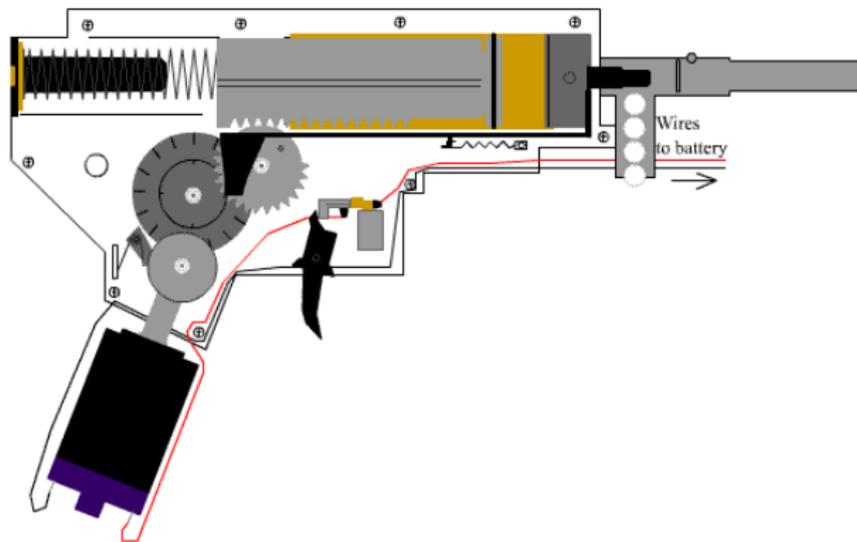


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!

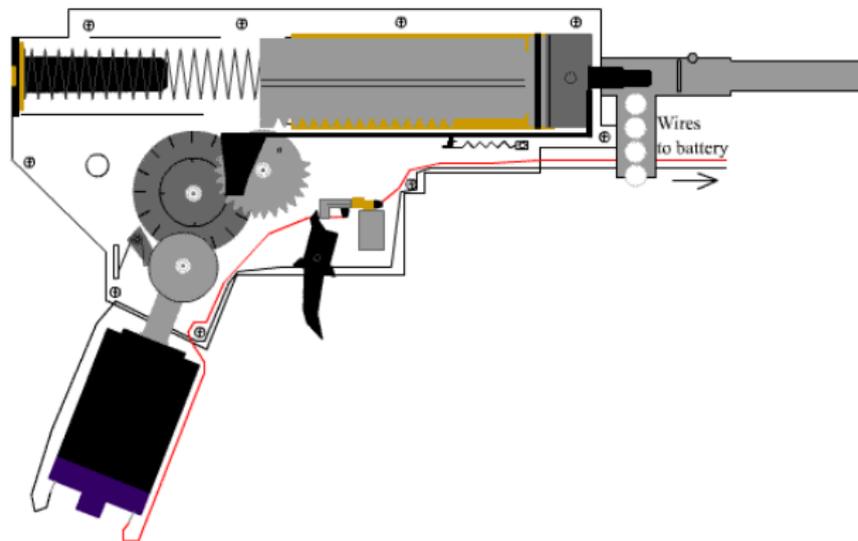


Images: Wikipedia (public domain)

The Tool

... what LPEG means.

Low **P**ressure **E**lectric **G**un — it's a softair tool!



Images: Wikipedia (public domain)

Ooops — wrong talk!

Peace!



LPeg Primer for AWK People

Hartmut Henkel, Oftersheim, Germany
N 49°22.275' E 008°35.553'

2nd International ConTEXt Conference, Bohinj
N 46°16.655' E 13°51.771'

22 August 2008

Outline

You will learn about...

LPeg — a pattern matching library for Lua
— written by Roberto Ierusalimsky, PUC Rio

- ▶ What's a PEG?
- ▶ Differences between (AWK) regular expressions and LPeg
- ▶ Simple workflow setup:
 - ▶ Loading the LPeg library
 - ▶ Reading and writing data — using Lua
 - ▶ LPeg key concepts: *match* and *capture*
 - ▶ LPeg's appetite & other beginner gotchas
- ▶ More on the power of *capture*

...with practical demonstrations.

Outline

Why this talk?

Why “LPeg primer for AWK people”?

What's LPeg?

What's a PEG?

Wikipedia:

*A **p**arsing **e**xpression **g**rammar, or PEG, is a type of analytic **f**ormal **g**rammar that describes a **f**ormal **l**anguage in terms of a set of rules for recognizing strings in the language.*

dict analytic:

...resolving into elements or constituent parts; ...

What's LPeg?

What's a PEG?

Wikipedia:

*A **p**arsing **e**xpression **g**rammar, or PEG, is a type of analytic **f**ormal **g**rammar that describes a **f**ormal **l**anguage in terms of a set of rules for recognizing strings in the language.*

dict analytic:

...resolving into elements or constituent parts; ...

LPeg is a PEG implementation (library) for Lua.

What's LPeg?

Notable differences: regular expressions vs. LPeg

- ▶ LPeg works in “**anchored mode**” — *the* initial hurdle

What's LPeg?

Notable differences: regular expressions vs. LPeg

- ▶ LPeg works in “**anchored mode**” — *the* initial hurdle
- ▶ LPeg patterns can be constructed — the great advantage

What's LPeg?

Notable differences: regular expressions vs. LPeg

- ▶ LPeg works in “**anchored mode**” — *the* initial hurdle
- ▶ LPeg patterns can be constructed — the great advantage
- ▶ LPeg has no notion of a record/line — multi-line blocks easy

What's LPeg?

Notable differences: regular expressions vs. LPeg

- ▶ LPeg works in “**anchored mode**” — *the* initial hurdle
- ▶ LPeg patterns can be constructed — the great advantage
- ▶ LPeg has no notion of a record/line — multi-line blocks easy
- ▶ LPeg is 8 bit clean — can handle arbitrary binary data

What's LPeg?

Notable differences: regular expressions vs. LPeg

- ▶ LPeg works in “**anchored mode**” — *the* initial hurdle
- ▶ LPeg patterns can be constructed — the great advantage
- ▶ LPeg has no notion of a record/line — multi-line blocks easy
- ▶ LPeg is 8 bit clean — can handle arbitrary binary data
- ▶ LPeg allows to formulate complex grammars — great power

Loading the LPeg library

Reading and writing data — using Lua

Simple workflow setup

LPeg key concepts: match

match

Simple workflow setup

LPeg key concepts: match

match

- ▶ Write up all needed terminals, e. g., keywords, fragments like space, lf, alpha, digit, backslash, leftbrace, rightbrace...

Simple workflow setup

LPeg key concepts: match

match

- ▶ Write up all needed terminals, e. g., keywords, fragments like `space`, `lf`, `alpha`, `digit`, `backslash`, `leftbrace`, `rightbrace`...
- ▶ Build your pattern (bottom-up, top-down)

Simple workflow setup

LPeg key concepts: match

match

- ▶ Write up all needed terminals, e. g., keywords, fragments like space, lf, alpha, digit, backslash, leftbrace, rightbrace...
- ▶ Build your pattern (bottom-up, top-down)
- ▶ Test your pattern with simple input

Simple workflow setup

LPeg key concepts: match

match

- ▶ Write up all needed terminals, e. g., keywords, fragments like space, lf, alpha, digit, backslash, leftbrace, rightbrace...
- ▶ Build your pattern (bottom-up, top-down)
- ▶ Test your pattern with simple input
- ▶ Add captures...

Simple workflow setup

Surprises, gotchas

Surprises, gotchas:

- ▶ LPeg works in *anchored mode!*

Does your pattern describe the input *from the beginning?*

Simple workflow setup

Surprises, gotchas

Surprises, gotchas:

- ▶ LPeg works in *anchored mode*!
Does your pattern describe the input *from the beginning*?
- ▶ Be sure that you really talk about patterns.
E. g., $1^0 \neq \text{lpeg.P}(1)^0$

Simple workflow setup

Surprises, gotchas

Surprises, gotchas:

- ▶ LPeg works in *anchored mode*!
Does your pattern describe the input *from the beginning*?
- ▶ Be sure that you really talk about patterns.
E. g., $1^0 \neq \text{lpeg.P}(1)^0$
- ▶ LPeg's greediness:
LPeg matches left to right the maximum possible string — even if the overall match would fail.
→ Write patterns that are safe against fall-through.

Simple workflow setup

Surprises, gotchas

Surprises, gotchas:

- ▶ LPeg works in *anchored mode*!
Does your pattern describe the input *from the beginning*?
- ▶ Be sure that you really talk about patterns.
E. g., $1^0 \neq \text{lpeg.P}(1)^0$
- ▶ LPeg's greediness:
LPeg matches left to right the maximum possible string — even if the overall match would fail.
→ Write patterns that are safe against fall-through.
- ▶ “loop body may accept empty string”

Simple workflow setup

LPeg key concepts: capture

capture

Danger!

Instead of a conclusion...

info m4:

Some people find 'm4' to be fairly addictive. They first use 'm4' for simple problems, then take bigger and bigger challenges, learning how to write complex sets of 'm4' macros along the way. Once really addicted, users pursue writing of sophisticated 'm4' applications even to solve simple problems, devoting more time debugging their 'm4' scripts than doing real work. Beware that 'm4' may be dangerous for the health of compulsive programmers.

Danger!

Instead of a conclusion...

info m4:

Some people find 'm4' to be fairly addictive. They first use 'm4' for simple problems, then take bigger and bigger challenges, learning how to write complex sets of 'm4' macros along the way. Once really addicted, users pursue writing of sophisticated 'm4' applications even to solve simple problems, devoting more time debugging their 'm4' scripts than doing real work. Beware that 'm4' may be dangerous for the health of compulsive programmers.

Same might easily hold for LPeg.